

Dissertation
zur Erlangung des Doktorgrades der Fakultät für Angewandte
Wissenschaften der Albert-Ludwigs-Universität Freiburg im Breisgau

3D Invariants for Automated Pollen Recognition

Olaf Ronneberger

2007

Dekan: Prof. Dr. Bernhard Nebel

Prüfungskommission: Prof. Dr. Rolf Backofen (Vorsitz)
Prof. Dr. Peter Thiemann (Beisitz)
Prof. Dr.-Ing. Hans Burkhardt (Gutachter)
Prof. Dr.-Ing. Friedrich M. Wahl (Gutachter)

Datum der Disputation: 24.9.2007

Acknowledgements

In the first place, I want to thank my supervisor Prof. Dr. Hans Burkhardt. All the time he supported my work with many valuable suggestions and fruitful discussions. Moreover, he created an excellent working environment, and he allowed me great latitude which have extremely stimulated me. Furthermore I am grateful to Prof. Dr. Friedrich Wahl for his agreement to act as the second referee, and I want to thank Prof. Dr. Backofen and Prof. Dr. Thiemann for spending their time on reading the thesis and asking many good questions during the disputation.

Such a large project like the pollenmonitor can never be realized alone. I am very thankful to all partners who have contributed to this project. The most prominent is Dr. Eckart Schultz from the German Weather Service, the “father” of the pollenmonitor. Without his optimism, his engagement, his endurance and his ability to convince the deciders at the various levels, the project would never have reached the present state. The second most important partner was Ulrich Heimann also from the German Weather Service. We three (Ulli, the pollenmonitor and I myself) have spent many days, evenings and weekends together, and nothing worked anymore each time one of us was indisposed.

On the algorithmic and the software side I have benefited a lot from all the other PhD students, regular students and HiWi’s at our institute. I am especially grateful to Qing Wang for the very fruitful team work within the pollen project, to Bernard Haasdonk for introducing me into the secrets of support vector machines including the many fruitful discussions, the proof-reading of this thesis and his suggestion for the alternative title “3D Invariants for Automated Pollen Recognition By Happy Smiley Rotations and Haarige Integrations”. Moreover this work benefited a lot from all the student projects and master theses which originated in our group and which I was responsible for during my PhD time. For their contributions I am indebted to (in chronological order) Rafael Baumgartner, Thorsten Schmidt, Anselm Vossen, Christoph Keller, Janis Fehr, Janina Schulz, Margret Keuper, and Mario Emmenlauer.

Furthermore I am grateful to Qing Wang, Janis Fehr, Janina Schulz, Thorsten Schmidt, Regula Gehrig, Eckart Schulz, and my father for proofreading of this thesis; Sabine Haxelmans and Roland Nitschke for the tedious recording of the confocal pollen data set; Jörg Mechnich, Florian Pigorsch, Christoph Keller, Jürgen Hense, Matthias Schlachter for writing and maintaining several programs and libraries for the hardware control, the data management and the visualization; Ildiko Buzakiss, Renate Smukalla and Ulrich Heimann for the manual labeling of the nearly 200,000 particles of the pollen monitor dataset; and last not least my family for all the support and for the frequent questions: when will the thesis be finally complete.

Freiburg, October 2007

Olaf Ronneberger

Abstract

In biomedical research we can see a clear tendency towards the study of cells and whole organisms in their natural 3D environment, away from the studies of isolated flat cells, that were grown in a synthetic 2D environment and were analyzed with 2D techniques. Another trend in biomedical research is the ever increasing need for high-throughput experiments, e. g., to retrieve statistically relevant information. The equipment for the automatic recording of great amounts of 3D volumetric data is available in many research laboratories today, but the recorded 3D data are still manually evaluated which has become the main bottleneck in many applications.

In this thesis the development of 3D invariants for the recognition of biological structures is described. These invariants are based on the Haar-integration-framework of [Schulz-Mirbach \(1995b\)](#). The invariance properties are reached by an integration over the desired transformation groups, e. g. rotation and translation. Several important aspects of the application of the Haar-integration framework to 3D volumetric data sets of biological structures are described. The most important extension is the introduction of deformation models such that the resulting features are robust to elastic deformations of the structures. Another important aspect is the reached robustness to even non-linear gray-scale transformations, that allow certain variations of the recording parameters between the training and the test objects. The direct computation of these invariants is computationally very expensive. Several new techniques are introduced that allow a fast computation of the invariants by means of the FFT, by the expansion of the integral into spherical harmonics series or by simultaneous computation of multiple invariants based on invertible vectorial kernel functions. Furthermore voxel-wise invariants are introduced for a simultaneous segmentation and recognition of 3D structures. Vectorial invariants are developed for a fast and reliable detection of spherical objects in cluttered environments.

A very challenging application that demands many of the requirements to be fulfilled which are given in the biomedical research is the recognition of pollen grains. The high number of different pollen grains from different plants contain very different kinds of structures that have to be identified. For a part of these structures clear one-to-one correspondences can be identified, while for the other part of these structures only the statistical properties match.

In the given application we use microscopically recorded images to recognize a real-world object. For the correct interpretation of the gray values it is important to understand the different effects within a microscope. The main four steps, illumination, interaction of the object with the light, transformation of the emitted light, and the recording of the light are explained.

In pollen recognition (as in many other applications) we should differentiate between those results that can be reached within a clean and well-controlled laboratory environment and those results of a real-world routine application. In this thesis two representative data sets for these scenarios are used: The first one (denoted as “confocal data set” here) is a typical laboratory data set: The pollen were collected directly from the corresponding plants. They were carefully prepared on one slide per taxon and were manually recorded as a full 3D volumetric data set with confocal laser scanning microscopy. Due to the high costs of such a system and the time-consuming operation, only a small data set containing 389 pollen grains of 26 different taxa (15 grains per taxon) was recorded. The preparation and recording applied here guarantees a 100% correct labeling of the pollen grains and contains the lowest possible degree of distortion due to optical effects. On the other hand the used samples do not represent all variations within each taxon, such as different genera, species or subspecies, different growth conditions of the plants, etc. Furthermore this data set does not contain deformed, contaminated or agglomerated pollen grains or pollen grains at different levels of degradation and it does not contain the vast amount of other particles that are found in real air samples.

The second data set (denoted as “pollenmonitor data set”) is a typical real-world data set. It was automatically collected, prepared and recorded by the first prototype of the pollenmonitor during the pollen season 2006 in Freiburg and Zürich. It contains about 22,750 pollen grains from 33 taxa together with about 170,000 other spherical airborne particles.

The detection of the pollen grains in the recorded volumetric data sets is done by vectorial voxel-wise invariants (named “MiSP detector” for Microscopical Spherical Particles). The precise segmentation of the detected pollen grains is performed by standard algorithms that were adapted to the given requirements. For the confocal data set a graph-cut-based segmentation was used. For the pollenmonitor data set a combination of a modified canny-edge detector, a model-based weighting of the found edges and a snake approach was applied. The feature extraction is done by the upper described invariants (named “MiSP invariants”) which exhibit a high robustness to elastic deformations and nonlinear gray-value transformations. At the same time these invariants are sensitive to subtle structural changes that cannot be modeled by deformations. The classification of the objects is performed with support vector machines that were optimized to deal with the very high dimensionality of the extracted feature vectors (e.g., 87,296 for the pollenmonitor data set).

The recognition rate on the confocal data set (389 pollen grains of 26 taxa) is very high. In a leave-one-out validation with a simple 1 Nearest Neighbor classifier only 3 of the 389 pollen grains were misclassified (99.2% recognition rate). With a support vector machine only 2 incorrectly classified pollen grains were obtained (99.5% recognition rate). In the second case the misclassified pollen were confused with another genus from the same family. The best recognition rates obtained by other approaches on typical “laboratory” data sets are 89% on a data set containing 3,800 pollen grains from 19 taxa, or 70% on a data set containing 16,220 pollen grains from 80 taxa.

For the rating of the results on the pollenmonitor data set another statistical measure that the recognition rate is useful. In pollen samples from the ambient air we are confronted with a very high number of particles other than pollen that have to be correctly rejected by the recognition system. Here the precision (fraction of correctly classified pollen within the objects, that were classified as pollen) is the most important statistical measure. The validation was done by splitting the pollenmonitor data set (22,750 pollen grains from 33 taxa + about 170,000 dust particles) into a training set and a test set. The training set contains only samples from Freiburg, while the test set contains samples from Freiburg and Zürich. The obtained precision for all 33 pollen taxa is 96.7% at a recall of 84.3%. The results for the five highly allergenic pollen taxa in this data set (that are monitored for the pollen forecasts) are 98.5% precision at a recall of 86.5%. The best results obtained by other approaches on a real-world data set (3104 pollen grains from 8 taxa + about 30,000 dust particles) is a precision of 30% at a recall of 64.9%.

The results obtained with the proposed MiSP invariants are significantly better than the state of the art in pollen recognition. Furthermore these techniques have proven to work not only on a small laboratory-type data set but also on a very large real-world data set. Especially the reached robustness to elastic deformations and the full 3D approach seems to be a key requirement for a reliable recognition of biological structures. We think that the new findings will be a valuable foundation for many further developments in the field of the recognition of biological structures in 3D volumetric data sets.

Zusammenfassung

In der bio-medizinischen Forschung ist eine klare Tendenz erkennbar, die weg von der klassischen 2D Analyse einzelner flacher Zellen in einer synthetischen zweidimensionalen Umgebung hin zur Untersuchung von Zellen und ganzen Organismen in ihrer natürlichen dreidimensionalen Umgebung geht. Außerdem wächst der Bedarf an Experimenten, die an einer großen Zahl gleicher Systeme durchgeführt werden, um z.B. Daten für statistische Aussagen zu gewinnen. Die apparative Ausstattung zur Aufnahme großer Mengen dreidimensionaler Volumendaten ist heutzutage in vielen Forschungslaboratorien verfügbar, aber die Auswertung dieser Daten erfolgt immer noch manuell. Diese zeitaufwändige Auswertung ist inzwischen der Haupt-Engpass in vielen derartigen Anwendungen.

In der vorliegenden Dissertation wird die Entwicklung von 3D Invarianten zur Erkennung von biologischen Strukturen beschrieben. Diese Invarianten wurden mit Hilfe eines Schemas entwickelt, das von Schulz-Mirbach (1995b) angegeben wurde und auf der Haar-Integration beruht. Die benötigten Invarianzeigenschaften werden dabei durch die Integration über die entsprechende Transformationsgruppe wie z.B. Rotation und Translation erreicht. Verschiedene Erweiterungen des Haar-Integrations-Schemas werden beschrieben, die für die Anwendung auf 3D Volumendaten biologischer Strukturen benötigt werden. Die wichtigste Erweiterung ist dabei die Einführung eines Deformationsmodells, so dass die resultierenden Merkmale robust gegenüber elastischen Deformationen der betrachteten Strukturen werden. Ein weiterer wichtiger Aspekt ist die erreichte Robustheit gegenüber nicht-linearen Grauwert-Transformationen, die gewisse Veränderungen der Aufnahmebedingungen zwischen den Aufnahmen der Trainings- und der Test-Objekte erlaubt.

Die direkte Berechnung dieser Invarianten ist extrem rechenaufwändig. Mehrere neue Verfahren zur schnellen Berechnung der Invarianten durch den Einsatz der FFT, durch die Entwicklung des Integrals in eine Reihe von Kugelfunktionen, oder durch die simultane Berechnung von mehreren Invarianten, basierend auf invertierbaren vektoriiellen Kernfunktionen, werden vorgestellt. Außerdem werden voxelweise berechnete Invarianten eingeführt, mit deren Hilfe die Segmentierung und Erkennung von 3D Strukturen in einem Schritt durchgeführt werden kann. Für eine schnelle und zuverlässige Detektion von kugelförmigen Objekten in sehr heterogenen Umgebungen werden vektorielle Invarianten entwickelt.

Eine sehr anspruchsvolle Anwendung, bei der viele der Anforderungen erfüllt werden müssen, die unter anderem in der bio-medizinischen Forschung auftreten, ist die Erkennung von Pollen. Die große Anzahl von verschiedenen Pollen der verschiedenen Pflanzen enthalten sehr unterschiedliche Typen von Strukturen, die identifiziert werden müssen. Für einen Teil dieser Strukturen können klare ein-zu-eins Korrespondenzen gefunden werden, während für einen anderen Teil der Strukturen nur die statistischen Eigenschaften übereinstimmen.

In der vorliegenden Anwendung soll ein reales Objekt anhand von mikroskopisch aufgenommenen Bildern erkannt werden. Für die korrekte Interpretation der Grauwerte ist es wichtig, die verschiedenen Effekte innerhalb des Mikroskops zu verstehen. Daher werden die vier wichtigsten Schritte, die Beleuchtung, die Wechselwirkung des Lichtes mit dem Objekt, die Transformation des emittierten Lichtes und die Aufnahme des Lichtes umfassend erläutert.

In der Pollen-Erkennung (so wie in vielen anderen Anwendungen) sollte man zwischen solchen Ergebnissen unterscheiden, die in einer sauberen, gut kontrollierbaren Labor-Umgebung erreicht werden können, und solchen Ergebnissen, die in einer realen Routineanwendung erreicht werden. In dieser Dissertation werden zwei repräsentative Datensätze für diese beiden Szenarien verwendet. Der erste (hier als "Konfokal-Datensatz" bezeichnet) ist ein typischer Labor-Datensatz: Die Pollen wurden direkt von den entsprechenden Pflanzen gesammelt und sorgfältig auf einem Objektträger pro Taxon präpariert. Von jedem Pollenkorn wurde manuell ein 3D Volumendatensatz mit einem konfokalen Laser-Scannig-Mikroskop aufgenommen. Aufgrund der hohen laufenden Kosten eines solchen Systems und der zeitaufwändigen Aufnahme wurde nur ein relativ kleiner Datensatz mit 389 Pollen von 26 verschiedenen Taxa (15 Pollen pro Taxon) erzeugt. Dieses Vorgehen garantiert ohne Ausnahme die ko-

rrekte Zuordnung der Pollen zu der entsprechenden Pflanzenart und die geringst mögliche Verfälschung der Aufnahmen durch optische Effekte. Andererseits werden in diesem Datensatz nicht alle möglichen Variationen innerhalb eines Taxons abgedeckt. Dies betrifft vor allem Variationen aufgrund von verschiedenen Genera, Arten oder Unterarten innerhalb des Taxons, oder aufgrund von verschiedenen Wachstumsbedingungen der Pflanzen. Desweiteren enthält dieser Datensatz keine deformierten, kontaminierten oder agglomerierten Pollen, sowie keine Pollen in verschiedenen Zersetzungs-Zuständen und vor allem nicht die enorm hohe Zahl anderer Partikel, die in echten Luftstaubproben auftreten.

Der zweite Datensatz (hier als "Pollenmonitor Datensatz" bezeichnet) ist ein typischer "Real-World"-Datensatz. Die Proben wurden in der Pollensaison 2006 vollautomatisch mit dem ersten Prototypen des Pollenmonitors gesammelt, präpariert und aufgenommen an den Standorten Freiburg und Zürich. Dieser Datensatz enthält insgesamt ca. 22.750 Pollen von 33 Taxa und etwa 170.000 andere rundliche luftgetragene Partikel.

Die Detektion der Pollen in den aufgenommenen volumetrischen Datensätzen erfolgt mit vektoriellen voxel-weisen Invarianten (hier als "MiSP detector" für "Microscopical Spherical Particles" bezeichnet). Für die nachfolgende Segmentierung der detektierten Partikel werden angepasste Standardverfahren eingesetzt: Eine graph-cut basierte Segmentierung für den Konfokal-Datensatz und eine Kombination aus einem modifizierten Canny-Edge-Detector, einer modellbasierten Gewichtung der gefundenen Kanten und einem Snake-Ansatz für den Pollenmonitor-Datensatz.

Für die Merkmals-Extraktion werden die oben beschriebenen Invarianten eingesetzt (hier als "MiSP invariants" bezeichnet), die eine hohe Robustheit gegenüber elastischen Deformationen und nicht-linearen Grauwerttransformationen aufweisen. Trotzdem sind diese Invarianten noch sensitiv für subtile strukturelle Änderungen, die nicht durch eine elastische Deformation modelliert werden können. Für die Klassifikation der Objekte werden Support-Vektor-Maschinen eingesetzt, die für die Verarbeitung von sehr hoch-dimensionalen Merkmalsvektoren optimiert wurden (z.B. 87.296 Dimensionen beim Pollenmonitor Datensatz).

Die erreichte Erkennungsrate auf dem Konfokal-Datensatz (389 Pollen aus 26 Taxa) ist sehr hoch. In einer "leave-one-out validation" mit einem einfachen 1-nächster-Nachbar Klassifikator wurden nur drei Pollenkörner falsch klassifiziert (99.2% Erkennungsrate). Beim Einsatz einer Support-Vektor-Maschine sinkt diese Zahl auf 2 falsch klassifizierte Pollenkörner (99.5% recognition rate), wobei diese zwar einem anderen Genus zugewiesen wurden, der aber zur selben Familie gehört. Die besten Erkennungsraten mit anderen Ansätzen auf typischen Labor-Datensätzen liegen bei 89% bei einem Datensatz mit 3.800 Pollen aus 19 Taxa oder 70% bei einem Datensatz mit 16.220 Pollen aus 80 Taxa.

Für die Bewertung der Ergebnisse auf dem Pollenmonitor-Datensatz ist ein anderes statisches Maß als die einfache Erkennungsrate sinnvoll. Pollenproben aus der Umgebungsluft enthalten eine sehr hohe Anzahl von anderen luftgetragenen Partikeln die von dem Erkennungssystem korrekt zurückgewiesen werden müssen. In diesem Fall ist die Präzision (Anteil der korrekt klassifizierten Pollen innerhalb aller Objekte, die als Pollen klassifiziert wurden) das wichtigste statistische Maß. Die Validierung des Systems auf dem Pollenmonitor-Datensatz (22.750 Pollen aus 33 Taxa + etwa 170.000 Staubpartikel) wurde durch eine Aufteilung in einen Trainings-Datensatz und einen Test-Datensatz durchgeführt. Der Trainings-Datensatz enthält nur Proben aus Freiburg, während der Test-Datensatz Proben aus Freiburg und aus Zürich enthält. Die erreichte Präzision bei der Betrachtung aller 33 Taxa ist 96.7% bei einer Erkennungsrate von 84.3%. Die Ergebnisse für die 5 hoch-allergenen Pollenarten (die für die Pollenflugvorhersagen relevant sind) liegen noch leicht darüber. Dort konnte eine Präzision von 98.5% bei einer Erkennungsrate von 86.5% erreicht werden. Die besten Ergebnisse, die mit anderen Ansätzen auf einem "Real-World"-Datensatz (3104 Pollen aus 8 taxa + etwa 30.000 Staubpartikel) erreicht wurden, liegen bei 30% Präzision bei einer Erkennungsrate von 64.9%.

Die erzielten Ergebnisse mit den vorgeschlagenen "MiSP invariants" sind signifikant besser als der Stand der Technik. Darüber hinaus konnte demonstriert werden, dass diese Techniken nicht nur auf kleinen Labor-Datensätzen sondern auch auf sehr großen "Real-World"-Datensätzen sehr gute Ergebnisse liefern. Insbesondere die erreichte Robustheit gegenüber elastischen Deformationen und der

3D Ansatz scheinen eine wesentliche Voraussetzung für die verlässliche Erkennung von biologischen Strukturen darzustellen. Wir glauben, dass die neuen Erkenntnisse eine wertvolle Grundlage für viele weitere Entwicklungen im Bereich der Erkennung von biologischen Strukturen in 3D Volumendaten sind.

Contents

1	Introduction	14
1.1	Motivation	14
1.1.1	The Need for Pattern Recognition in 3D Volumetric Data Sets	14
1.1.2	Pollen Forecasts for Allergic People	15
1.1.3	Shortcomings of the Manual Pollen Counting	16
1.1.4	Goals	16
1.2	State of the Art	17
1.2.1	Automated Pollen Recognition	17
1.2.2	The Haar-Integration Framework	22
1.3	Main Contributions	22
1.4	Structure	25
1.5	Notation	27
2	Microscopic Imaging	28
2.1	Optics	28
2.1.1	Illumination	28
2.1.2	Interaction of Light with Objects	29
2.1.3	Optical Transformation of the Emitted Light	30
2.2	Recording of the Light	33
2.2.1	Charge Coupled Devices (CCD) Sensor	33
2.3	Reduction of Out-Of-Focus Light	35
2.3.1	Pure Hardware Approach: Confocal Microscopy	35
2.3.2	Pure Software Approach: Deconvolution	37
2.3.3	Combined Hardware+Software Approach: Structured Light	37
2.4	Improvement of the Dynamic Range (HDR Images)	38
2.4.1	Computing the Number of Photons from the Gray Values	41
2.4.2	Compensation of Uniform Bleaching Effects and Offsets	42
2.4.3	Compensation of Nonuniform Bleaching Effects	42
2.5	Improvement of Homogeneity	44
2.5.1	Compensation of Errors from Illumination and Recording Step (Static Shading Correction)	44
2.5.2	Compensation of Inhomogeneities due to Out-Of-Focus Dust (Dynamic Shading Correction)	44
3	Data Sets	46
3.1	Confocal Data Set	46
3.1.1	Collection, Preparation and Recording	46
3.1.2	Representative Table	48
3.1.3	Left-out Taxa	53
3.1.4	Discussion	54
3.2	Pollenmonitor Data Set	55
3.2.1	Collection	55
3.2.2	Preparation	57

3.2.3	Microscopic Recording	58
3.2.4	Evaluated Area per Air Sample	61
3.2.5	Labeling	61
3.2.6	Discussion	64
4	The Haar Integration (HI) Framework	70
4.1	Invariance by Integration over the Transformation Group	70
4.2	Direct Computation Using General N-Point-Kernel Functions	71
5	Analysis of the Properties of the HI framework	73
5.1	Correct Sampling of the Transformation Group for the Computation of Invariants	73
5.1.1	Consequences of Under-Sampling	76
5.2	Maximal Information that can be Sensed with Two-Point-Kernels and the Euclidean Transformation Group	77
5.2.1	The Gray Value Co-Occurrence Distribution “GVCD”	77
5.2.2	Using the GVCD to Estimate the Robustness of the Invariants	80
5.2.3	Using the GVCD to Find Appropriate Kernel Functions	80
5.2.4	Using the GVCD to Analyze the Completeness of the Invariants	80
6	Extensions of the HI framework	82
6.1	Constraints for Transformations, Kernel Functions and Data Sets	82
6.2	Categorization of the Transformation Groups	82
6.2.1	Case 1: Group of Rotations or Cyclic Shifts	83
6.2.2	Case 2: Group of Euclidean Transformations, Scalings and Similar	84
6.3	Ensuring a Finite Result for Infinite Transformations	84
6.3.1	Implicit Limitation of the Integral by Appropriate Kernel Functions	84
6.3.2	Explicit Limitation of the Integral by Partial Haar Integration	84
6.3.3	Explicit Limitation of the Integral by Probability-Weighted Haar Integration	86
6.3.4	Implicit Limitation of the Integral by the Use of Synthetic Channels	86
6.4	Multi-Scale Approach to Enhance Richness of Sparse Kernel Functions	88
6.5	Integration of Prior Knowledge into the HI framework	88
6.5.1	Constraints of the Different Techniques	89
6.6	Parameterized Kernel Functions	90
6.7	Voxel-Wise Invariants for Simultaneous Segmentation and Recognition	91
6.7.1	Limitations of Centered Two-Point-Kernels for Voxel-Wise Invariants	92
6.8	Voxel-Wise Vectorial Invariants to Detect Spherical Structures	92
7	Fast Computation of the Invariants	95
7.1	Separable N-Point-Kernel Functions	95
7.2	Using the Equivalence of Kernels	95
7.3	Categorization of the Methods	96
7.4	Two-Point-Kernels with 3D Euclidean Transformations	96
7.4.1	Alternative Approach	98
7.4.2	Comparison of the Computational Costs of the Two Approaches	99
7.5	Three-Point-Kernels with 3D Translations and Planar Rotations	99
7.6	N-Point-Kernels with 3D Translations and Planar Rotations	102
7.7	Three-Point-Kernels with 3D Euclidean Transformations	104
7.7.1	Convolution on the 2-Sphere	107
7.8	Variations of the Kernel Angles	108
7.9	Fast Computation of Voxel-Wise Invariants with Separable Kernels	109

7.10	Fast Computation of Voxel-Wise Invariants with Sparse Vectorial Kernels	109
7.11	Computational Costs	110
8	Detection and Segmentation	112
8.1	2D Segmentation using Morphological Techniques	112
8.1.1	Preprocessing of Data Sets with Particles in Different Planes	115
8.2	Detection with Voxel-Wise Invariants (MiSP Detector)	115
8.3	2D Boundary Extraction with Snakes	118
8.4	3D Segmentation with Graph-Cuts	118
9	Specialized Invariants for Microscopical Spherical Particles (MiSP)	121
9.1	The Transformation Group for MiSP Invariants	121
9.1.1	Invariance to Translation by Normalization	122
9.1.2	Invariance to Rotation.	122
9.1.3	Selection of an Appropriate Model for the Deformations	123
9.1.4	Integration of the Deformation Models into the Transformation Group	124
9.2	The Kernel Functions for MiSP Invariants	125
9.2.1	Robustness to Arbitrary Gray Value Transformations.	125
9.2.2	Synthetic Channels with Segmentation Results.	127
9.2.3	Resulting Kernel Functions.	128
9.3	Resulting Haar Integral	129
9.4	Fast Simultaneous Computation of the Invariants	130
9.5	Interpretation	132
9.6	Pseudo Code	137
10	Experimental Results	138
10.1	Relevant Statistical Measures	138
10.1.1	Estimation of the Pollen Concentration in the Air	138
10.1.2	Discussion	142
10.1.3	A Better Measure for the False-Positives	142
10.2	Results on the Confocal Data Set	143
10.2.1	Euclidean Invariants on the Confocal Data Set	143
10.2.2	Graph-Cut Segmentation on the Confocal Data Set	146
10.2.3	MiSP Invariants on the Confocal Data Set	150
10.2.4	Comparison to the State of the Art	154
10.3	MiSP Detector on Pollenmonitor Data Set	156
10.4	MiSP Invariants on Pollenmonitor Data Set	161
10.4.1	Features	161
10.4.2	Training and Test Data Sets	161
10.4.3	Optimization for High Precision or High Recall	162
10.4.4	Results	163
10.4.5	Detailed Results	165
10.4.6	Comparison to the State of the Art	173
10.5	Summary of Further Results	178
10.5.1	Deconvolution	178
10.5.2	Use of Additional Microscopical Techniques	179
10.5.3	Automatic Kernel Selection with the GVCD	179
10.5.4	Voxel-wise Invariants for Segmentation and Recognition of Cell Nuclei	181

11 Summary, Conclusion and Outlook	183
11.1 Summary and Conclusions	183
11.1.1 Conclusions and Intermediate Results that led to the Final MiSP Invariants	184
11.1.2 Segmentation	186
11.1.3 Application of MiSP Invariants to Pollen Recognition	186
11.1.4 High Dimension of the Feature Vectors	188
11.1.5 Final Conclusions	189
11.2 Outlook	189
A Highly Allergenic Pollen Taxa	191
A.1 Artemisia (Beifuß)	192
A.2 Alnus (Erle)	194
A.3 Alnus viridis (Alpen-/ Grün-Erle)	196
A.4 Betula (Birke)	198
A.5 Corylus (Hasel)	200
A.6 Poaceae (Süßgräser)	202
A.7 Secale (Roggen)	206
A.8 Ambrosia (Traubenkraut)	208
B Moderately Allergenic Pollen Taxa	210
B.1 Carpinus (Hainbuche)	210
B.2 Ostrya Carpinifolia (Hopfenbuche)	212
B.3 Quercus (Eiche)	214
B.4 Fraxinus (Esche)	216
B.5 Plantago (Wegerich)	218
B.6 Rumex (Ampfer)	220
C Lowly Allergenic Pollen Taxa	222
C.1 Castanea (Eßkastanie)	222
C.2 Fagus (Rotbuche)	224
C.3 Populus (Pappel)	226
C.4 Salix (Weide)	228
D Non-Allergenic Pollen Taxa	230
D.1 Acer (Ahorn)	230
D.2 Humulus (Hopfen)	232
D.3 Chenopodium (Gänsefuß)	234
D.4 Compositae (Korbblüter)	236
D.5 Cruciferae (Kreuzblütengewächse)	238
D.6 Aesculus (Roßkastanie)	240
D.7 Juglans (Walnuß)	242
D.8 Larix (Lärche)	244
D.9 Picea (Fichte)	246
D.10 Pinus (Kiefer)	248
D.11 Platanus (Platane)	250
D.12 Taxus (Eibe)	252
D.13 Tilia (Linde)	254
D.14 Ulmus (Ulme)	256
D.15 Urtica (Brennnessel)	258
D.16 Cyperaceae (Sauergräser)	260

D.17 Sambucus (Holunder)	262
E Categorization of Pollen Taxa into Different Allergenic Levels	264
Bibliography	266

1 Introduction

1.1 Motivation

1.1.1 The Need for Pattern Recognition in 3D Volumetric Data Sets

All structures in the real world are (at least) three-dimensional. Every two-dimensional representation of these 3D structures involves a substantial loss of information. While there are several applications where we can ensure in advance that only irrelevant information is lost, we can identify many important applications where this loss of information may have serious consequences. One prominent example of this kind are the medical applications, where wrong conclusions may be drawn, e. g., from conventional 2D radiographs. Consequently, the medicine is one of the driving forces for the development of 3D techniques, such as CT (“computed tomography”) or MRT (“magnetic resonance tomography”) which allow to record the three dimensional distribution of the quantity which the particular technique is sensitive to, like the absorption of X-rays or the density of protons.

Within the last decade the biomedical research has become another driving force for the development of 3D sensors (especially in the 3D microscopy). While in this field wrong conclusions due to imperfect 2D representations usually have not such serious consequences as in medicine, we can see a clear tendency towards the study of cells and cell clusters including whole organisms in their natural 3D environment, away from the studies of isolated flat cells, that were grown in a synthetic 2D environment and were analyzed with 2D techniques.

Another trend in biomedical research is the ever increasing need for high-throughput experiments, e. g., to retrieve statistically relevant information. The equipment for the automatic recording of great amounts of 3D volumetric data is available in many research laboratories today, but the recorded 3D data are still manually evaluated which has become the main bottleneck in many applications.

A closer analysis of the requirements for an automated analysis of the microscopical 3D data sets reveals significant differences between biomedical research and typical medical applications. In medical applications, the orientation of the object in the data set is mostly well known or can be computed by means of registration techniques, if, e. g., two data sets (usually from the same individual at the same time) need to be aligned. Furthermore the medical data sets are usually interpreted by human specialists, such that the research mainly focuses on the preprocessing (e. g. denoising), the visualization, the registration and the segmentation of these data sets, e. g. in order to extract physical parameters like the size or the precise location of an object.

In contrast to this no prominent direction usually exists in cell and tissue samples. So normalization approaches that try to rotate the recorded object into a standard orientation are impractical in most cases. Furthermore no one-to-one correspondences between the corresponding structures of two different individuals exist in many cases. The biological structures only share the same development process but the resulting patterns, e. g., a venation pattern or the surface pattern on a pollen grain are never identical but have only the same statistical properties.

Another important feature in the biomedical research compared to the typical medical application is the great diversity of the examined organisms, the frequent change of the structures that have to be recognized, and the information that needs to be extracted.

Therefore, the main challenge to the pattern recognition in this field is not only to cope with the large variations that are found in biological structures but also the need for a fast and easy adaption of the system to new problems.

Regarding the mentioned shortcomings of the approaches that are based on registration and normalization, we anticipate that feature based approaches are better suited to fulfill the outlined requirements. So what we need is a general framework that allows to construct features which exhibit the particular invariance and the robustness that are specified by the given application.

A framework which has this potential has been introduced by [Schulz-Mirbach \(1995a\)](#). Since then it has been continuously extended and has been successfully adapted to many different applications. The method allows to construct any number of invariants of a data set by Haar integration over the desired transformation group. A brief description of this framework (abbreviated as HI-framework in the following) is given in chapter 4ff.

A very challenging application which demands to fulfill many of the aforementioned requirements is the recognition of pollen grains. Taking the recognition of pollen grains as a reference application has several advantages.

- There is a clear long term need for the automation of pollen counting which will be outlined in the next section. This need is independent of current research problems.
- While in many other applications a half-automated solution will be sufficient to answer the underlying questions, the pollen recognition system must operate fully autonomously, comparable to the challenges in very-high-throughput experiments.
- The high number of different pollen grains from different plants contain very different kinds of structures that have to be identified. For a part of these structures clear one-to-one correspondences can be identified, while for other structures only the statistical properties match.

1.1.2 Pollen Forecasts for Allergic People

About 15%–20% ([Winkler et al., 2001](#)) of the human population is allergic to pollen. Even though there are several pharmaceuticals on the market that are able to reduce or suppress the symptoms, many people prefer to keep away from high pollen concentrations by avoiding certain regions or by staying indoor during the daily pollen concentration peak(s). The allergic potential of the six most important allergenic pollen taxa in Germany is listed in table 1.1. A complete categorization of the pollen taxa into different allergenic levels is provided in appendix E.

Table 1.1: Allergenic potential of the most important pollen taxa in Germany

main flowering preiod	botanic name	Pollen Taxa		Allergic Stress at conventionally measured pollen concentration [1/m ³]		
		English name	German name	low	moderate	high
January–March	<i>Alnus</i> ^(A.2)	Alder	Erle	1-10	11-100	>100
January–March	<i>Corylus</i> ^(A.5)	Hazel	Hasel	1-10	11-100	>100
March–April	<i>Betula</i> ^(A.4)	Birch	Birke	1-10	11-50	>50
May–August	<i>Poaceae</i> ^(A.6)	Gras	Süßgräser	1-5	6-30	>30
June	<i>Secale</i> ^(A.7)	Rye	Roggen	1-2	3-6	>6
July–September	<i>Artemisia</i> ^(A.1)	Mugwort	Beifuß	1-2	3-6	>6

It should be noted that the most important pollen taxon of these six taxa is the *Poaceae*^(A.6) (grass), because 80% of the pollen-allergy sufferer (including the author himself) are allergic to this type of pollen.

1 Introduction

The state of the art of measuring the pollen concentration in the air is still the time consuming and expensive manual pollen counting¹ done by experienced microscopists. For this technique air samples are taken every day by means of a volumetric spore trap, the trapped particles are manually prepared on microscopical slides and are evaluated under the microscope. Due to the high costs only very small areas on these slides are inspected (usually an area that contains the particles of 1 m³ of air), and the countings are performed only every second day in Germany and only once a week in Switzerland.

1.1.3 Shortcomings of the Manual Pollen Counting

Late Availability and Varying Quality of the Data

The main shortcomings of the current technique are the late availability of the data and (in Germany) its varying often unknown quality. These two facts make the pollen forecasts more difficult and reduce their reliability.

In Germany the pollen countings at the about 45 stations are executed by about 100 different persons with varying experience. In Switzerland the pollen are counted centrally in Zürich and Payerne by long-experienced experts. At the beginning of the project we performed a direct comparison between the two main experts who were instructed to label independently the same pollen grains on a computer-controlled microscope such that a one-to-one correspondence of the labels for each pollen grain was guaranteed. The result was a nearly 100% correspondence of the labels. The only disagreements were found for some *Taxus*^(D,12) or *Juniperus* pollen grains which can hardly (if at all) be distinguished by visual inspection which both are non-allergenic by the way. This 100% recognition rate for all relevant pollen taxa by human experts gave reason to define this as the “golden standard” and to base our training on manually labeled samples.

Insufficient Evaluated Air Volumes

Considering the allergenic potential of some pollen taxa (see table 1.1), the evaluated air volume of 1 m³ in the routine environment is in most situations much too low to provide statistically reliable data (see (Comtois et al., 1999) for an extensive experimental analysis). To give an optimistic estimate of the expected uncertainty, we assume a spatial and temporal constant pollen concentration in the air and a perfect sampling of all particles contained in 1 m³ of air. In this ideal case the statistics can be modeled by a Poisson distribution, or by the inverse Poisson distribution, if the true pollen concentration shall be predicted from the counted number of pollen grains. Especially for the very low pollen concentrations the uncertainty of the predicted value is unacceptably large. If the taken 1 m³-air-sample contains e. g. 5 pollen grains, the true concentration in the air will be between 1.8 and 10.9 pollen/m³ (95% confidence interval). Some additional results from the inverse Poisson distribution are listed in table 1.2. Please note that these calculations provide only the lowest estimate of the uncertainty. In the real world, we find highly temporal variations of the pollen concentrations over the day. So the uncertainty for a prediction of the current pollen concentration from these data is even much higher.

1.1.4 Goals

Among the shortcomings listed in the previous section the most important one is smallness of the evaluated air volume. While the volume can easily be enlarged in scientific studies by evaluating larger areas of the slides (as long as sufficient personal capacities are available) it is not affordable to significantly enlarge the volume in the daily routine countings, keeping in mind that a reduction of the uncertainty by a factor of 2 needs an enlargement of the air volume by a factor of 4. For a sufficient

¹A standardized protocol about manual pollen sampling and counting is available on the internet by the Pan- American Aerobiology association <http://www.paaa.org>

Table 1.2: Uncertainty for the true pollen concentration if only 1m^3 of the air sample is evaluated, modeled by the inverse Poisson distribution.

counted pollen grains within a 1m^3 -air-sample	true pollen concentration [pollen/ m^3]	
	(68% confidence interval)	(95% confidence interval)
0	0 – 1.15	0.0 – 3.0
1	0.27 – 2.5	0.04 – 4.8
2	0.86 – 3.9	0.3 – 6.4
3	1.6 – 5.1	0.7 – 7.9
4	2.3 – 6.4	1.2 – 9.4
5	3.0 – 7.6	1.8 – 10.9
10	7.1 – 13.5	5.0 – 17.6
20	15.8 – 24.8	11.7 – 28.9
50	43.2 – 57.4	37.4 – 65.2
100	90.3 – 110.3	81.6 – 120.9

temporal resolution the evaluated air volume has to be even larger, and a real-time acquisition of the current pollen concentration is not possible at all by means of manual countings.

All these facts gave reason for the development of an automated pollen recognition system to be directly integrated into a fully automated pollenmonitor that provides the current pollen concentration online for the pollen-allergy sufferer.

The system will be based on microscopical image recognition which has several important advantages over other approaches like fluorescence spectra (Takahashi et al., 2000), antibody staining (Razmovski et al., 2000), Raman spectroscopy (Ivleva et al., 2005) or photoacoustic microscopy (Miyamoto and Hoshimiya, 2006). Firstly, the microscopical image recognition allows a very easy and very efficient quality control of the results, because it is based on the world-wide established standard, namely the visual pollen recognition, secondly, it can be easily extended to other airborne particles of interest (e. g., other pollen taxa in other countries, fungal spores or other particle types, as long as they are accessible by light microscopy) just by extending the software, and thirdly, a microscopical image recognition system can be built at affordable costs which is essential for the application within a big network.

1.2 State of the Art

1.2.1 Automated Pollen Recognition

The idea to automate the identification and counting of pollen grains and spores has been considered to be beyond any realization for many years. Nevertheless the automation has been proposed nearly 40 years ago by J. R. Flenley (Flenley, 1968). The complexity of the problem may be one of the reasons why only a few research groups world wide are working on this subject.

Massey University in New Zealand

The oldest and still active group with the most publications concerning pollen recognition, e. g., (Flenley, 1968; Langford et al., 1986, 1990; Stillman and Flenley, 1996; Li and Flenley, 1999; Zhang et al., 2004; Treloar et al., 2004; Li et al., 2004; Hodgson et al., 2005; Allen et al., 2006) is the group around John R. Flenley at the Massey University in New Zealand. They published in 1986 the first paper

1 Introduction

concerning automated recognition of fossil pollen in SEM recordings by an analysis of their surface texture (Langford et al., 1986). A comprehensive summary of the various studies by this group has been published 2004 in a series of articles "Towards automation of palynology" in the Journal of Quaternary Science (Treloar et al., 2004; Li et al., 2004; Zhang et al., 2004). The recent publications by this group incorporate the recognition also of fresh airborne pollen (Hodgson et al., 2005) (see also (Zhang et al., 2005) of Y. Zhang, a former member of this group). Furthermore they have built an automated low-cost microscope "AutoStage" with dark field illumination, dedicated to pollen recognition (Allen et al., 2006).

The tools developed by this group are mainly intended to be applied to the recognition of pollen in fossil samples which are prepared from a suspension. So the authors of (Allen et al., 2006) create microscopical slides with at most 500 pollen per slide to avoid clumping. This allows a simple segmentation by computing an edge image and the application of some morphological operators.

Currently (Allen et al., 2006) use a set of 43 features (Gabor, Wavelet and Haralick features for texture description, and geometric, histogram and Hu moment features for shape description), which have turned out to give the best performance during their earlier research.

For the classification, a multilayer perceptron is applied. All published results are based on 2D images of pure pollen samples containing manually collected pollen grains. In (Hodgson et al., 2005) the training was done with 64 pollen grains from 4 taxa. The test with 120 pollen grains yielded a recognition rate of 96%. In another experiment they obtained a recognition rate of 89% on a data set containing 3800 pollen grains (2850 training and 950 test objects) from 19 species (Allen et al., 2006). On a selected subset of this data set containing only 7 species, a recognition rate of 98% was reached.

University of Sydney, Australia

In the late 90'th a big commercially oriented Australian project aiming at the automation of pollen recognition has been undertaken during 2 years with 24 involved persons. Due to patent issues, the researchers were not allowed to publish their results. The only available information is contained in abstracts from conferences (Jones et al., 1999; Jones, 2000) and some personal communication by Allen Jones. The group has used a database containing 16,220 transmitted light microscopic images (2D) of individual pollen grains from 80 plant species, recorded from manually prepared pure pollen slides. A set of 20 morphological features, 70 Fourier descriptors and 800 Haralick texture features were extracted. A subset of 200 of these features has been used to setup a series of binary tree classifiers. By splitting the data set into 2/3 training and 1/3 test objects, mean recognition rates of 76.3% on genus level (46 taxa) and 70.7% on species level (all 80 taxa) have been obtained.

Bangor University , UK

Ian France worked on the recognition of fossil pollen during his PhD between 1994 and 2000 (France et al., 1997, 2000). He has been involved in the set up of the only publicly available database of pollen images, the "Bangor/Aberystwyth Pollen Image Database" (Duller et al., 1999a,b). The published results (France et al., 2000) incorporate three of the available 10 species in that database. For segmentation the authors use a standard threshold on the shading corrected image. The feature extraction and classification is done with a neural network called "Paradise Network", which includes a feature extraction layer, a pattern detection layer and a classification layer. The feature extraction layer mainly computes horizontal and vertical gradients. The pattern detection layer classifies small patches at certain positions. These positions are varied within a certain range, to become robust to small deformations. Finally the classification layer assigns the objects to clusters, that were found during training with an unsupervised clustering. These clusters were manually assigned to one of the given pollen classes. The reported mean recognition rate on the dataset with 204 pollen from 3 taxa was 81.7% with

a mean precision of 96.7%. Nothing is communicated about the validation scheme used, so maybe the identical data sets have been used for training and testing.

INRIA, Sophia-Antipolis in France

The third big group (in chronological order), who worked intensively on the automation of pollen recognition of fresh airborne pollen is the group around Monique Thonnat at INRIA, Sophia-Antipolis in France. They mainly tried to transfer a human pollen classification scheme (based on the number of porates or colporates, thickness of the exine, and the like) one-to-one to the computer. Additionally some global measures like the mean color of the object and several features deduced from the shape of the pollen have been used. The first investigations deal with pollen recognition in 2D images (Mazière, 1997) recorded with transmitted light microscopy. Within an European project A.S.T.H.M.A. (“Advanced System of Teledetection for Healthcare Management of Asthma”) the authors transferred this approach to 3D image stacks (Bonton et al., 2002; Boucher et al., 2002; Boucher and Thonnat, 2002) from real air samples. All these studies have been done on one data set containing 350 pollen grains from 30 taxa. The best reported recognition rates are 77% in a leave-one-out test (Boucher et al., 2002).

University of Vigo in Spain

A further project on “Automatic classification and count of pollen from digital optical microscope images” was founded by the Vigo University in Spain yielding first publications in 2004 (Rodríguez-Damian et al., 2004; Sa-Otero et al., 2004). The group around Eva Cernadas García concentrates their work on 3 species of the Urticaceae-family. The used database contains scanned analog photographs of about 100 pollen grains of each species from manually prepared pure pollen samples (Rodríguez-Damian et al., 2006).

For the detection of the pollen grains, they use the standard circular hough transform on a pre-processed Sobel edge image. For the final contour extraction they tested a morphological approach, a snake-contour and a convex-hull approach.

The feature extraction is done by classical textural and shape features: Geometrical features from the binary mask, central and Hu Moments on the contour and on the gray values, Fourier descriptors of the contour and Haralick texture features as well as gray level run length statistics.

For classification they tested a minimum distance classifier with feature selection, a multi-layer-perceptron and a support vector machine with the RBF (radial basis function) kernel. The best reported result using a leave-one-out validation was a recognition rate of 89%.

Delft University of Technology, Netherlands

In 2006, the first results of a group around R. P. W. Duin and E. A. Hendriks were published (Chen et al., 2006). They use 3D transmitted light microscopic stacks with a low resolution in the third direction (8 planes, recorded at $8\mu\text{m}$ spacing). Their data set contains 254 pollen grains from 3 allergenic pollen taxa (birch (*Betula*^(A.4)), grass (*Poaceae*^(A.6)) and mugwort (*Artemisia*^(A.1))) which had been recorded from purified samples of stained pollen.

For segmentation and feature extraction, standard techniques have been applied: A threshold segmentation, 19 shape features from the binary mask and 14 basic gray level features like the distance to the mean histogram of each taxon. Aside from these features, the authors developed special porate and colporate detectors. They use a circular hough transform to find the porate of a grass pollen, a cross correlation based template matching to find the porate of a birch pollen, and a measure for the regularity of the gray values along the cell wall to detect the colporates of mugwort pollen. These suggested porate and colporate detectors work only with pollen in polar view.

A so-called “linear normal classifier” has been applied yielding a recognition rate of 97.2% by use of a 5-fold cross-validation. Without the special detectors, the recognition rate drops to 94.9%.

Caltech and New York University, US

The only group so far that has published recognition results on real-world air samples with a reasonable number of pollen (3686 pollen grains belonging to 27 species), is the pollen group of the California Institute of Technology in a cooperation with the New York University (Ranzato et al., 2007).

Their data set consists of air samples that have been collected with the standard Burkard spore trap and have been recorded by use of 2D transmitted light microscopy. Due to the great amount of other dust particles and the partially imperfect preparation with lots of bubbles, a simple threshold segmentation is not possible here.

Instead the DOG (“difference of Gaussian”) salient point detector is applied to detect possible centers of pollen grains. From these centers the greatest gradients are searched on rays in 8 directions which allows a robust fit of a circle to the found locations. This circle is used to define the bounding box. As a next step bubbles are identified by their high mean brightness and rejected using a threshold. Particles with overlapping bounding boxes are also rejected. With this approach about 83% of the pollen grains are correctly detected.

The feature extraction is prepared by smoothing with 4 differently sized Gaussians and by extraction of 9 rotational invariant features computed at each of the pixels from the partial derivatives (like gradient magnitude, Laplacian, etc.). After a nonlinear mapping of the resulting values, the final features are computed by an average over all the single values in the image. The combination of three different mapping rules have been identified to provide good results. According to the first two mapping rules values above, respectively below the background gray value are eliminated, and according to the third rule the gray values below the background gray value are inverted. The resulting feature vector has $4 \times 9 \times 3 = 108$ components.

For classification the 108-dimensional feature vector is projected to a 14-dimensional subspace found by a Fisher’s Linear Discriminants Analysis. Then each class is modeled by a mixture of Gaussians. The final classification is performed with a Bayes classifier.

Several classification results are listed in (Ranzato et al., 2007): On a data set with manually segmented pollen (denoted as “expert patches”) from the 8 most frequent pollen types (3104 pollen grains) without dust particles a recognition rate of 78,2% is reached. By the usage of the “good patches” out of the results of the automatic segmentation of these 8 pollen types together with dust particles (1:10) a recognition rate of 64,9% has been obtained. Finally the authors reduced this data set to the three most frequent pollen types that are found in January (alder (*Alnus*^(A.2)), ash (*Fraxinus*^(B.4)) and pine (*Pinus*^(D.10))) which increased the recognition rate to 83%.

For a pollen recognition system that operates on real air samples, a further statistical measure of the performance is important, namely the precision. It is not reported in the article, but it can be computed from the provided confusion tables. Assuming that the relative numbers of “good patches” per species are proportional to the total numbers per species, we obtain a mean precision of about 30% for the experiment with 8 pollen types and about 54% for the experiment with 3 pollen types.

Others

Apart from the big efforts on this subject which are listed above, a couple of approaches have been published only as a kind of “notice of intend” with no or very limited results. E. g., Albuquerque Araujo et al. (2001a,b) report the successful discrimination of two single pollen grains. Other studies are dedicated to very particular questions, e. g., for the probability of oil deposits in sediment samples (Ujii, 2001). In such cases the average brightness of the pollen grains seems to deliver sufficient information. In several other papers, the pollen were only used as a test of particular techniques, e. g., Dahme et al. (2006) used the pollen textures from a database of neotropical pollen (Bush and Weng, 2007) to test their texture recognition algorithms.

Discussion

20 years after the first attempts on the automation of pollen recognition have been published, many groups have ceased their work on this subject after several years of research. The most difficult step from the clean world of manually prepared pure pollen samples to the dirty world of real air samples was performed only by the INRIA group and by the Caltech group (see table 1.3 and table 1.4 for a condensed overview)

Table 1.3: Condensed Overview over other approaches to pollen recognition (**Pure Pollen Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate
Massey University New Zealand								
(Hodgson et al., 2005)	pure	2D	morph.	texture, shape	neural net.	184	4	96%
(Allen et al., 2006)	"	"	"	"	"	1,400	7	98%
"	"	"	"	"	"	3,800	19	89%
University of Sydney Australia								
(Jones, 2000)	pure	2D	?	texture, shape	binary trees	16,220	46	76.3%
							80	70.7%
University of Vigo Spain								
(Rodríguez-Damian et al., 2006)	pure	2D	morph, snake	texture, shape	neural net, SVM	100	3	89%
Delft University of Technology Netherlands								
(Chen et al., 2006)	pure	3D	thresh.	texture, shape, specialized	"linear normal classifier"	254	3	97.2%

pure: pure pollen samples
SVM: Support Vector Machine

Table 1.4: Condensed Overview over other approaches to pollen recognition (**Real-world Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate	precision
Bangor University UK									
(France et al., 2000)	pollen +debris	2D	thresh.	neural net.	neural net.	204	3	81.7%	96.7%
INRIA, Sophia-Antipolis France									
(Boucher et al., 2002)	air	3D color	morph. +color	shape, color, specialized	Mahalanobis	350	30	77%	?
Caltech and New York University, US									
(Ranzato et al., 2007)	air	2D	DOG + circle fit	texture	Fisher linear discr. + GMM	3104	8	64,9%	30%

DOG: Difference of Gaussians
GMM: Gaussian mixture model

A second important step (in our opinion) is to use the full 3D information instead of just a 2D image of the middle layer of the pollen. One of the shortcomings of 2D approaches is the appearance of defocused particles when two differently sized particles lie side-by-side on the sample. Even more important is the problem that even pollen experts are not able to make an unequivocal classification of the pollen taxon, just based on the plain 2D information. Only the INRIA group and the Delft group

have recorded 3D data with the reserve that the number of planes of the Delft group might be still too low for a human expert.

One critical point in all the presented studies is the fact that training and test pollen are taken from the same microscopical slide. This is particularly critical with datasets that consist of one slide per pollen species. Features, that are sensitive to the mean gray value, to signal-to-noise ratio and to similar image properties may be biased by the characteristics of the preparation of the slide or the recording parameters. Such critical preparation and recording characteristics are, e. g., the age of the pollen grains, the consistency of the embedding medium, the current degree of degradation of the microscopic lamp, the current adjustment of aperture stop or of the Köhler illumination, etc. A highly-sophisticated classifier that operates on such features may then use the slide characteristics instead of the desired characteristics of the pollen grains to make its decision.

Also for mixed pollen slides and real-world air samples, the recording of training and test pollen from the same slide can bias the results to higher recognition rates. In such a setup, e. g., the mean gray value of a particle is a very discriminative feature – which is usually not the case on separately prepared and recorded slides.

From the image recognition point of view one can summarize that all the approaches extract only global statistics of the gray value distributions within the pollen grains apart from the shape features, and that each of the approaches contains at least a few features that are sensitive to the mean gray value (if not the mean gray value itself is used as a feature). Two groups (the INRIA group and the Delft group) have developed specialized detectors for certain structures in certain pollen taxa at certain orientations.

According to our experience the “global statistics” work quite well on pure pollen data sets. On real air samples with very high numbers of non-pollen particles, always lots of other particles appear that have very similar global statistics as the pollen. The very high number of false positives in the Caltech study seems to confirm this experience.

1.2.2 The Haar-Integration Framework

As mentioned previously, the Haar-Integration (HI) framework, introduced by [Schulz-Mirbach \(1995b\)](#) offers a simple but very powerful way of a general feature extraction. The method has initially been described for two-dimensional image data ([Schulz-Mirbach, 1995a](#); [Burkhardt and Siggelkow, 2001](#)), but it has been straightforward to extend the HI framework to three-dimensional volumetric data ([Schael, 1996, 1998](#); [Schael and Siggelkow., 2000](#)). Even though the computational complexity of the direct computation is linear in the number of pixels or voxels of the image, the pre-factor is usually so high that suitable algorithms for a faster computation are needed. For this purpose [Siggelkow and Schael \(1999\)](#) have proposed a method that is based on a Monte-Carlo integration and that allows a fast estimation of these invariants. Other extensions of the HI framework include the construction of feature histograms for content-based image retrieval ([Siggelkow, 2002](#)), the use of relational kernels for texture defect detection ([Schael, 2004](#)), the integration of transformation knowledge into the kernel functions ([Haasdonk, 2005](#)) or their combination with salient point detectors for the extraction of local features ([Halawani, 2006](#)).

1.3 Main Contributions

In this thesis the development of robust 3D invariants for the recognition of biological structures in volumetric data sets is described. The chosen application is the recognition of pollen grains continuously sampled from the ambient air for a fully automated online pollen monitoring. Together with partners from the German Weather Service, from Fraunhofer institutes and from industry the first prototype of such a fully automated pollenmonitor has been built. The results obtained during a routine-like

operation of this machine in 2006 are presented in this thesis.

The most important contribution is the development of algorithms that have proven to work not only on small data sets recorded in a well-controlled laboratory environment, but on a very large data set recorded in a real-world routine environment.

A pollen recognition system can never be reduced to the plain segmentation and image recognition part, in other words, the comprehension of the steps taken to provide the initial data have to be taken into account as well, i.e., the collection, the preparation, the microscopy, the recording and the pre-processing. The aspects of representative sampling of the pollen grains and of the preparation of the slides are out of scope in a thesis in computer science. However, the influences of the image acquisition to the further image recognition are so evident, that a pattern recognition task cannot be solved without controlling every aspect of the image acquisition system (see figure 1.1)

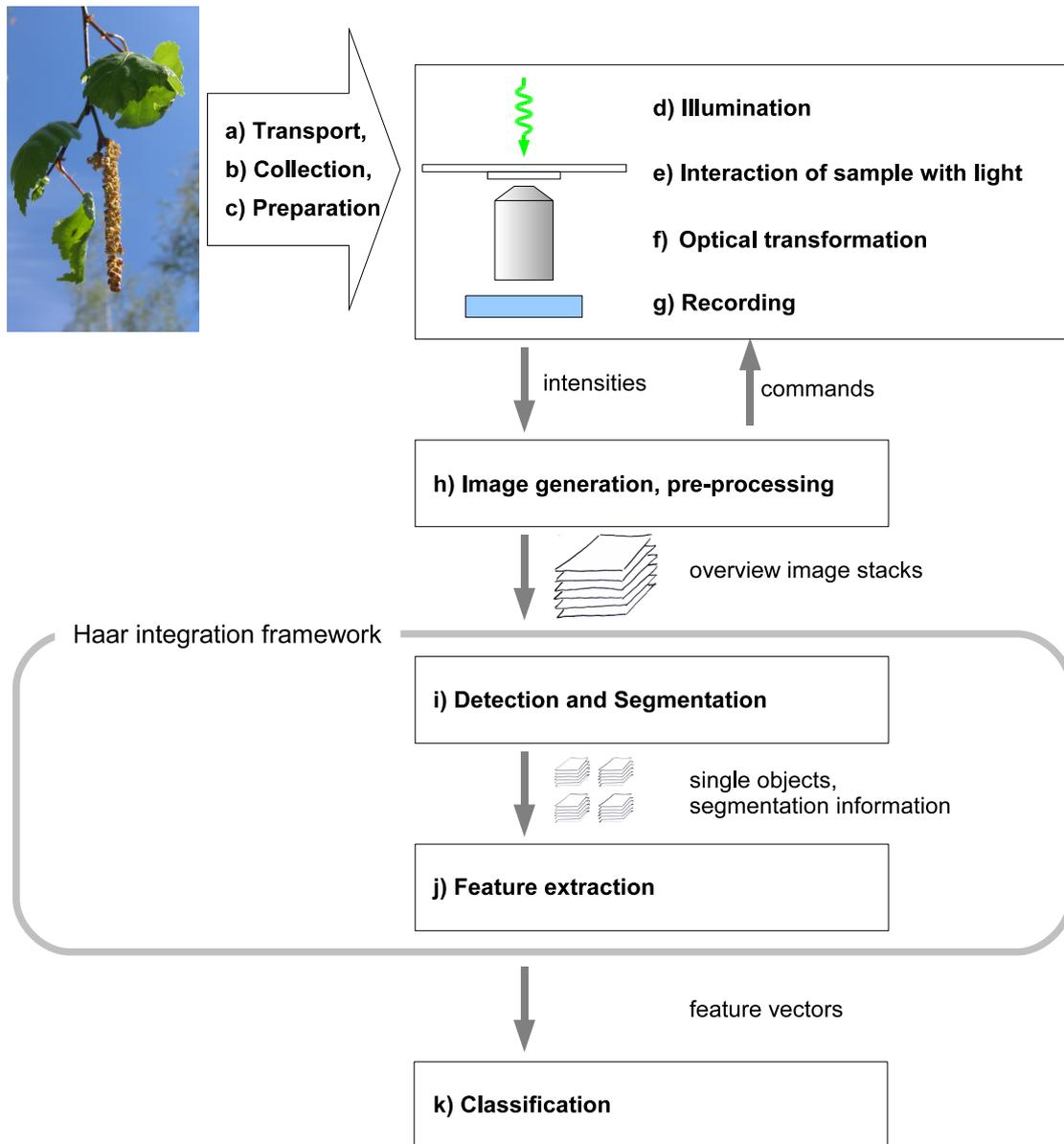


Figure 1.1: Pollen recognition scheme

According to this scheme the main contributions can be grouped into four categories, the pre-processing, the detection and segmentation, the feature extraction and the classification.

Image Generation and Pre-Processing

The work that has directly been associated with the image acquisition system includes

- Experiments in order to decide, whether the deconvolution of stacks recorded by a conventional fluorescence microscope can deliver appropriate image quality for pollen recognition together with Thorsten Schmidt ([Schmidt, 2003](#))²
- Improved algorithms for the reconstruction of images that have been recorded with structured light techniques together with Anselm Vossen ([Vossen, 2004](#))³ and the application of structured light techniques to the recognition of pollen in honey samples together with Raphael Baumgartner ([Baumgartner, 2005](#))⁴
- Computation of high dynamic range images from multiple fluorescence recordings with different shutter times and dynamic bleaching-correction.

Detection and Segmentation

Major efforts have been done to establish a fast, robust and accurate detection and segmentation of biological objects in the recorded 3D stacks. This includes

- A fluorescence-based detection and coarse segmentation of arbitrarily shaped objects with a subsequent refinement based on the transmitted light channel using morphological techniques
- The development of vectorial voxel-wise invariants together with Janina Schulz and their application to cells in the Arabidopsis root ([Schulz, 2005](#); [Schulz et al., 2006](#))⁵
- A detection of spherical pollen in the transmitted light channel based on vectorial voxel-wise-invariants and the subsequent refinement using a modified canny edge detector, gradient vector flow and a fast FFT-based snake algorithm together with Qing Wang ([Ronneberger et al., 2007](#)).
- A full 3D segmentation of confocally scanned pollen based on a graph cut algorithm of [Boykov and Kolmogorov \(2004\)](#)

Feature Extraction

The developed 3D invariants are based on the Haar-Integration framework of [Schulz-Mirbach \(1995b\)](#).

- Meeting the requirements that arise by processing real biological objects recorded in 3D with a confocal laser scanning microscope. Primarily this is the introduction of a multi-scale approach together with the fast algorithm for the computation of two-point kernels based on the FFT ([Ronneberger et al., 2000, 2002a,b](#))
- The definition of the information that can be sensed by a two-point kernel and theoretical considerations how this information can be used for an automatic selection of kernel functions. A first feasibility study on an automated selection of kernel functions has been done together with Thorsten Schmidt ([Schmidt, 2004](#))⁶

²The student project of [Schmidt \(2003\)](#) was initiated and supervised by the author

³The student project of [Vossen \(2004\)](#) was initiated and supervised by the author

⁴The master thesis of [Baumgartner \(2005\)](#) was initiated and supervised by the author

⁵The master thesis of [Schulz \(2005\)](#) was initiated and supervised by the author

⁶The master thesis of [Schmidt \(2004\)](#) was initiated and supervised by the author

- The extension to voxel-wise invariants and a fast algorithm to compute 3-point kernels by means of a Spherical Harmonics expansion (Ronneberger et al., 2005a). This algorithms have been applied to the simultaneous segmentation and recognition of cells in confocal recordings together with Janis Fehr (Fehr, 2004; Fehr et al., 2005)⁷
- The introduction of a deformation model, synthetic channels and parameterized kernel functions to compute great multitudes of deformation-robust invariants with high localization in the object space that are still sensitive to fine-structural changes within the objects (Ronneberger et al., 2007)
- The elimination of the dependencies of the invariants on the absolute gray value and high robustness to gray scale transformations, including even nonlinear transformations.

Classification

For the classification of the objects an own highly customizable and extremely optimized support vector machine library – the libsvm (support vector machine template library), based on the publicly available libsvm from Chi-Jen Lin (Chang and Lin, 2001) has been developed. The optimization includes the handling of very high-dimensional feature vectors (e. g. 87,304 dimensions in (Ronneberger et al., 2007)), the full caching of the kernel matrix, and a highly optimized cross-validation, as well as a leave-one-out validation and grid search capabilities for finding the best training and kernel parameters.

1.4 Structure

The structure of this thesis is mainly oriented at the steps that influence the pollen recognition from the microscopical recording until the final classification result.

- **Microscopic Imaging:** In chapter 2 the required theory of the microscopic imaging is outlined. This includes the illumination, the interaction of the sample with the light, the optical transformation of the light and the recording (steps d-g in figure 1.1). Additionally a part of the applied 3D image generation and the post-processing techniques (step h in figure 1.1) are explained.
- **Data Sets:** In chapter 3 the two data sets used within this thesis are presented. The first data set is a typical laboratory data set containing high quality pollen grains recorded with a confocal laser scanning microscope. The second data set is a typical real-world data set containing a huge amount of airborne particles including pollen grains, that have been automatically collected, prepared and recorded with the first prototype of a pollenmonitor in 2006.
- **Theoretical Foundations:** In the chapters 4, 5, 6 and 7 the theoretical foundations for the 3D invariants are described. This includes
 - A short introduction to the Haar integration framework in chapter 4.
 - The analysis of the properties of the HI framework in chapter 5. E. g., the correct sampling of the transformation group and the analysis of the maximal information that can be sensed with two-point-kernels.
 - Description of the developed extensions to the HI framework in chapter 6. This includes the analysis of existing techniques and the presentation of new techniques to ensure the finiteness of the integral, different techniques in order to integrate prior knowledge into the HI framework, the introduction of voxel-wise invariants for simultaneous segmentation and recognition as well as vectorial voxel-wise invariants for the detection of spherical structures.

⁷The master thesis of Fehr (2004) was initiated and supervised by the author

- Several algorithms for a fast computation of the invariants are presented in chapter 7. This includes algorithms that make use of fast FFT-based convolutions and that use a fully rotation invariant approximation of the kernel functions by truncated Fourier- or Spherical Harmonics series expansions. For the vectorial voxel-wise invariants a fast simultaneous computation at multiple positions is presented that is based on an “inversion” of the kernel function.
- **Detection and Segmentation:** In chapter 8 the segmentation techniques are described that were applied to the pollen samples. This includes standard morphological techniques used within the first feasibility studies. The final segmentation is based on a detector using vectorial voxel-wise invariants and a subsequent application of snakes to find the contour. For the confocal recordings a graph-cut based segmentation is used.
- **Specialized Invariants for Microscopical Spherical Particles (MiSP):** In chapter 9 the derivation of the finally applied 3D invariants is described. The chapter starts with the definition of the required invariance properties based on a hypothetical elastic registration between the test and the training sample. In the following it is shown, how all aspects of such a registration can be modelled by an invariant approach. Finally a fast computation scheme for this class of invariants is outlined and a brief interpretation for a more intuitive understanding of the properties of these invariants is provided.
- **Experimental Results:** In chapter 10 and the appendices A, B, C and D the experiments and the obtained results are described. After a short analysis of the relevant statistical measures the results on the “confocal data set” and on the “pollenmonitor data set” are given and compared to the state of the art. If we only provide statistical results obtained with the new techniques on a former unknown data set, a reader has no chance to form an own opinion on the capabilities and shortcomings of these techniques. In order to provide the needed information extensive appendices have been created, containing automatically selected representative examples of the true-positive, the rejected, the false-negative and the false-positive pollen grains for each pollen taxon. At the end of chapter 10 several further results are summarized that have not been integrated into this thesis due to time and space limitations.
- **Summary, Conclusion and Outlook:** In chapter 11 we summarize the obtained practical and theoretical results and draw the conclusions. In the summarization several intermediate results are outlined that have led to important conclusions during our research on automated pollen recognition. Finally an outlook to the next challenges in this research area is provided together with some ideas how the obtained results may contribute to their solution.

1.5 Notation

The used notation is similar to that used in physics. E. g., in most cases, fields are represented by capital letters, while single values or vectors are represented by lower case letters.

x	single (possibly complex) scalar value
$\mathbf{x} = (x_1, x_2, \dots, x_n)$	n-dimensional vector
X	n-dimensional continuous scalar field
\mathbf{X}	n-dimensional continuous vector field
$X(\mathbf{x})$	scalar value (e. g., a gray value) at position \mathbf{x} of scalar field X
$\mathbf{X}(\mathbf{x})$	vectorial value (e. g., a gradient) at position \mathbf{x} of vector field \mathbf{X}
$\mathcal{F}\{\cdot\}$	Fourier transform
$\hat{A} := \mathcal{F}\{A\}$	Fourier transformed scalar field
$\hat{A}(\mathbf{k})$	Fourier coefficient at position \mathbf{k}
$C = A \circ B \Leftrightarrow C(\mathbf{x}) = A(\mathbf{x}) \circ B(\mathbf{x})$	element-wise binary operations of two scalar fields, where $\circ \in \{+, -, \cdot, /\}$
$\delta(\mathbf{x})$	Dirac delta function
\mathbf{R}_φ	rotation matrix
x^*	complex conjugate
$\Re(x)$	real part of a complex value
$\Im(x)$	imaginary part of a complex value
$\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right)$	del operator, represented by the nabla symbol
$\nabla_{\mathbf{x}'} = \left(\frac{\partial}{\partial x'_1}, \frac{\partial}{\partial x'_2}, \frac{\partial}{\partial x'_3} \right)$	del operator with respect to \mathbf{x}'
∇X	vector field containing gradients of scalar field X
$\nabla X(\mathbf{x})$	gradient at position \mathbf{x}

2 Microscopic Imaging

To interpret a gray value in a microscopic image and its relation to the real microscopic object, it is important to understand the main four steps in a microscope equipped with a camera

1. The generation and optical transformation of the illuminating light (e. g., directional, spectral and spatial distribution)
2. The interaction of the light with the object and its surrounding (e. g., absorption, reflection, refraction, diffraction, fluorescence)
3. The optical transformation of the emitted light (e. g., magnification, spatial filtering, filtering of wavelengths, etc.)
4. The recording of the light at certain positions, which is usually the “conversion” of photons to electrons, the transport of the electrons to the AD-converter and the AD-conversion of the resulting voltages.

The effects of all these steps have to be taken into account when trying to recognize an object on the basis of the recorded data. In the following sections the possible variations between images taken from the same object are described.

2.1 Optics

2.1.1 Illumination

There are lots of possible variations in the illumination part of the microscope. Most of them can and should be fixed by proper adjustment of the microscope, so that the intensity, as well as the spectral and spatial distribution of light is the same when recording the reference objects and the test objects. If it is not possible to ensure a constant light intensity, at least the ratio between the different wavelengths (or colors) of the light emitted by the light source should not change, otherwise the gray image of multicolored objects will change unpredictably. On the other hand, if we can control the spectrum in a reproducible way (e. g. with different filters) we can utilize this to sense different properties of the objects by taking multiple images with different settings.

Another effect which influences the integrated light intensity is the 50Hz oscillation of the intensity when using a lamp running on AC (e. g. HBO-50 fluorescence lamp from Zeiss). The only way to get reproducible images here is to adjust the exposure time to multiples of 20ms, so that the integrated light intensity does not depend on the exposure start time.

Even though we use the Köhler illumination, the distribution of the light intensity is not perfectly uniform. This shading should be corrected in the resulting image to ensure that the objects gray values do not change when the object moves from the middle of the visible region to the border.

Most microscopes have also the aperture stop integrated into the illumination part. Thus a change of the aperture will change the whole three dimensional spatial light distribution and therefore will change the appearance of the objects. As we are not able to compensate those effects at acceptable time and costs when using transmitted light microscopy (see next section), we have to ensure that the reference and test objects are recorded with the same aperture. Here again we can utilize this effect to sense different properties of the objects provided we are able to control the aperture in a reproducible way.

2.1.2 Interaction of Light with Objects

In the interaction of light with the object we take three different classes of effects into account. The first is absorption which only changes the intensity of the light rays but not their direction. If we illuminate the object with white light and if the absorption of a region depends on the wavelength, we see the region at a certain color. The amount of the transmitted light is proportional to the intensity of the incident light, which means that the absorption factor of two consecutive regions is the product of the single absorptions factors (see figure 2.1)

Light absorption of particles that are far out of focus results in slightly darker regions in the image and can therefore be modeled as a variation of the illumination. This also has to be taken into account when a shading correction is performed.

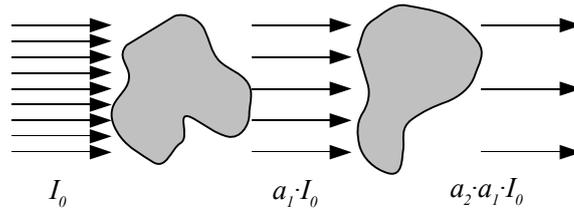


Figure 2.1: Absorption of light: the absorption factor of two consecutive objects is the product of the single absorption factors a_1, a_2

The second class of effects results from the variation of the speed of light, respectively, the refraction index within the light paths. This results in a refraction and reflection of the light rays at the gradients of the refraction index (see figure 2.2)

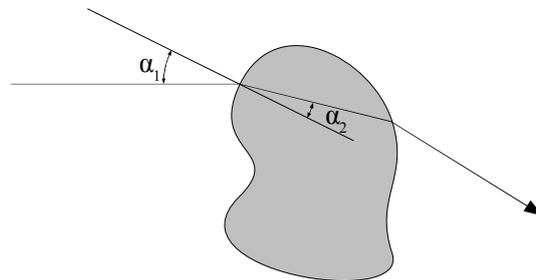


Figure 2.2: refraction due to different refraction indices of the object and the ambient medium

In figure 2.3 this effects can be observed especially at the three pores of the *Betula*^(A.4) pollen which are nearly identical in reality but produce very different gray images in the camera plane.

For a known 3D object these effects can be simulated by a ray tracing approach. However, the inverse problem (computing the 3D object from one or more images) is nearly impossible to solve at acceptable time and costs.

The third class of effects is the primary fluorescence of certain molecules within the object. The absorption of a photon raises a covalent bond of the molecule into an excited state. Via multiple vibrational sub-levels it falls back into its ground state and emits a lower-energy photon (at a larger wavelength) into a random direction. Usually only a very small fraction of the bonds within the molecule fulfill these conditions, which results in a very low intensity of the emitted light compared to the incident light. Furthermore there is a certain probability that the bond in the excited state breaks, which destroys the molecule and of course its fluorescence properties (“bleaching”). Every excited molecule can be considered as an omni-directional small light source (see figure 2.4). If the absorption and the

2 Microscopic Imaging

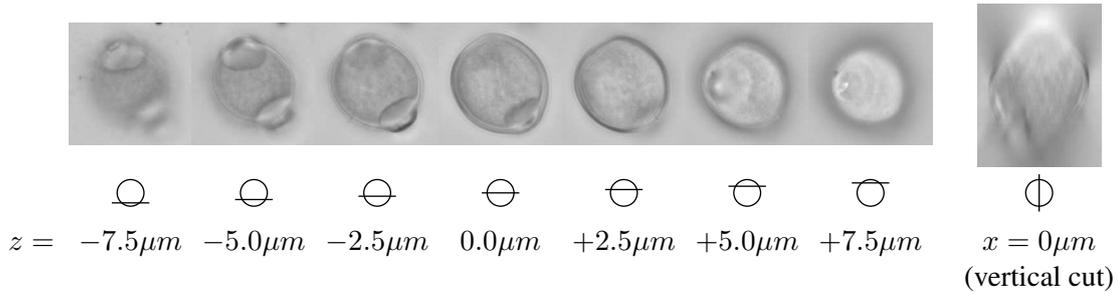


Figure 2.3: transmitted light microscopic images of a *Betula*^(A.4) pollen grain recorded with a 100x oil-immersion objective at different vertical positions. The rightmost image shows a vertical cut through the recorded 3D data set which was reconstructed from 256 single images.

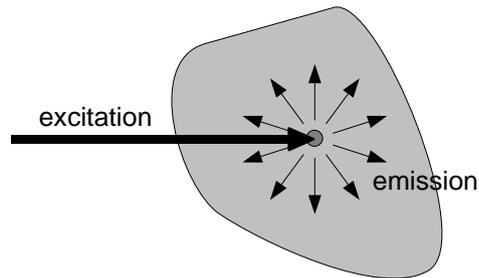


Figure 2.4: Primary fluorescence: each molecule can be considered as a small light source.

variance of the refraction indices in the object are negligible, the resulting image is just a sum of all these small lights, which simplifies the interpretation significantly. Figure 2.5 shows an example of a pollen grain recorded with fluorescence microscopy.

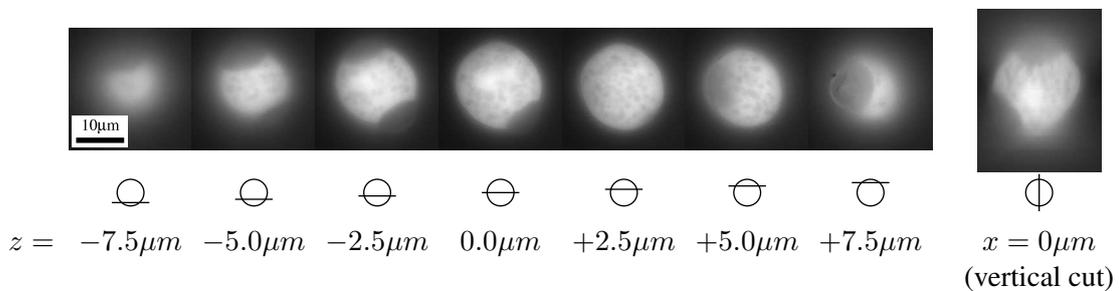


Figure 2.5: fluorescence microscopic images of a *Betula*^(A.4) pollen grain recorded with a 100x oil-objective at different vertical positions. The rightmost image shows a vertical cut through the recorded 3D data set which was reconstructed from 256 single images.

2.1.3 Optical Transformation of the Emitted Light

After the interaction between the light and the object, the emitted light field is optically transformed by the lenses and filters of the microscope (see figure 2.6).

The effects of the optical transformation of the emitted light will be modeled within three steps. A magnification step, a spectral filtering step and a spatial filtering step.

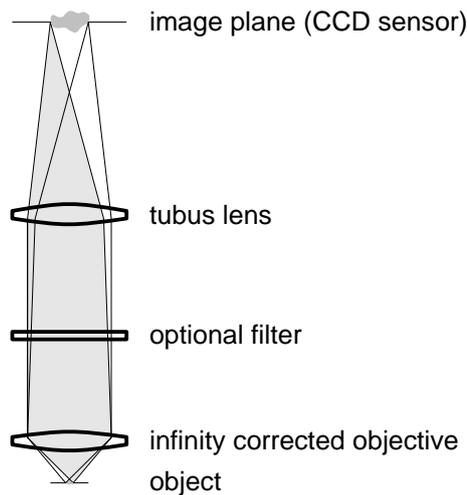


Figure 2.6: Microscopical optics for digital image recording

The effect of the magnification step is obvious: The light that originates from two points in the object plane with distance d_{obj} is projected onto two points in the image plane with distance d_{img} . The factor between these two distances is the magnification factor m , such that

$$d_{\text{img}} = m \cdot d_{\text{obj}} . \quad (2.1)$$

If the light is recorded by a CCD sensor (as depicted in figure 2.6) no ocular is used. In this case the magnification factor m is just the magnification factor that is written on the objective. E. g., if we use a 20x objective, the image of an object with $1\mu\text{m}$ diameter will have $20\mu\text{m}$ diameter in the image plane on the CCD sensor. And if we use a CCD sensor with a pixel size of $6.45\mu\text{m}$ the diameter of the object in this recorded image will be 3.1 pixel.

The spectral filtering step influences the intensity of the light depending on its wavelength (or color). This is necessary, e. g., for fluorescence microscopy where an appropriate long-pass filter guarantees that only the emitted fluorescence light reaches the image plane, but not any reflected excitation light.

While the magnification and the spectral filtering are the wanted effects of a microscope, the imaging process always introduces a spatial filtering of the image. The main cause of this spatial filtering is the wave nature of light, as well as the fact that the structural dimensions under consideration are in the order of the wavelength (the wavelength of visible light is between $0.4\mu\text{m}$ and $0.8\mu\text{m}$, the smallest considered pollen grains have a diameter of $10\mu\text{m}$)

The origin of the spatial filtering is explained by means of wave-optics. In the wave-optical model each point of the object is viewed as a source of a Huygens spherical wave. The waves that originate from a very small object (idealized as a spatial delta function) are shown in figure 2.7.

Only a part of the spherical wave that originates from the point source is captured by the objective. The half-angle of the light cone that enters the lens, α , is an important characteristic of an objective and is quantified by the numerical aperture

$$\text{NA} = n \sin \alpha , \quad (2.2)$$

wherein n is the refraction index of the immersion medium (1.0 for air up to 1.56 for oils). So in objectives with a high numerical aperture more of the emitted light is collected which results in brighter images. Especially for low intensities in fluorescence microscopy this is an important fact.

Another feature of an objective that depends on the numerical aperture is its ability to truly reconstruct the original light field in the image plane. This can be visualized by drawing Huygens waves

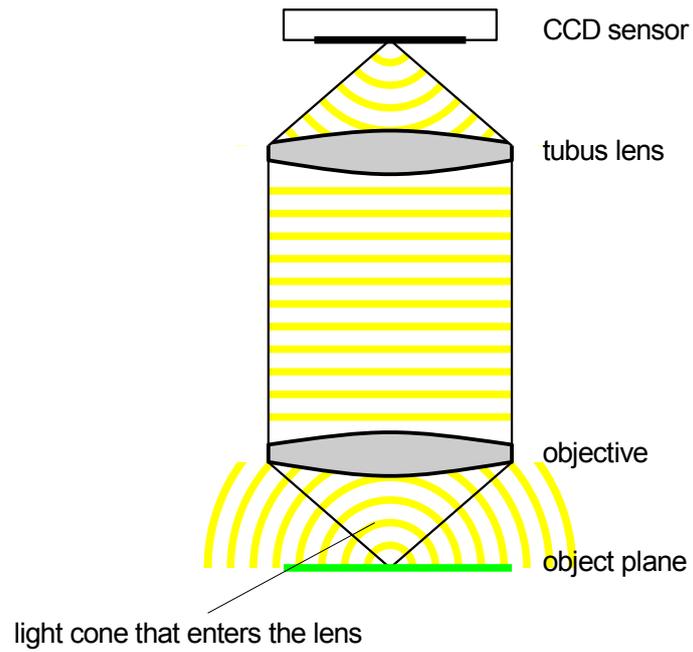


Figure 2.7: Spatial filtering effects in a microscopical setup. Only a part of the emitted light can be captured for the reconstruction of the object in the image plane.

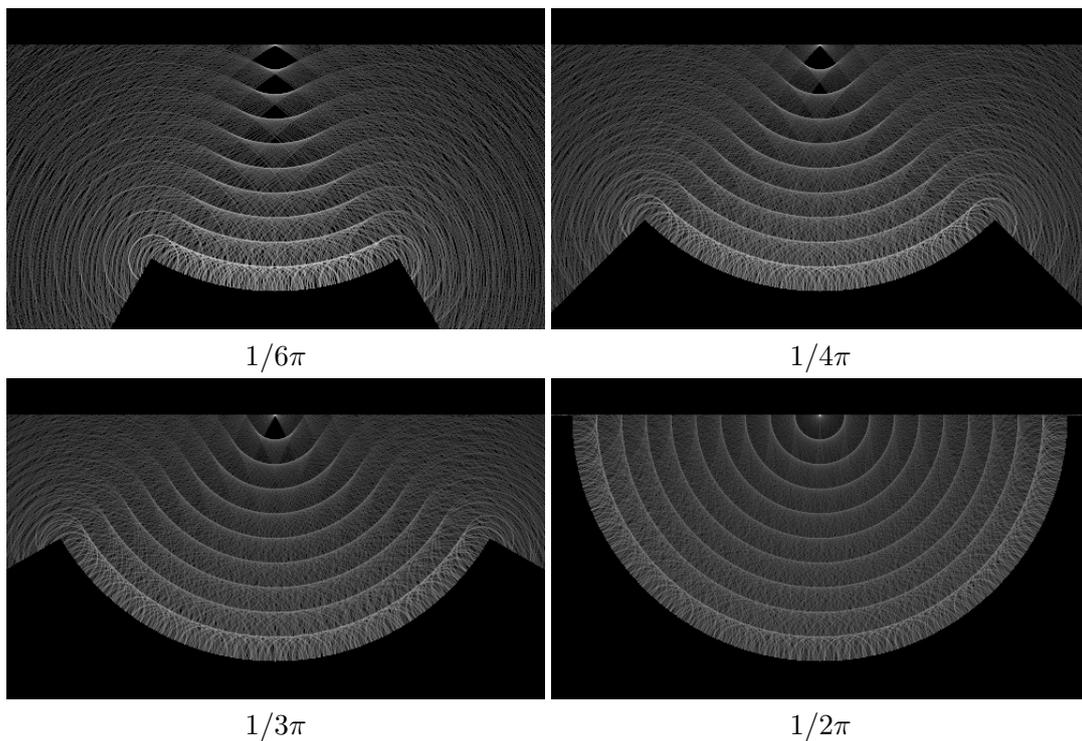


Figure 2.8: Reconstruction of the light field around the image plane for different apertures, depicted by Huygens spherical waves. The half angle of the entering light cone is written below each image.

(see figure 2.8). When the optical setup of figure 2.7) is used, the virtual origins of the Huygens waves are located on a sphere around the image of the original point source.

It is clearly seen in this illustration that the reconstruction of the spatial delta function in the image plane becomes sharper when the aperture is increased. In a first order approximation this effect can be modeled by the convolution of the ideal 3D intensity distribution I_{ideal} around the image plane with the so-called point-spread-function (PSF) H_{psf} of the optical system.

$$I_{\text{real}} = I_{\text{ideal}} * H_{\text{psf}} \quad (2.3)$$

2.2 Recording of the Light

For the correct interpretation of a digital image it is important to take the different effects into account that arise during the conversion of light intensities (or photon counts) into digital gray levels.

2.2.1 Charge Coupled Devices (CCD) Sensor

In a conventional microscope the light in the image plane is usually recorded with a charged coupled device “CCD” (see figure 2.9).

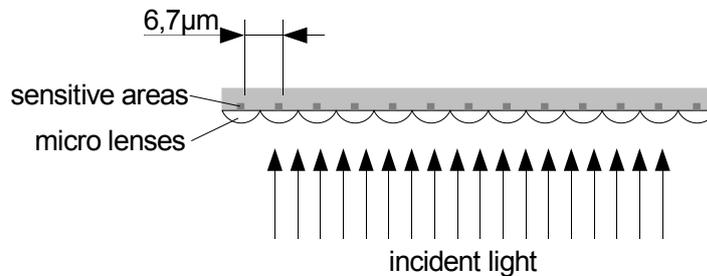


Figure 2.9: CCD sensor

On a CCD sensor the incident light is sampled with regularly spaced light sensitive areas. To fulfill the requirements due to the sampling theorem (and hence be able to reconstruct the original signal from the discrete samples) the user must ensure by optical low-pass-filtering that the intensity distribution does not contain spatial frequencies higher than the Nyquist critical frequency.

$$f_c = \frac{1}{2\Delta x} \quad (2.4)$$

with Δx = distance between to CCD elements. Many CCD sensors are equipped with micro lenses in front of each pixel to collect as much of the incident light as possible. A positive side effect is that these lenses already introduce a certain low-pass filtering that reduces the under-sampling-effects if the image contains too high spatial frequencies.

Besides these spatial sampling effects there are several further effects that influence the final digital gray level. The incident light can be best modeled by single photons here. These photons are “converted” to electrons, the resulting charge is read out and is converted to a digital gray value (see figure 2.10).

During this “conversion” the following effects have to be taken into account. The first is the quantum nature of light: The light intensity corresponds to a certain number of photons per second. The temporally and spatially occurrence of the photons is random. So if we measure a light intensity, this corresponds to the counting of photons within a certain area and a certain time interval. The counting

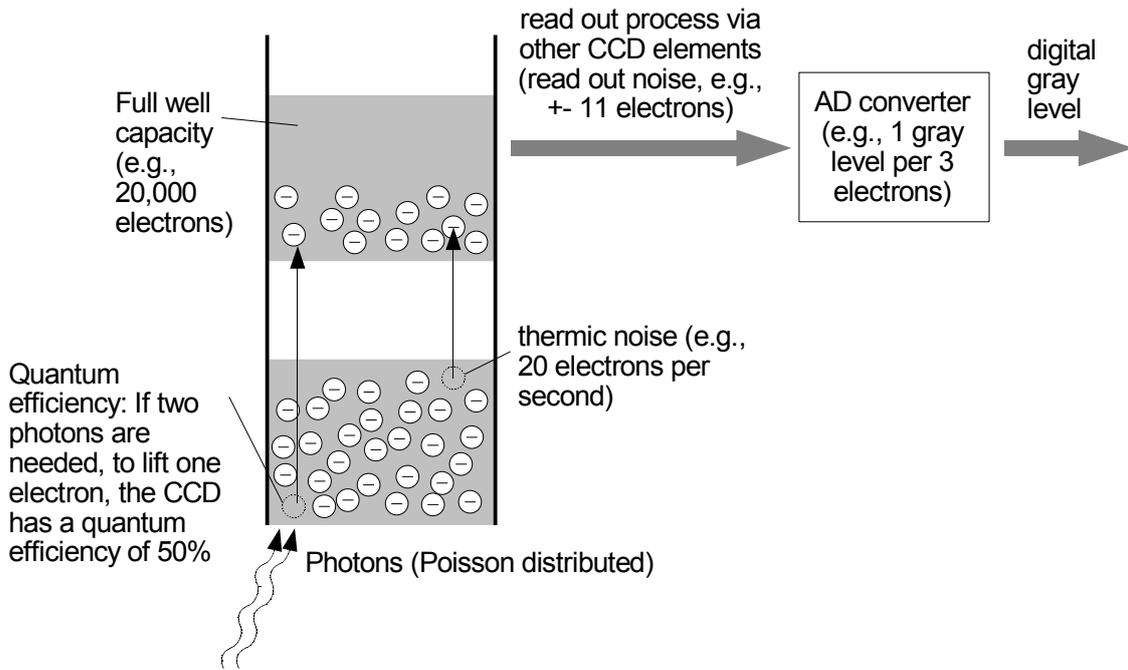


Figure 2.10: CCD element: Important effects during “conversion” of photons to digital gray levels

results are Poisson distributed in this case. The “Poisson noise”, i.e. the variance of the resulting countings is equal to the number of the counted photons N_{phot} . So the relative error is

$$\frac{\Delta N_{\text{phot}}}{N_{\text{phot}}} = \frac{\sqrt{N_{\text{phot}}}}{N_{\text{phot}}} = \frac{1}{\sqrt{N_{\text{phot}}}} \quad (2.5)$$

For low intensities, short illumination times or small CCD elements, this error is the dominant error source in the whole conversion pipe. E. g., a gray value that corresponds to 100 photons has an error of 10%. For high intensities, long illumination times or large CCD elements this error is usually negligible compared to the other errors within the process. E. g., for a gray value that corresponds to 20,000 photons this error is only 0.71%.

The first step in the CCD element is the “conversion” of the incoming photons to electrons by means of the photoelectric effect, i.e. by lifting electrons into a potential well. An important property of the CCD sensor is its quantum efficiency α_{qe} , namely the fraction of photons which successfully lift electrons. For the estimation of the error due to this step, the probability α_{qe} can be combined with the probability of the occurrence of a photon. So the number of photo-generated electrons N_e for a given light intensity is also Poisson distributed such that the quantum efficiency just increases the relative error by a certain factor

$$\frac{\Delta N_e}{N_e} = \frac{\sqrt{\alpha_{\text{qe}} \cdot N_{\text{phot}}}}{\alpha_{\text{qe}} \cdot N_{\text{phot}}} = \frac{1}{\sqrt{\alpha_{\text{qe}}}} \cdot \frac{\Delta N_{\text{phot}}}{N_{\text{phot}}} \quad (2.6)$$

E. g., a quantum efficiency of 0.5 will increase the relative error by a factor of $1/\sqrt{0.5} = \sqrt{2}$.

Another cause for lifting electrons into the potential well are the thermal fluctuations. The number of electrons due to this effect depends on the temperature of the CCD element and the exposure time, but is independent of the incident light intensity. Therefore the rate of thermally lifted electrons is called the dark current. To reduce the resulting “dark noise” the CCD sensor can be cooled.

An important property of a CCD element is its full well capacity which denotes the maximum number of photo-generated electrons that can be stored in the potential well. At higher exposures the sensor is saturated. In applications which allow to increase the light intensity (e. g. in transmitted

light microscopy) or allow to increase the exposure time, the full well capacity is usually the limiting factor for the overall signal-to-noise ratio.

After the illumination of the CCD elements, the collected charges in each of the potential wells are transported to the AD-converter. This read-out process introduces several error sources that are usually described by one combined read-out-noise. This read-out-noise is independent of the light intensity and the exposure time. It can be modeled by a Gaussian distribution. It is important to note that the read-out-noise produces also negative charges in contrast to the aforementioned error sources, i. e. with zero light intensity also negative gray values may occur.

The last error source in the whole pipe is the quantization error introduced by the conversion of the charges into digital values. This error is proportional to the quantization step that is chosen for the AD conversion. The number of quantization steps is usually somewhere between 256 (using 8 bit numbers) and 16,384 (14 bit numbers). The size of the quantization step can be controlled by the gain. Some cameras also allow to adjust the offset, which specifies the gray value for zero light intensity. If the image is recorded for a further analysis with image processing techniques, it is recommendable to set this offset to a value greater than zero. In this case, the “negative intensities” that may originate from the read-out-noise at very low light intensities, can be properly recorded, and the average over multiple pixels will return an unbiased result.

Another source of errors on CCD sensors are production errors which result in imperfect CCD elements. In most such cases we are confronted with “dead pixels” which always return a zero gray value independent of the light intensity, and “hot pixels” which always return the highest possible gray value. Such pixel defects are usually found in cheap CCD’s that are used in consumer cameras. Cameras that are designed for scientific applications usually contain selected CCD’s without any such defect.

2.3 Reduction of Out-Of-Focus Light

The main error source in microscopical images is the unwanted contribution of non-focused planes to the image of the focused plane due to the optical transfer function (OTF) of the optics.

To eliminate or at least reduce these effects, one can use either confocal microscopy which eliminates unwanted light by hardware components providing images with the highest possible quality at very high costs, though.

A fully software–implemented alternative are deconvolution techniques (e. g., Wiener filter) which remove the light dispersion by post-processing the digital images taken with a conventional fluorescence microscope.

A mixture of hardware and software solutions is a structured illumination with a consecutive post-processing of the digital images (e. g. with Zeiss “ApoTome”).

Instead of the efforts to avoid or to compensate the non-focus contributions in the images, one may appropriately adapt the feature extraction and classification to be robust against blurring due to the 3D point spread function.

2.3.1 Pure Hardware Approach: Confocal Microscopy

In conventional microscopy the whole sample volume is illuminated with uniform intensity. Out-of-focus points become only darker due to the point spread function of the detection optic.

In confocal microscopy (see figure 2.11) the sample is illuminated through a pinhole that is in the conjugated focal plane (“con-focal”) with respect to the detector pinhole.

The central effect is that object points other than the point of interest experience lower intensity illumination such that the light emitted from these points is less than in conventional microscopy. To record the whole volume, the sample must be scanned in all three dimensions with this focused point.

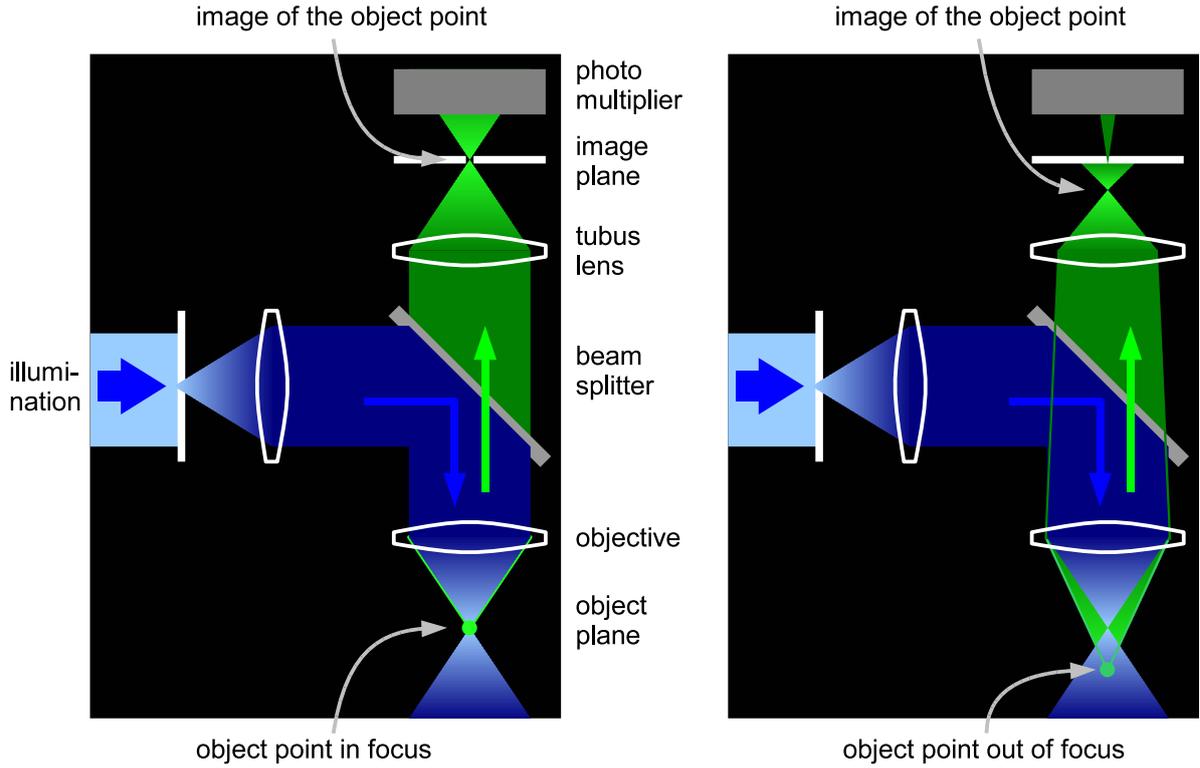


Figure 2.11: The confocal principle. The illumination is limited by a pinhole located in the conjugated focal plane of the detector pinhole. Object points outside the focal plane are illuminated with less intensity.

Quite similar to the conventional microscopy the confocal microscopy can also be described by Fourier optics. The intensity distribution of the illumination light I_{illum} is modeled by the convolution of the ideal image of the illumination pin-hole with the point-spread function,

$$I_{\text{illum}} = \delta * H_{\text{psf}} = H_{\text{psf}} . \quad (2.7)$$

The light I_{obj} emitted from the respective object point is proportional to the local density of the object D_{obj} and the local illumination intensity, i. e.

$$I_{\text{obj}} = D_{\text{obj}} \cdot I_{\text{illum}} = D_{\text{obj}} \cdot H_{\text{psf}} . \quad (2.8)$$

The intensity distribution I_{image} around the image of the focused point is modeled by the convolution of the ideal image of the object points with the point-spread function,

$$I_{\text{image}} = I_{\text{obj}} * H_{\text{psf}} = (I_{\text{ideal}} \cdot H_{\text{psf}}) * H_{\text{psf}} . \quad (2.9)$$

Finally the resulting signal must pass through the detector pinhole which is located in the coordinate origin.

$$I_{\text{image}}(\mathbf{0}) = \int_{\mathbb{R}^3} I_{\text{ideal}}(\mathbf{x}) \cdot H_{\text{psf}}(\mathbf{x}) \cdot H_{\text{psf}}(\mathbf{x}) d\mathbf{x} . \quad (2.10)$$

To obtain the full 3D data set the object is moved and the resulting intensity at the detection pinhole

$I_{\text{scan}}(\mathbf{t})$ is recorded as a function of the location \mathbf{t} of the focal point within the scanned object:

$$\begin{aligned} I_{\text{scan}}(\mathbf{t}) &= \int_{\mathbb{R}^3} I_{\text{ideal}}(\mathbf{x} + \mathbf{t}) \cdot H_{\text{psf}}(\mathbf{x}) \cdot H_{\text{psf}}(\mathbf{x}) d\mathbf{x} \\ &= \left(I_{\text{ideal}} * (H_{\text{psf}} \cdot H_{\text{psf}}) \right) (\mathbf{t}) . \end{aligned} \quad (2.11)$$

So the ideal light distribution is convolved with the squared point spread function. The decay of the PSF in axial direction is approximately $\propto 1/r^2$. So in conventional microscopy the contribution of out-of-focus signals decay $\propto 1/r^2$. In confocal microscopy this contribution decays much faster namely $\propto 1/r^4$.

The most common variant of such a confocal microscope is a confocal laser scanning microscope (abbreviate with LSM in the following). Here the pinhole and the lens in the illumination part are replaced with a laser that already emits parallel light.

2.3.2 Pure Software Approach: Deconvolution

A pure software based approach to the reconstruction of the true intensity distribution is the deconvolution. As the obtained intensity distribution can be modeled by the convolution of the true intensity distribution with the point spread function of the optical system, see eq. (2.3), one can try to perform a reconstruction by a deconvolution with the PSF.

Various deconvolution techniques have been studied together with Thorsten Schmidt. For more information please refer to (Schmidt, 2003) and the summary in section 10.5.1.

The main disadvantage of the deconvolution techniques is that it is principally not possible to reconstruct those spatial–frequency components that have not been transmitted by the optical system due to its finite bandwidth, and that it is always necessary to record and process the whole volume even if only one slice is needed.

2.3.3 Combined Hardware+Software Approach: Structured Light

An interesting approach that combines additional hardware in the microscope with an appropriate post-processing of the recorded data is the optical sectioning by structured light (Neil et al., 1997).

For this technique a grid is placed in the illumination path, such that it is projected into the object plane (see figure 2.12).

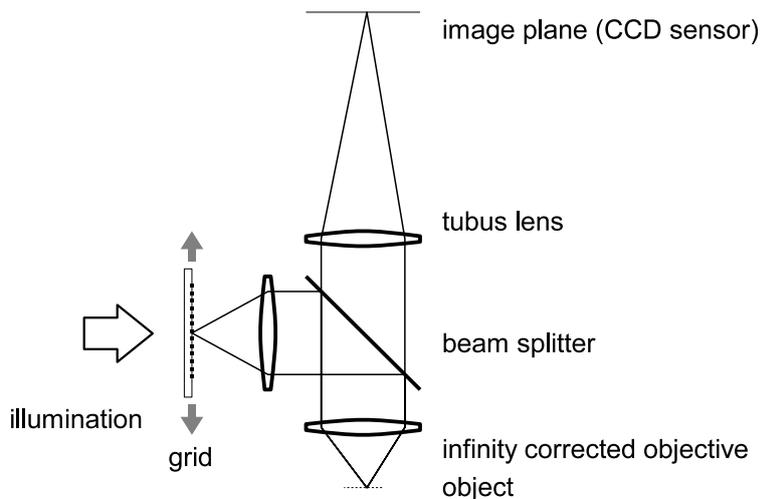


Figure 2.12: Optical sectioning by structured light

This results in a grid shaped illumination pattern in the focal plane. With increasing distance to the focal plane this grid pattern becomes defocussed resulting in a more and more homogeneous illumination.

The resulting intensity distribution in the image plane can therefore be interpreted as a sum of the spatially modulated signal originating from the focus plane and the non-modulated signal originating from the out-of-focus regions.

Now multiple images are taken at different translations of the grid. The reconstruction algorithms then separate the modulated signal (originating from the focal plane) from the non-modulated signal.

The advantages of this technique is the improved z-resolution (compared to conventional microscopy and the deconvolution techniques), and the much faster recording times (compared to a confocal microscope).

The most important disadvantage is that the object must not change between the different recordings, e. g. due to movements or due to bleaching effects.

The imperfect reconstructions with the techniques described in [Neil et al. \(1997\)](#) (these algorithms leave visible stripes in the images) gave reason for further analysis of the effects and the search for better reconstruction algorithm together with Anselm Vossen ([Vossen, 2004](#)).

Meanwhile better algorithms are available ([Schaefer et al., 2004](#)) and are implemented in the commercial software provided by Zeiss for their structured light extension, called “ApoTome”.

2.4 Improvement of the Dynamic Range (HDR Images)

The intensity of the primary fluorescence exhibits great variations between the different pollen taxa. Even pollen grains of the same taxon on one and the same slide may show significantly different fluorescence intensities due to different effects that may have destroyed the fluorophores within the molecules.

To record fluorescence from all particles with a reasonable signal-to-noise ratio for both the very weak and the very strong fluorescent objects within one image, the same image is recorded repeatedly while either the laser power (LSM) or the shutter time of the CCD-camera (conventional fluorescence microscopy) is varied.

The final goal of the fluorescence recording is to measure the fluorophore density at each position within a given object. In thick objects, a significant fraction of the excitation photons and of the emitted photons may be absorbed on the way to and back from the examined position. This results in a darker appearance of structures located deeper in the object, even if they have the same fluorophore density. In a usual microscopical setup we cannot measure these absorption effects. Therefore we combine all these effects in the following analysis and just consider the measurable “apparent fluorophore density in the focal point” D . The number of photons $N_i(\mathbf{x})$ measured at position \mathbf{x} in recording i is proportional to this apparent fluorophore density,

$$N_i(\mathbf{x}) = c_i \cdot D(\mathbf{x}) + \Delta N_i(D(\mathbf{x})) \quad (2.12)$$

with a factor c_i depending on all microscopical settings, which influence either the number of the excitation photons or the number of the measured photons, e. g. the illumination time, the laser power, the aperture of the objective, the pinhole diameter, the selected filters and beam splitters, etc. ΔN_i is a noise term, that also depends on the selected microscopical settings and the apparent fluorophore density.

As mentioned above, the best way to control c_i is to keep all microscopical settings constant and to vary only the laser power or the illumination time. Both parameters allow a fast, well-defined and reproducible control of the number of photons that reach the sensor. In the following we only refer to the “laser power” in order to keep the description brief and intuitive. Of course any other parameter, that allows a well-defined control of c_i can be used instead.

Both, the respective CCD element and the photo-multiplier can measure the light intensity only up to a certain number of photons N_{sat} . Exceeding this number the CCD sensor saturates and the photo-multiplier becomes nonlinear. This is illustrated for a hypothetical sensor that is fully saturated at 1000 photons and a hypothetical fluorophore that emits 1000 photons at density 1 and laser power 1 (figure 2.13). In confocal microscopy there are four main sources of noise (Wilhelm et al., 2006):

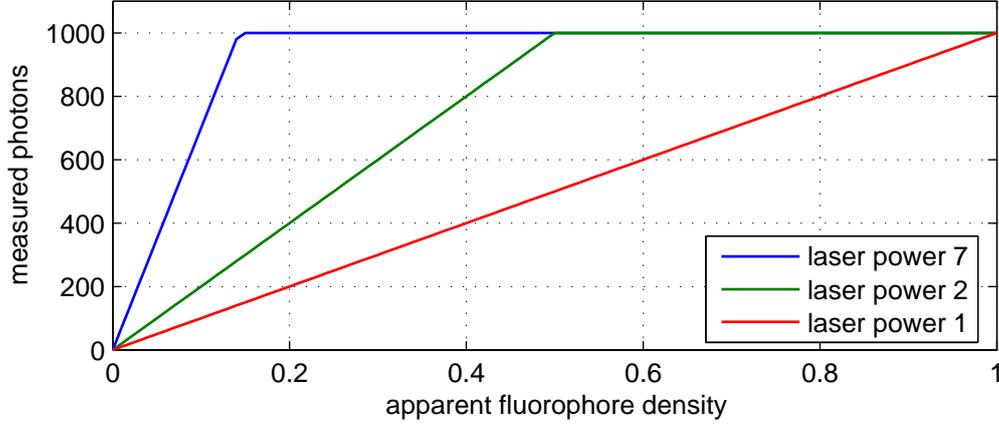


Figure 2.13: Model for number of measured photons for different fluorophore densities and different laser powers. The assumed sensor is fully saturated at 1000 photons

1. **The laser noise** q caused by random fluctuations of the population of the excited state in the laser medium. This noise is proportional to the number of detected photons N
2. **The shot noise (Poisson Noise)** $\Delta N_{\text{Poisson}} = \sqrt{N}$ caused by the quantum nature of light.
3. **The secondary emission noise** α_{se} caused by random variation of photoelectron multiplication at the dynodes of a photo multiplier tube.
4. **The dark noise** due to random generation of “dark electrons” N_d . Its standard deviation is about $\sqrt{N_d}$

The total error for a given number of photoelectrons N (photons that reached the sensor when a photo-multiplier is used) is therefore (Wilhelm et al., 2006)

$$\Delta N = \alpha_{\text{se}} \cdot \sqrt{(N + N_d)(1 + q^2)} . \quad (2.13)$$

The signal-to-noise ratio

$$R = \frac{N}{\Delta N} \quad (2.14)$$

obtained with different laser powers is depicted in figure 2.14 for the parameters that have been used in an example in (Wilhelm et al., 2006), namely $N_d = 100$, $\alpha_{\text{se}} = 1.2$, and $q = 0.05$

For the reconstruction of one high dynamic range image from multiple recordings, a weighted average over the measured non-saturated intensities is computed for each position \mathbf{x} . We define

$$\begin{aligned} n &: \text{number of recordings} \\ p_i &: \text{laser power of the } i\text{th recording, beginning with the highest} \\ N_i(\mathbf{x}) &: \text{number of detected photons at position } \mathbf{x} \text{ in the } i\text{th recording} \\ R_i(\mathbf{x}) &: \text{computed signal-to-noise ratio at position } \mathbf{x} \text{ in the } i\text{th recording} \\ W(\mathbf{x}) &= \left\{ i \mid (1 \leq i \leq n) \wedge (N_i(\mathbf{x}) < N_{\text{sat}}) \right\} \\ &\quad \text{indices of well exposed recordings at position } \mathbf{x}. \end{aligned} \quad (2.15)$$

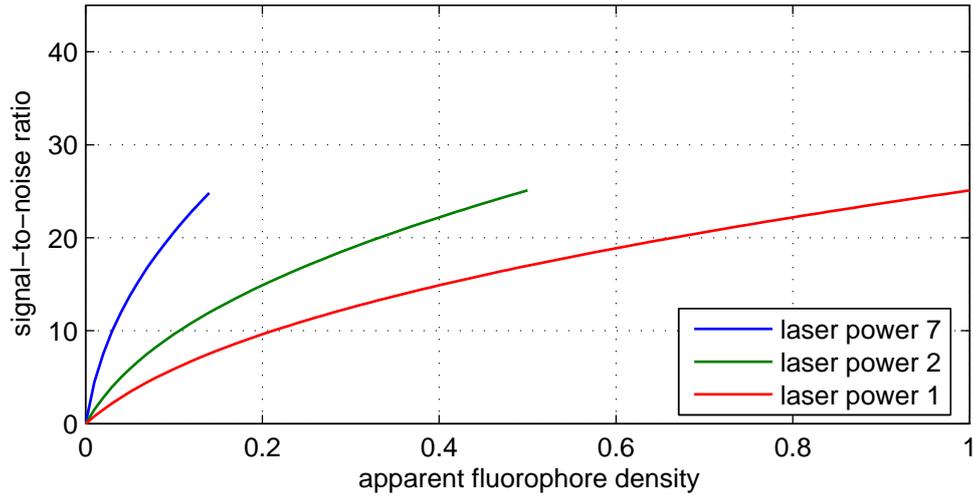


Figure 2.14: Signal-to-noise ratio of measured fluorophore density for different laser powers, according to the noise model of (Wilhelm et al., 2006)

The best way to estimate the number of photons $N_{\text{HDR}}(\mathbf{x})$ that have been received by the virtual non-saturation-limited sensor at the highest laser power p_1 is to compute the weighted average of the recorded intensities,

$$N_{\text{HDR}}(\mathbf{x}) = \frac{1}{\sum_{i \in W(\mathbf{x})} R_i^2(\mathbf{x})} \cdot \sum_{i \in W(\mathbf{x})} R_i^2(\mathbf{x}) \cdot \frac{p_1}{p_i} \cdot N_i(\mathbf{x}) . \quad (2.16)$$

Finally one obtains for the signal-to-noise ratio at the location \mathbf{x} within the resulting image

$$R_{\text{HDR}}(\mathbf{x}) = \sqrt{\sum_{i \in W(\mathbf{x})} R_i^2(\mathbf{x})} . \quad (2.17)$$

This is shown in figure 2.15 as a function of the apparent fluorophore density for the parameters given above. It is interesting to compare the signal-to-noise ratio between the considered HDR image

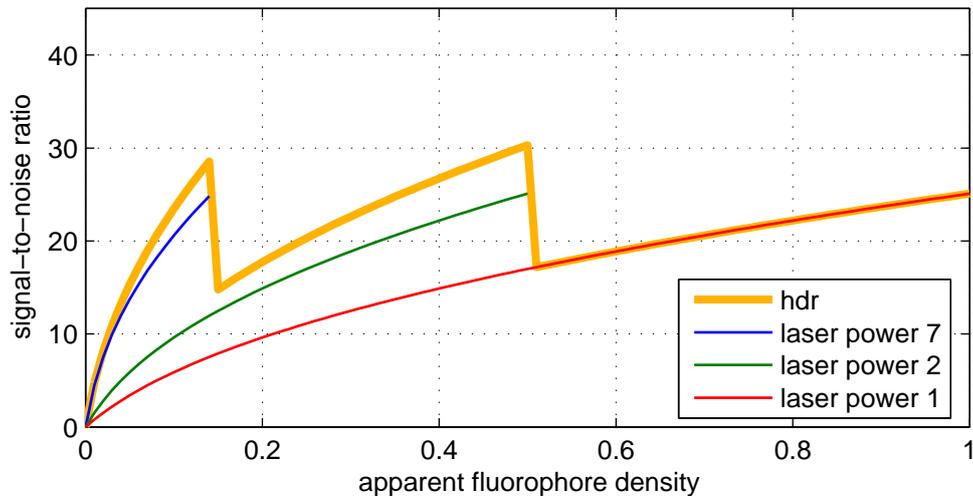


Figure 2.15: Signal-to-noise ratio of computed HDR-image

and the conventional average over a 3-fold recording at the lowest laser intensity which takes the same time as the multiple exposures for the HDR image (see figure 2.16 for the considered LSM example): While the SNR for medium fluorophore densities is comparable for both techniques, the HDR image

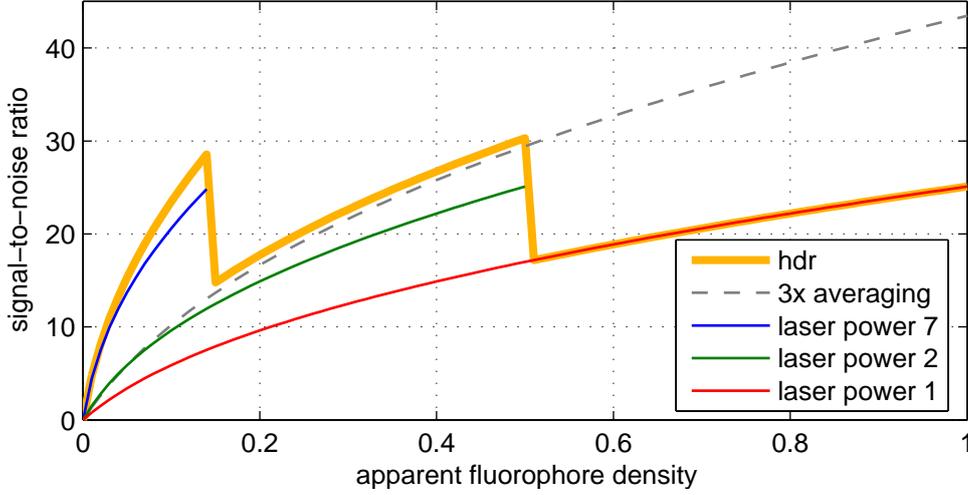


Figure 2.16: Comparison of Signal-to-noise ratio from the HDR image to a conventional 3-fold averaging at the lowest laser power.

has a significantly better SNR in dark regions at the expense of a lower SNR in bright regions. This more homogeneous SNR of a HDR image is preferable, especially for image analysis tasks, because structures that are located at different depths and therefore are subject to different absorption of the excitation and emitted photons, can be compared with less bias than in a conventional image.

2.4.1 Computing the Number of Photons from the Gray Values

In order to apply the HDR formula (2.16) in a real setup, the ratio a between the required “number of photons” and the recorded gray value

$$N(\mathbf{x}) = a \cdot I(\mathbf{x}) \quad (2.18)$$

is needed. This ratio is usually unknown, and therefore must be determined experimentally.

The required information can be extracted from multiple recordings I_j , $j = 1 \dots m$ of the same sample that contains all gray values. For each pixel, the mean and standard deviation are computed,

$$I_\mu(\mathbf{x}) = \frac{1}{m} \sum_{j=1}^m I_j(\mathbf{x})$$

$$I_\sigma(\mathbf{x}) = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (I_j(\mathbf{x}) - I_\mu(\mathbf{x}))^2}, \quad (2.19)$$

which are inserted into (2.13)

$$a \cdot I_\sigma(\mathbf{x}) \stackrel{!}{=} \alpha_{se} \cdot \sqrt{(a \cdot I_\mu(\mathbf{x}) + N_d)(1 + q^2)}$$

$$\Rightarrow 0 \stackrel{!}{=} a^2 \cdot I_\sigma^2(\mathbf{x}) - a \cdot I_\mu(\mathbf{x}) \cdot \alpha_{se}^2 (1 + q^2) - N_d \cdot \alpha_{se}^2 (1 + q^2). \quad (2.20)$$

Thus the value of a can be determined by a least-square fit.

2.4.2 Compensation of Uniform Bleaching Effects and Offsets

First HDR experiments both, with images recorded by a LSM and by a CCD camera mounted to a conventional fluorescence microscope resulted in factors between the exposure times, respectively between the laser intensities which did not comply with equation (2.16).

One of the main reasons for this discrepancy is the bleaching of the fluorophores by the illumination which directly changes the fluorophore density. The second reason are the various effects, that add a fixed offset to the recorded gray values such that a zero of detected photons does not result in a zero gray value. Especially with CCD sensors (their noise is caused significantly by the electrical processing, and therefore may lead to negative output), the zero photon point is mapped intentionally to a positive gray value to avoid systematic errors at low intensities.

If we can assume, that the bleaching causes a decrease of the fluorophore density by a factor which is independent of the position in the image, the HDR reconstruction is still possible. Using this assumption we can estimate the bleaching factor from those regions that are well exposed in at least two recordings. If we furthermore assume that the gray value offsets are independent of the position in the image, we can rewrite the HDR equation (2.16) to

$$N_{\text{HDR}}(\mathbf{x}) = \frac{1}{\sum_{i \in W(\mathbf{x})} R_i^2(\mathbf{x})} \cdot \sum_{i \in W(\mathbf{x})} R_i^2(\mathbf{x}) \cdot (s_i \cdot N_i(\mathbf{x}) + b_i) . \quad (2.21)$$

The factor s_i combines the bleaching factor and the laser power factor. This factor s_i and the offset b_i of the recording with the highest laser power are set to $s_1 = 1$ and $b_1 = 0$, such that the gray values in the reconstructed image correspond to this setting. Let

$$\mathbf{P}_i = \{ \mathbf{x} \mid N_i(\mathbf{x}) < N_{\text{sat}} \} \quad (2.22)$$

be the set of well exposed positions in recording i . Then the parameters s_2 and b_2 can be found by a linear regression

$$s_2 \cdot N_2(\mathbf{x}) + b_2 \stackrel{!}{=} N_1(\mathbf{x}) \quad \forall \mathbf{x} \in \{\mathbf{P}_1 \cap \mathbf{P}_2\} \quad (2.23)$$

and iteratively the parameters s_i and b_i for $i = 3 \dots n$ from the successive recordings,

$$s_i \cdot N_i(\mathbf{x}) + b_i \stackrel{!}{=} s_{i-1} \cdot N_{i-1}(\mathbf{x}) + b_{i-1} \quad \forall \mathbf{x} \in \{\mathbf{P}_i \cap \mathbf{P}_{i-1}\} . \quad (2.24)$$

After the s_i and the b_i have been determined, the final HDR image can be computed by equation (2.21).

2.4.3 Compensation of Nonuniform Bleaching Effects

While the uniform bleaching correction according to the previous section provides satisfactory results on stained samples, it was not able to reconstruct HDR images from unstained real air samples without visible discontinuities. The main reason for this was found to be the different bleaching characteristics of the diverse types of natural fluorophores in such samples. So the previous assumption that factor s_i is independent of the position in the image is not fulfilled.

To keep the model simple, all the offsets b_i are set to zero. So we are confronted with the task to estimate the combined bleaching and laser power factors, denoted by $S_i(\mathbf{x})$, in those regions, where one of the images is saturated, from those regions, where both images are well exposed. For this purpose we compute the ‘‘factor image’’ for two successive recordings, e. g. the first two:

$$S_2(\mathbf{x}) = \begin{cases} N_1(\mathbf{x}) / N_2(\mathbf{x}), & \text{if } \mathbf{x} \in \{\mathbf{P}_1 \cap \mathbf{P}_2\} \\ \text{missing value,} & \text{else .} \end{cases} \quad (2.25)$$

The task to find the missing values in $S_2(\mathbf{x})$ can now be formulated as an interpolation problem. See fig 2.17 for an illustration. Several interpolation algorithms have been tested for this purpose. The requirements to the algorithms are

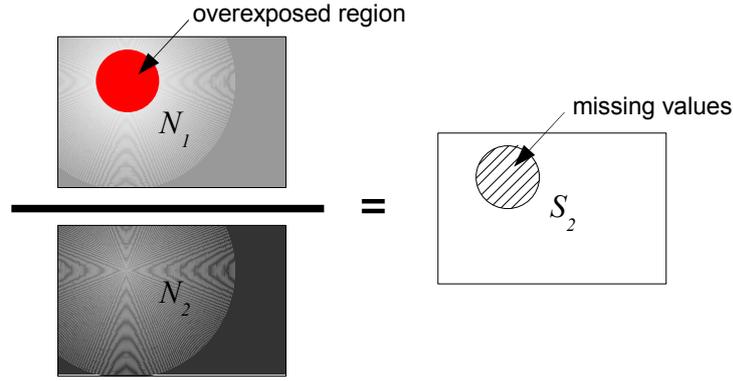


Figure 2.17: In samples with nonuniform bleaching, the factor $S_2(\mathbf{x})$ depends on the position in the image. For the HDR reconstruction it must be appropriately interpolated in the “missing values” regions.

- The resulting array must be continuous
- The experimental errors cause some scatter of the raw values in S_2 which may be reduced by smoothing before or as part of the interpolation. In the first case the smoothing algorithm must be able to cope with the missing values.
- The missing value regions are not compact, they usually have very random borders.
- For the application in the pollenmonitor the interpolation algorithm must be fast (only a few seconds for three stacks with about 100 mega voxels each) and must not use too much memory.

An obvious approach which meets these requirements is a weighted average over the non-missing-values in the surrounding of a missing value. With a mask

$$M(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \{\mathbf{P}_1 \cap \mathbf{P}_2\} \\ 0 & \text{else,} \end{cases} \quad (2.26)$$

and a smoothing kernel $K(\mathbf{x})$ the interpolated array can be computed as

$$S_{2,\text{interp.}} = \frac{(M \cdot S_2) * K}{M * K}, \quad (2.27)$$

wherein $(M \cdot S_2)$ denotes an element-wise multiplication $(M \cdot S_2)(\mathbf{x}) := M(\mathbf{x}) \cdot S_2(\mathbf{x})$. By the use of the separable kernel

$$K(x, y, z) = \alpha^{|x|} \cdot \alpha^{|y|} \cdot \alpha^{|z|} \quad \text{with } 0 < \alpha < 1, \quad (2.28)$$

which decays exponentially along each coordinate axis and which can be effectively computed in the image domain by a recursive scheme, the timing constraints can be fulfilled on a down-scaled array. While this interpolation scheme works well for small regions of missing values, the results are disappointing within large regions: It was not possible to find a value of α that smoothly interpolates such large regions while the important local variations of S_2 are retained. Furthermore numerical problems occurred within such large regions.

After testing several alternative interpolation schemes, the following procedure was found to best fulfill the above criteria: The scattering of the factor array S_2 from N_1 and N_2 is reduced by taking the local maxima in a certain rectangular surrounding $|\mathbf{q}|_\infty < k$

$$S_{2a}(\mathbf{x}) = \frac{\max \left\{ N_1(\mathbf{x} + \mathbf{q}) \mid |\mathbf{q}|_\infty < k \wedge (\mathbf{x} + \mathbf{q}) \in \{\mathbf{P}_1 \cap \mathbf{P}_2\} \right\}}{\max \left\{ N_2(\mathbf{x} + \mathbf{q}) \mid |\mathbf{q}|_\infty < k \wedge (\mathbf{x} + \mathbf{q}) \in \{\mathbf{P}_1 \cap \mathbf{P}_2\} \right\}}. \quad (2.29)$$

A further advantage of the local maximum is, that it selects the pixels nearest to the boundary of the “missing value” regions. Together with this local maximum extraction a sub-sampling of the arrays is performed to reduce the further computational costs. The next step is the line-wise linear interpolation of the missing values from their nearest left and right neighbors $\ell(\mathbf{x})$ and $\mathbf{r}(\mathbf{x})$ with “non-missing” values. Using

$$\ell \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \max\{q \mid q \leq x_1 \wedge (q, x_2, x_3) \in \{\mathbf{P}_1 \cap \mathbf{P}_2\}\} \\ x_2 \\ x_3 \end{pmatrix} \quad (2.30)$$

and

$$\mathbf{r} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \min\{q \mid q \geq x_1 \wedge (q, x_2, x_3) \in \{\mathbf{P}_1 \cap \mathbf{P}_2\}\} \\ x_2 \\ x_3 \end{pmatrix} \quad (2.31)$$

the line-wise interpolation can be written as

$$S_{2b}(\mathbf{x}) = \frac{\|\mathbf{x} - \mathbf{r}(\mathbf{x})\| \cdot S_{2a}(\ell(\mathbf{x})) + \|\mathbf{x} - \ell(\mathbf{x})\| \cdot S_{2a}(\mathbf{r}(\mathbf{x}))}{\|\mathbf{r}(\mathbf{x}) - \ell(\mathbf{x})\|} \quad \forall \mathbf{x} \notin \{\mathbf{P}_1 \cap \mathbf{P}_2\} . \quad (2.32)$$

After smoothing with the cheap recursive filter mentioned above

$$S_{2c} = S_{2b} * K , \quad (2.33)$$

and up-scaling to the original size, the resulting array S_{2c} can be used for the HDR reconstruction. The factor arrays S_{ic} for the successive recordings are computed accordingly.

2.5 Improvement of Homogeneity

2.5.1 Compensation of Errors from Illumination and Recording Step (Static Shading Correction)

Most of the errors that occur in the recording step (CCD dark signal, photo response non-uniformity, dead pixels, digitalization offset, inhomogeneous illumination) do not vary from image to image and can therefore be compensated by a pixel-wise calibration. As CCD’s are linear over a wide range of intensities, we can limit this calibration to a linear compensation. To measure the offset and the factor (true gray value / measured gray value) it is sufficient to record an empty sample with different exposure times and perform a linear regression for each pixel resulting in an “offset image”, a “gain image” and a “correlation coefficient image”, that shows the linearity between the exposure times and the measured gray value. The detection of defect pixels can be done by searching outliers in these three images.

2.5.2 Compensation of Inhomogeneities due to Out-Of-Focus Dust (Dynamic Shading Correction)

Additionally to the static inhomogeneities in the illumination of the sample, there can be absorption of light by out-of-focus dust or other particles. As this “background” changes from image to image, we need some more prior knowledge to separate it from the focussed particles.

With the following assumptions

- All wanted particles (e. g. pollen) together cover less than half of the pixels of the image
- the maximum area of a wanted particle is known

- the background changes only slightly within the diameter of a particle

we can estimate the background gray value for each pixel by computing the median gray value of an area around this pixel that is twice as large as the maximum particle area (see figure 2.18).

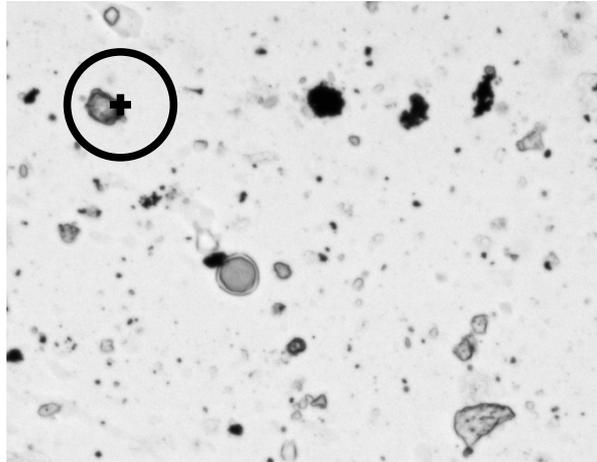


Figure 2.18: To estimate the background gray value at the position of the cross, the median of all gray values within the circle is computed

3 Data Sets

3.1 Confocal Data Set

3.1.1 Collection, Preparation and Recording

The “confocal data set” was recorded in order to do the first feasibility studies for pollen recognition from 3D volumetric data. We selected confocal laser scanning microscopy to ensure the highest available data quality.

The pollen grains were manually collected from the plants and embedded in glycerine to avoid fluorescence background from the embedding media. They were glued directly to the cover slip. A spacing ring ensured that they were not deformed by the preparation (see figure 3.1).

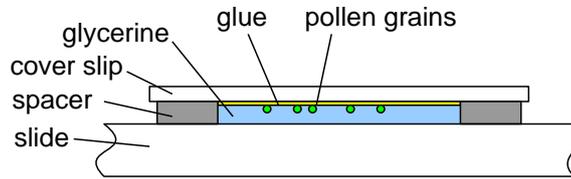


Figure 3.1: Preparation of pollen grains for confocal recording

For each of the 28 most common pollen taxa in Germany one slide was prepared and about 15 individual pollen grains were scanned with a confocal laser scanning microscope (Zeiss LSM 510) equipped with a Plan-Neofluar 40x/1.3 oil-immersion objective. The used excitation wavelength was 488 nm and the emitted light (using the natural fluorescence of pollen grains) was recorded above 515 nm (Zeiss filter set 09). The voxel-size was set to $0.20\mu\text{m} \times 0.20\mu\text{m} \times 0.20\mu\text{m}$ to receive cubic voxels. The laser power and photo multiplier gain were adjusted such that no overexposure occurred, the data were digitized with 8 bit per gray value.

Due to the preparation of pure samples, no additional labeling from pollen experts was needed. An example for the recorded raw data is shown in figure 3.2.

It is obvious that the outer layer, the *exine*, of the pollen grain show only a very low fluorescence compared to the inner parts. As these outer structures are crucial for the identification of the pollen grains, the first pre-processing step of the data was usually a gamma correction (with $\gamma \approx 2$) of the gray values,

$$v_{\text{corr}} = 255 \cdot (v/255)^{(1/\gamma)} . \quad (3.1)$$

After that correction the overall structure of the pollen grain is much better visible (figure 3.3), and the brightness of the structures corresponds better to their importance for the classification.

The mean diameter of the recorded pollen grains range from $15\mu\text{m}$ (*Urtica*^(D.15)) to $85\mu\text{m}$ (*Larix*^(D.8)) resulting in volumetric data sets from about 60^3 voxels to 425^3 voxels.

For the two largest pollen taxa (*Larix*^(D.8) and *Pinus*^(D.10)) significant problems occurred during recording. Due to absorption effects, the lower layers of the pollen grains were nearly invisible such that these two pollen taxa were excluded from the further analysis. The remaining largest pollen grains were *Secale*^(A.7) with a mean diameter of $50\mu\text{m}$ (about 250^3 voxels).

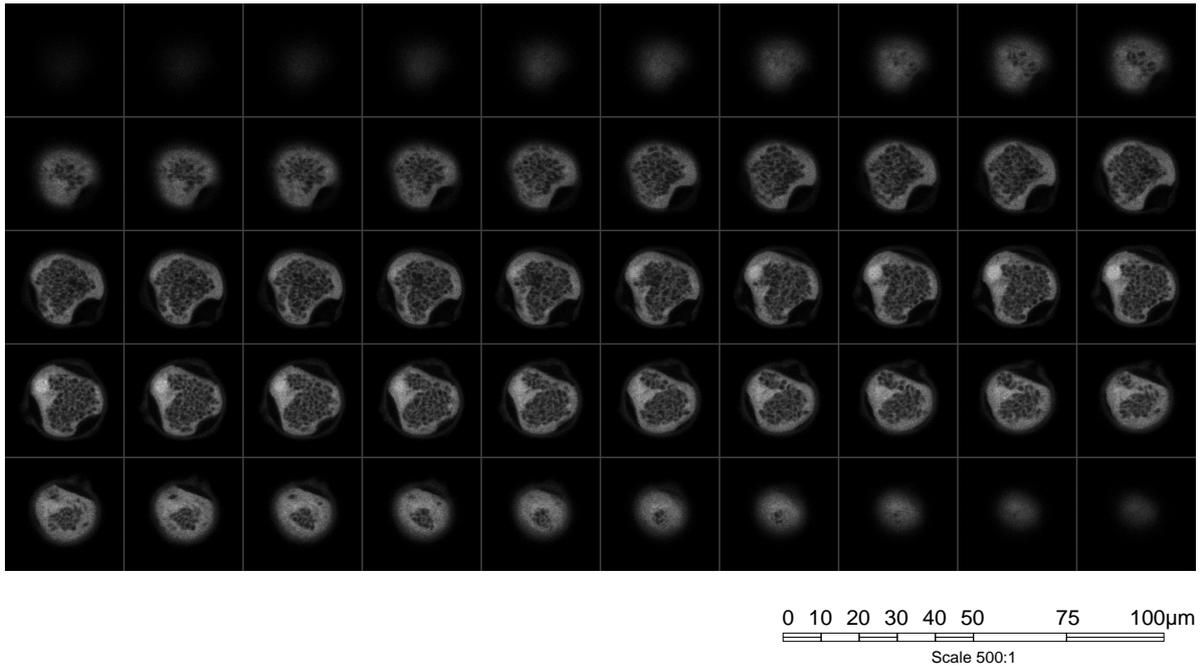


Figure 3.2: Raw data from the confocal laser scanning microscope. Every second slice of a recorded *Betula*^(A.4) pollen is shown (scale 500:1).

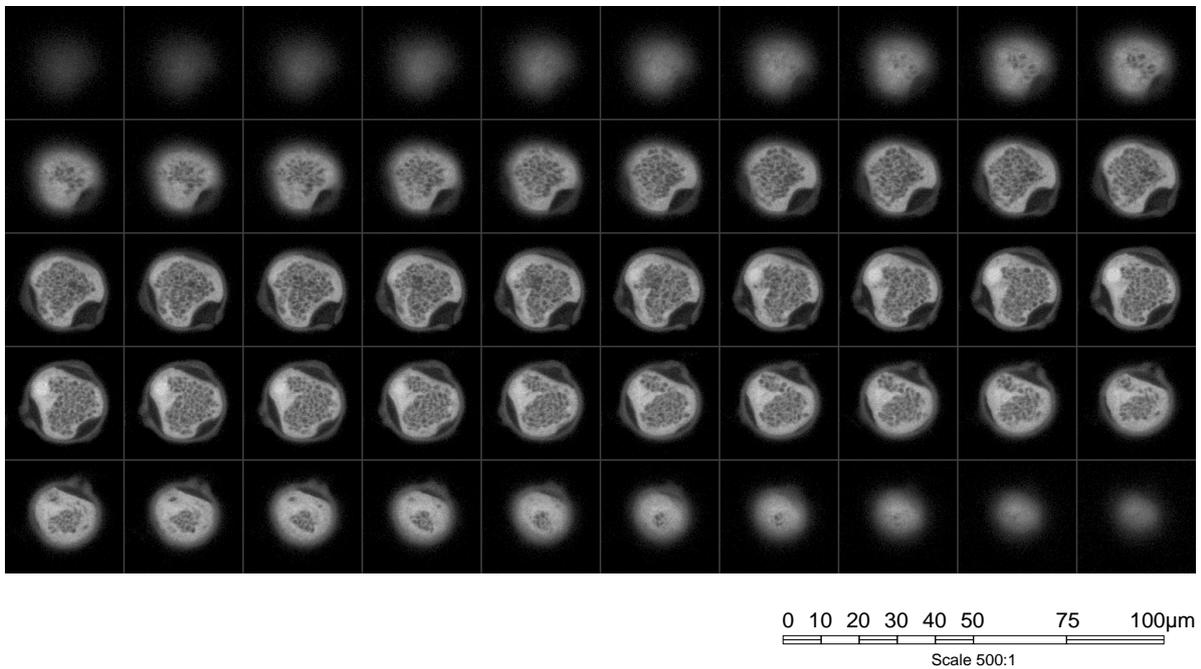


Figure 3.3: Gamma corrected data set of *Betula*^(A.4) pollen (scale 500:1). After a gamma correction with $\gamma = 2$ the important outer structures become brighter.

3.1.2 Representative Table

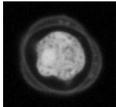
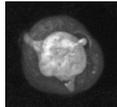
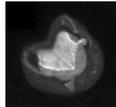
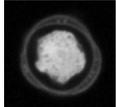
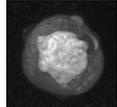
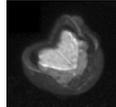
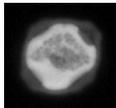
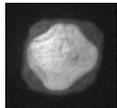
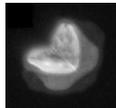
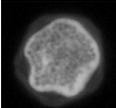
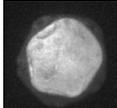
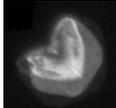
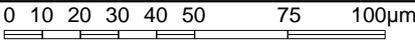
In the following table a short overview over the 26 used pollen taxa is given. For every taxa the two representative pollen grains are shown in different representations. The complete data set is presented in the Appendices [A](#) (highly allergenic pollen taxa), [B](#) (moderately allergenic pollen taxa), [C](#) (lowly allergenic pollen taxa) and [D](#) (non-allergenic pollen taxa). There a better impression of the inner class variations can be received. (In the electronic version of this document you can click on the pollen names to jump directly to the corresponding page in the Appendix).

In the table the following data are shown:

- **central plane:** The central plane was selected by the center of mass of the gray values in the whole dataset. For a less noisy display, the mean of this plane and its two neighboring planes was computed.
- **shaded volume rendering:** The shaded volume rendering is based on the maximum intensity projection of the gray values combined with a maximum intensity projection of a very simple illumination model using two light sources and ambient light. The reflection properties of each voxel were defined by the gradient direction and the gradient magnitude. No absorption model was applied such that in some renderings small bright spots appear from the reflection at strong gradients within the inner structures.
- **shaded volume rendering of cropped pollen:** Finally to see the inner structure of the pollen, a part of the volumetric data set was set to zero and also processed with the shaded volume renderer. This cropping needed some smooth blending at the cut surfaces to avoid interference patterns originating from the voxel grid.

The pollen grains are sorted according their their allergenic level (see Appendix [E](#) for a complete list and a discussion of this classification).

All images are presented here at the same scale (500:1) such that 1 mm in the image corresponds to $2\mu\text{m}$ in reality.

Highly Allergenic Pollen Taxa			
latin name (German name)	central plane	shaded volume rendering	shad. vol. rend. cropped pollen
<i>Artemisia</i> ^(A.1) (Beifuß)			
			
<i>Alnus</i> ^(A.2) (Erle)			
			
 Scale 500:1			

Highly Allergenic Pollen Taxa

latin name (German name)	central plane	shaded volume rendering	shad. vol. rend. cropped pollen
<i>Alnus viridis</i> ^(A.3) (Alpen-/ Grün-Erle)			
<i>Betula</i> ^(A.4) (Birke)			
<i>Corylus</i> ^(A.5) (Hasel)			
<i>Poaceae</i> ^(A.6) (Süßgräser)			
<i>Secale</i> ^(A.7) (Roggen)			

0 10 20 30 40 50 75 100µm

Scale 500:1

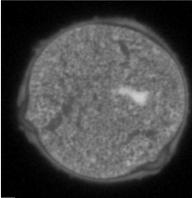
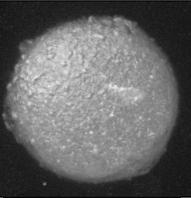
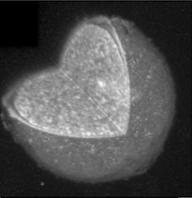
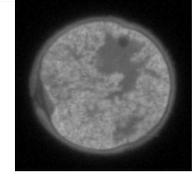
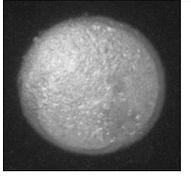
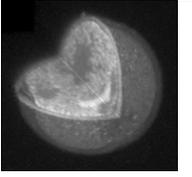
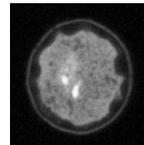
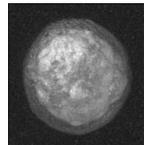
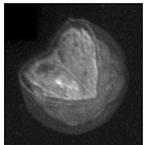
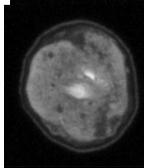
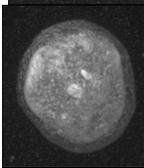
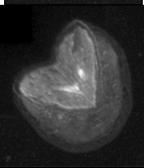
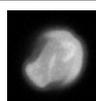
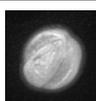
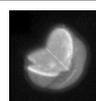
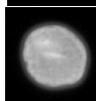
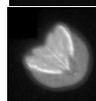
Moderately Allergenic Pollen Taxa

latin name (German name)	central plane	shaded volumeshad. rendering	shad. vol. rend. cropped pollen
<i>Carpinus</i> ^(B.1) (Hainbuche)			
<i>Quercus</i> ^(B.3) (Eiche)			
<i>Fraxinus</i> ^(B.4) (Esche)			
<i>Plantago</i> ^(B.5) (Wegerich)			
<i>Rumex</i> ^(B.6) (Ampfer)			

0 10 20 30 40 50 75 100µm

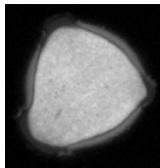
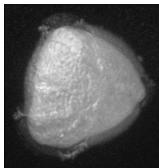
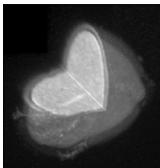
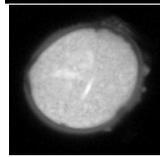
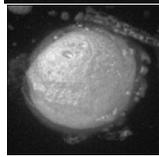
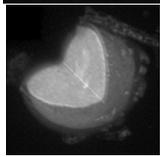
Scale 500:1

Lowly Allergenic Pollen Taxa

latin name (German name)	central plane	shaded volume rendering	shad. vol. rend. cropped pollen
<i>Fagus</i> ^(C.2) (Rotbuche)			
			
<i>Populus</i> ^(C.3) (Pappel)			
			
<i>Salix</i> ^(C.4) (Weide)			
			

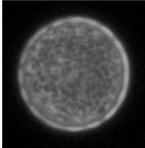
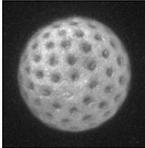
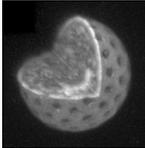
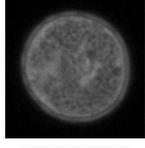
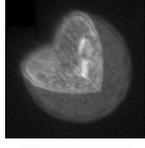
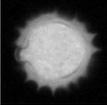
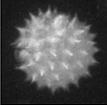
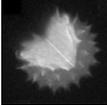
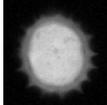
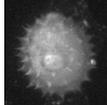
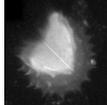
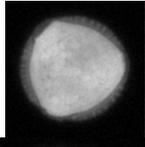
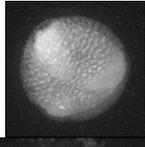
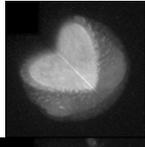
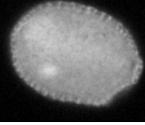
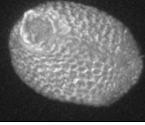
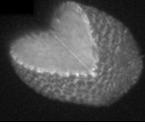
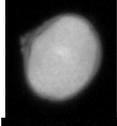
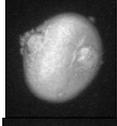
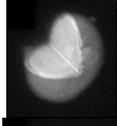
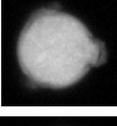
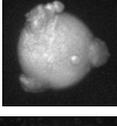
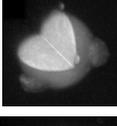
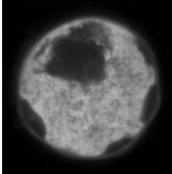
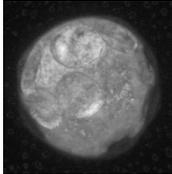
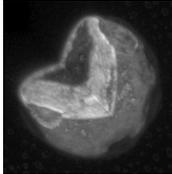
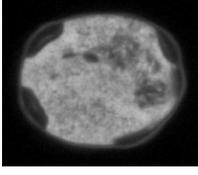
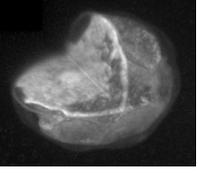
0 10 20 30 40 50 75 100µm
Scale 500:1

Non-Allergenic Pollen Taxa

latin name (German name)	central plane	shaded volume rendering	shad. vol. rend. cropped pollen
<i>Acer</i> ^(D.1) (Ahorn)			
			

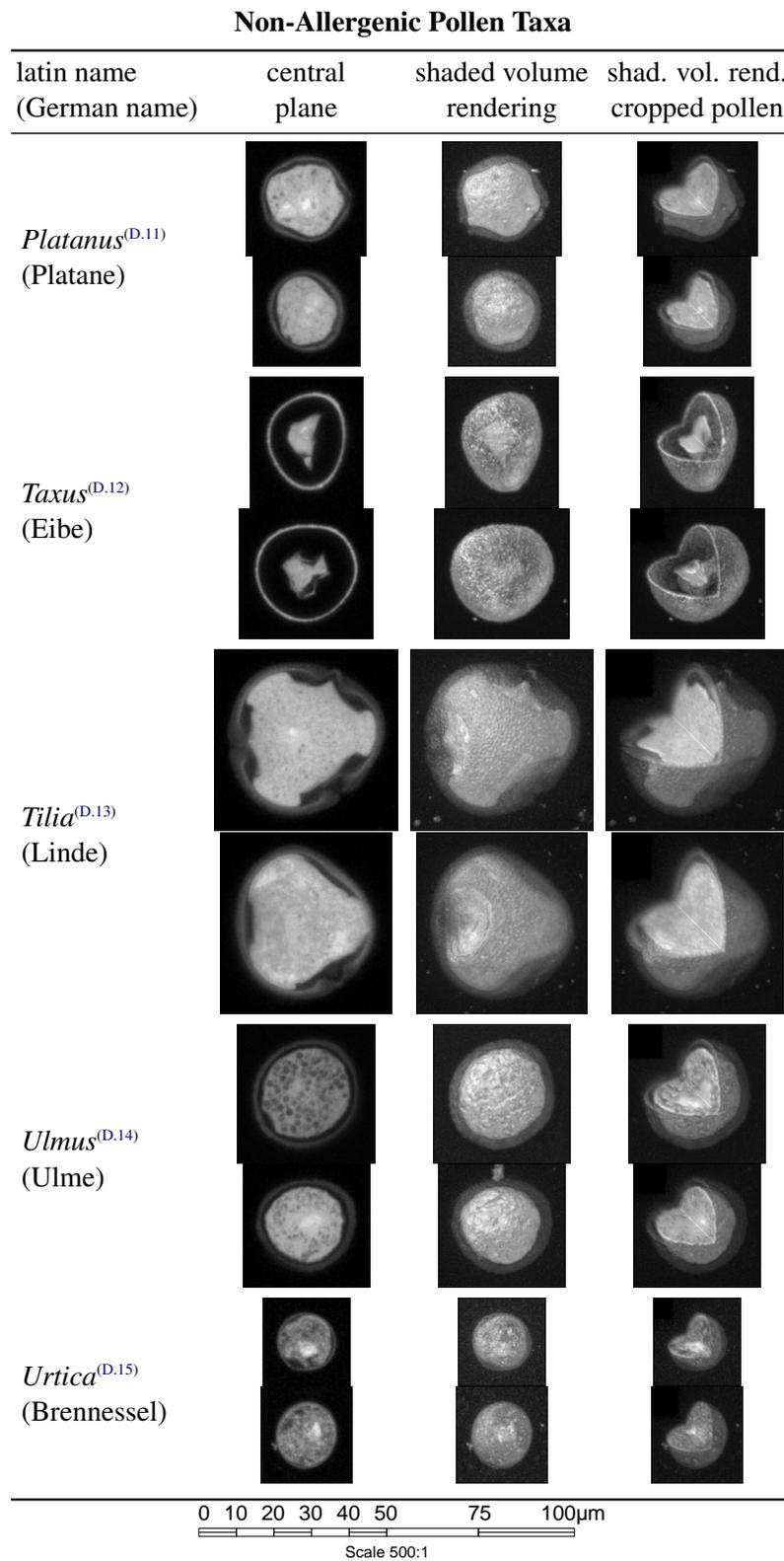
0 10 20 30 40 50 75 100µm
Scale 500:1

Non-Allergenic Pollen Taxa

latin name (German name)	central plane	shaded volume rendering	shad. vol. rend. cropped pollen
<i>Chenopodium</i> ^(D.3) (Gänsefuß)			
			
<i>Compositae</i> ^(D.4) (Korbblüter)			
			
<i>Cruciferae</i> ^(D.5) (Kreuzblüten- gewächse)			
			
<i>Aesculus</i> ^(D.6) (Roßkastanie)			
			
<i>Juglans</i> ^(D.7) (Walnuß)			
			

0 10 20 30 40 50 75 100µm

Scale 500:1

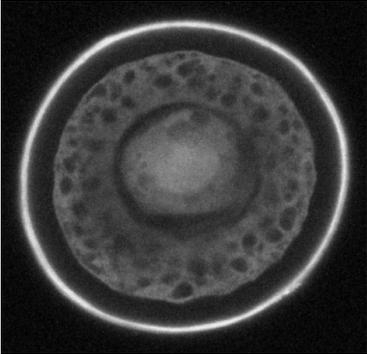
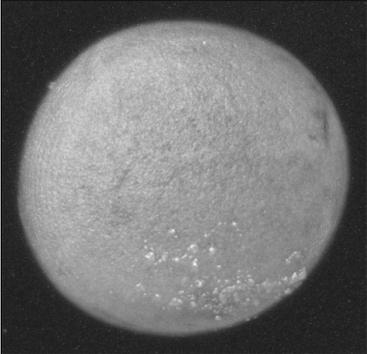
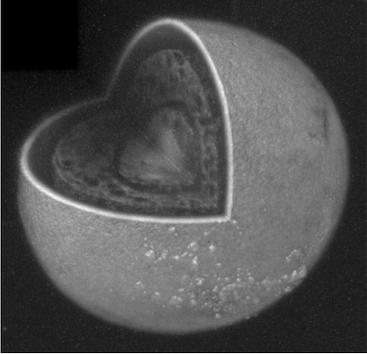
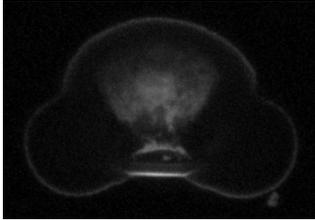
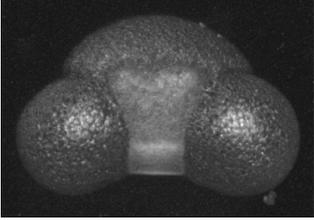
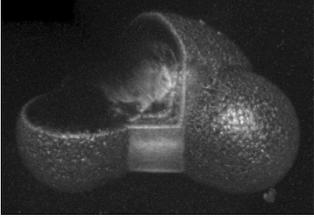


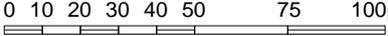
3.1.3 Left-out Taxa

The *Larix*^(D.8) and *Pinus*^(D.10) pollen grains were too big for the selected confocal setup and showed too strong absorptions. They are not allergenic and they can be classified very easy just by the size (*Larix*^(D.8)), or the non-spherical shape (*Pinus*^(D.10)). So they were not included in the experiments. Note

that they are presented here at the same scale (500:1) as the pollen in the previous table.

Non-Allergenic Pollen Taxa (not used due to bad image quality)

code latin name (German name)	central plane	shaded volume rendering
<i>Larix</i> ^(D.8) (Lärche)		 
<i>Pinus</i> ^(D.10) (Kiefer)		 



0 10 20 30 40 50 75 100µm

Scale 500:1

3.1.4 Discussion

Even though we tried to make the preparation and the recording as reproducible as possible, we cannot fully exclude that the recorded data may contain a certain systematic bias, that might lead to positively biased results. Furthermore the pollen grains for each taxon were only collected from a single plant (or very limited number of plants) and do not show all the inner class variations, that occur on a real air sample.

There are two reasons why we did not fully discard these data. The most probable systematic bias

(a certain gray value offset and a gray value scaling) was removed from the data by normalization (for the global invariants), and by designing features that are invariant to such variation (MiSP invariants). The second even more important reason is that all other publications concerning pollen recognition are based on similarly produced data (see section 1.2.1). So for a reasonable comparison of our algorithms to the published “state of the art” we should use data which is comparable to the data used in these studies.

3.2 Pollenmonitor Data Set

The “*pollenmonitor data set*” contains the air samples that were recorded with the first prototype of the pollenmonitor between March, 10th and September, 22nd 2006 in Freiburg and Zürich.

3.2.1 Collection

The pollenmonitor uses a two stage virtual impactor for taking the air samples (see figure 3.4). It

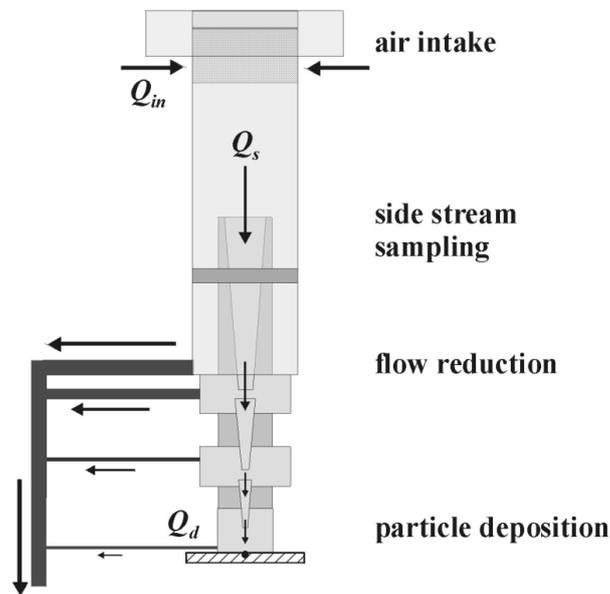


Figure 3.4: (From (Scharring et al., 2006) with kind permission of the authors) Schematic of the design of the sampling and deposition method used in the pollenmonitor. The overall sample flow is maintained by a fan. The internal flows are controlled by orifices. From the overall air stream $Q_{in} = 1167$ l/min a representative side stream of $Q_s = 100$ l/min is taken isokinetically and subsequently concentrated up resulting in a final flow rate of $Q_d = 2$ l/min.

collects the particles of 100 l air per minute and deposits them smoothly (at only 1.7 m/s) on the surface of the embedding medium (glycerine gelatine) on the sample carrier. The sample carrier is a metal plate with 24 holes of about 12 mm diameter (see figure 3.5). Each hole is covered at its bottom with a cover slip, building a small chamber, and filled with glycerine gelatine as embedding medium (see figure 3.6).

After impaction the collected particles cover a circular area with about 5 mm radius. For more details refer to (Scharring et al., 2006). From these parameters we can compute the desired conversion factor c_{air} between the area on the sample and the corresponding air volume as

$$c_{air} = \frac{100 \text{ l/min}}{\pi \cdot (5 \text{ mm})^2} = \frac{6 \text{ m}^3/\text{hour}}{78.54 \text{ mm}^2} = 0.07639 \frac{\text{m}^3}{\text{mm}^2 \cdot \text{hour}} \quad (3.2)$$

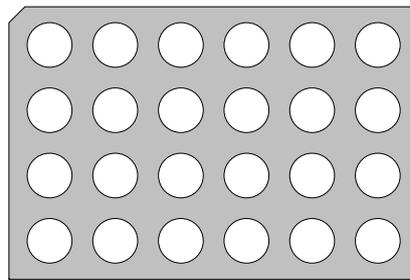


Figure 3.5: Sample carrier of the pollenmonitor containing 24 small chambers for hourly air samples.

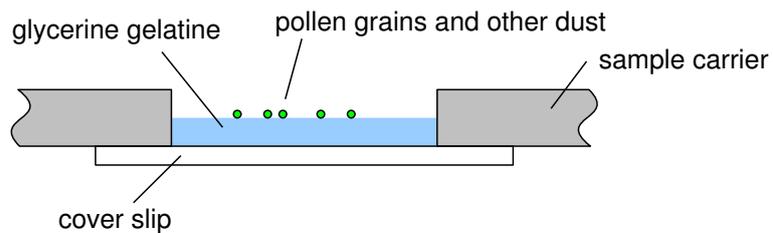


Figure 3.6: Side view of one chamber of the sample carrier (not true to scale)

The sampling duration was usually one hour. In rare cases the sampling duration was reduced in order to avoid too high particle densities.

During 2006 the pollenmonitor was placed in the measurement field of the German weather service in Freiburg downtown (see figure 3.7). At this location some local pollen sources, e. g., the *Urtica*^(D.15) (in German: “Brennessel”) located in the background of the image, may have influenced the total number of collected pollen grains.



Figure 3.7: The pollenmonitor in the measurement field of the German Weather Service in Freiburg. The image was taken on April 28th, 2006.

In the weeks 10-18 the 24-hour-operation of the pollenmonitor was tested by taking one sample every second hour. Later in the year the sampling rates were reduced to about 3 air samples per day just to collect enough pollen for building a complete reference data base. The dates and times of the

day, when the samples were taken are shown in figure 3.8.

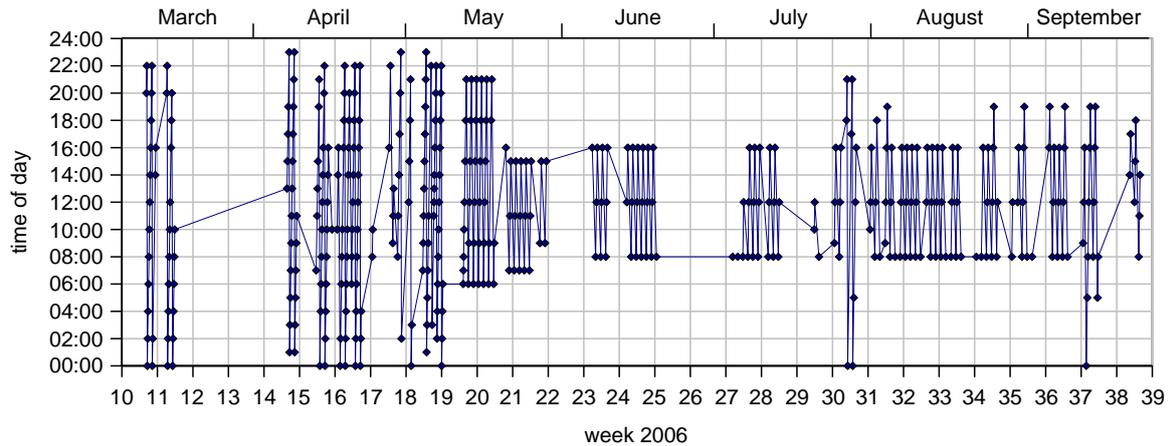


Figure 3.8: Date and time of the day of all air samples within the “pollenmonitor data set”. The samples of the second half of week 18 were taken during a four-day-demonstration of the machine in Zürich

3.2.2 Preparation

In the pollenmonitor an automatic preparation unit prepares the air samples for the microscopical evaluation. To keep the system as simple as possible, the impactor places the pollen grains on the surface of the embedding medium, and the cover slip is already integrated as bottom of the sample carrier, such that the only operation for the preparation is the melting of the glycerine gelatine (see figure 3.9)

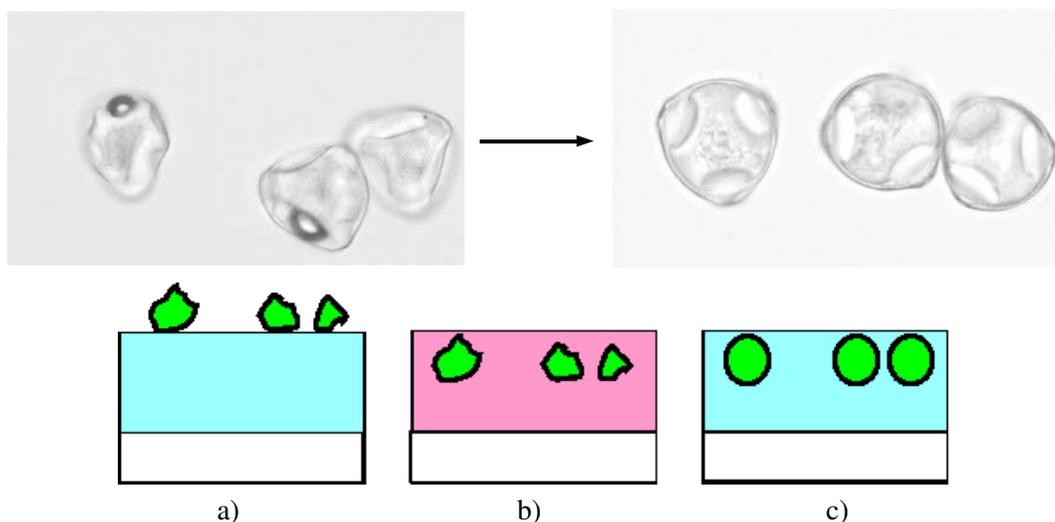


Figure 3.9: Preparation of air samples in the pollenmonitor (from (Scharring et al., 2006) with kind permission of the authors): a) The dried up pollen grains from the air are deposited on the surface of the embedding medium (Glycerine gelatine). b) The embedding medium is heated and melts, such that the pollen grains sink into it. c) After a short period, the pollen grains swelled to their original shape due to the absorption of the water from the embedding medium.

For the data set recorded in 2006, the preparation of the air samples in the pollenmonitor was done immediately after the sampling time ended. In 2005 we noticed significant changes in the morphology of the pollen grains, when the preparation was done several weeks later. To build a reference data base, that can be used in a routine environment it is important to keep this parameters identical to the later application.

3.2.3 Microscopic Recording

The readily prepared sample were microscopically scanned with the integrated inverse microscope. It uses a high-end Zeiss objective “Plan-Apochromat” with 20x magnification and a numerical aperture of 0.8. The light source for the transmitted light images was a green LED (wavelength 533 nm; power 35 mW), and the excitation light for the fluorescence recordings was obtained from a violet LED (wavelength 380 nm; power 85 mW). The broad band emitted light from the pollen grains was recorded above 455 nm.

For the recording a digital firewire camera (AVT Dolphin F-145B from Allied Vision Technologies) equipped with a 2/3" CCD sensor with 1392×1040 pixels and a cell size of $6,45 \mu\text{m} \times 6,45 \mu\text{m}$ was used. The shutter time for the transmitted light images was about 0.4 ms allowing the full frame rate of 15 fps. The fluorescence recordings were done at 3 different shutter times (50 ms, 150 ms and 400 ms) to cover the high dynamic range of the fluorescence signals. For a faster transfer of the images only 8bit of the available 10bit were used. But even in 8bit mode the main noise source was still the photon noise and readout-noise from the CCD sensor such that the left out 2 bits would not have contributed any valuable information.

Due to performance reasons, all stacks were recorded at the highest possible frame rate of the camera. For this purpose, the camera was started in continuous mode, and every time, when the first firewire packet of a new frame arrived at the computer (signaling that the exposure time was over) the command to move to the next z-position was sent to the translation stage.

From each air sample a certain number of multi-channel image stacks were recorded. Until the mid of the year the stacks for each channel were recorded sequentially (see algorithm 3.1) which is less critical to timing, but the channels in the resulting data set are not well-aligned.

Algorithm 3.1 Sequential Scheme for recording a multi-channel image stack within the pollenmonitor.

Output: 1 transmitted light stack and 3 fluorescence image stacks

- 1: **Switch** to **transmitted light** mode
 - 2: **Set** the desired **stack center** in the air sample according to the current grid position
 - 3: **Record** about 100 levels for a **focus stack** with reduced camera resolution (2x2 binning) around the current position (0.1 ms shutter time, $5 \mu\text{m}$ step size); Find the **sharpest level** in the focus stack and compute its z-coordinate
 - 4: **Set** the desired **stack center** in the air sample with the corrected z-coordinate
 - 5: **Record** all 70 levels for the **transmitted light image stack** (0.4 ms shutter time, $1,5 \mu\text{m}$ step size)

 - 6: **Switch** to **fluorescence** mode
 - 7: **Record** all 70 levels for the **“long” fluorescence image stack** (400 ms shutter time, $1,5 \mu\text{m}$ step size)
 - 8: **Record** all 70 levels for the **“middle” fluorescence image stack** (150 ms shutter time, $1,5 \mu\text{m}$ step size)
 - 9: **Record** all 70 levels for the **“short” fluorescence image stack** (50 ms shutter time, $1,5 \mu\text{m}$ step size)
-

While at the begin of the year the alignment of the channels was satisfactorily, it became worse with the increasing abrasion of the translation stage mechanics. Due to this fact and the significant

progresses in the segmentation algorithms that made the fluorescence image less important, several air samples were scanned with transmitted light microscopy only, resulting in single-channel image stacks.

In the mid of the year we finished the new version of the recording software with an interleaved recording of the channels for each image stack (see algorithm 3.2). Here the correspondence of the channels is much better.

Algorithm 3.2 Interleaved scheme for recording a multi-channel image stack within the pollenmonitor.

Output: 1 transmitted light stack and 3 fluorescence image stacks

- 1: **Find best focus** as described in sequential recording scheme
 - 2: **Move Stage** to begin of stack
 - 3: **for all** levels **do**
 - 4: **Switch** to **transmitted light** mode
 - 5: **Record** level i for the **transmitted light image stack** (0.4 ms shutter time)
 - 6: **Switch** to **fluorescence** mode
 - 7: **Record** level i for the **“short” fluorescence image stack** (50 ms shutter time)
 - 8: **Record** level i for the **“middle” fluorescence image stack** (150 ms shutter time)
 - 9: **Record** level i for the **“long” fluorescence image stack** (400 ms shutter time)
 - 10: **Move Stage** $1.5 \mu\text{m}$ in z-direction
 - 11: **end for**
-

The arrangement of all the inner components of the pollenmonitor are shown in figure 3.10.

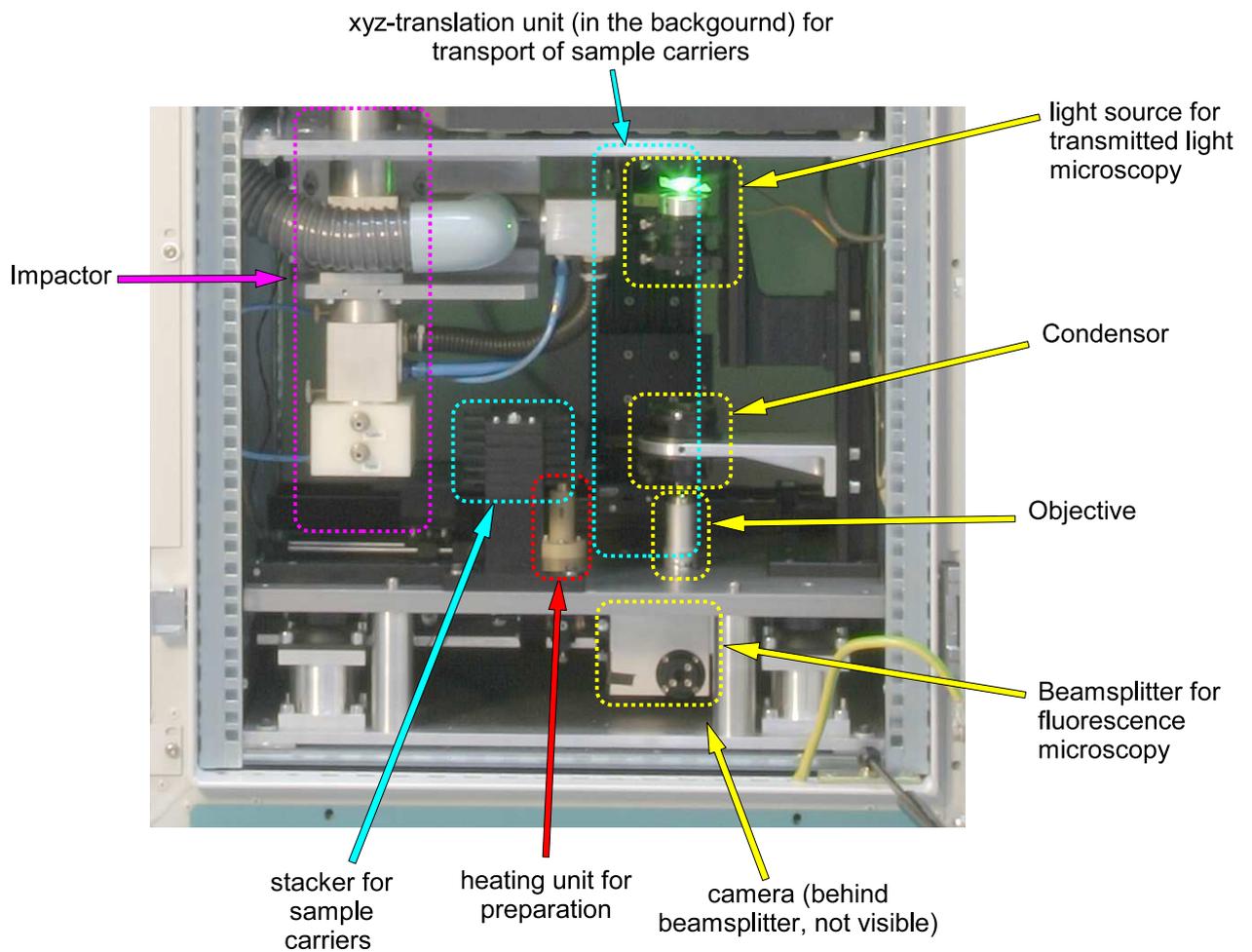


Figure 3.10: Inner components of the pollenmonitor

3.2.4 Evaluated Area per Air Sample

The number of recorded stacks per air sample was varied between 5×5 stacks and 15×15 stacks, according to the number of the air samples taken at this day, the needed scanning time (depending on the number of fluorescence channels that were recorded), the needs for more (or less) examples of the current pollen taxa, etc. (see figure 3.11).

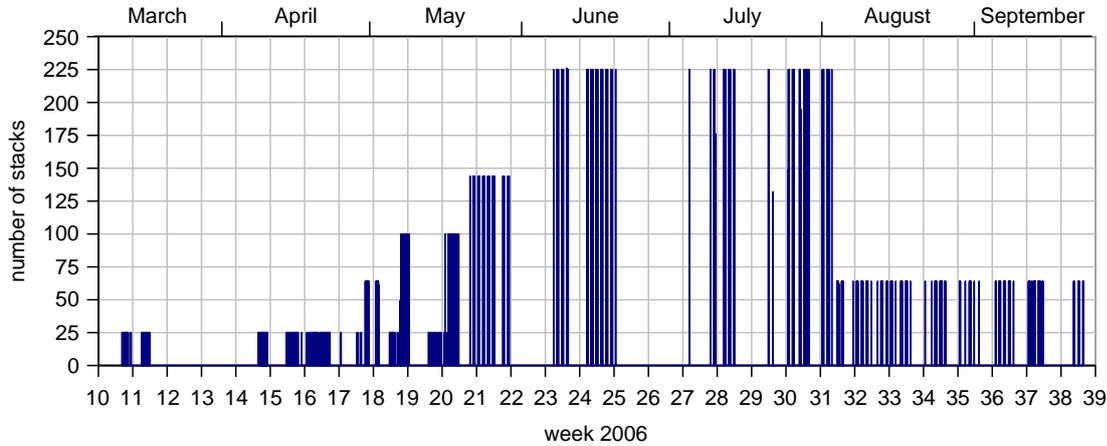


Figure 3.11: Area (in number of stacks) that was scanned from each air sample. This area was varied over the year due to several reasons (see text)

Each stack contains $1392 \times 1040 \times 70$ voxels of size $0.3225 \mu\text{m} \times 0.3225 \mu\text{m} \times 1.5 \mu\text{m}$. To avoid partially cropped pollen only those particles whose center has a minimum distance of 30 pixels ($\approx 10 \mu\text{m}$) to the x/y borders are taken into account, which reduces the evaluated area to $(1392 - 60) \times (1040 - 60) \text{ pixels}^2 = 0.1358 \text{ mm}^2$. Using the conversion factor c_{air} from (3.2) and the usual sampling time of 1 hour the particles on a sample with 5×5 stacks originate from 0.26 m^3 air, and the particles on a sample with 15×15 stacks originate from 2.33 m^3 air. An example for an air sample taken at July 25th, 2006, 16.00h that was scanned with 15×15 stacks is shown in figures 3.12, 3.13 and 3.14. The total number of evaluated stacks is 27,280, which corresponds to an air volume of 282.94 m^3

3.2.5 Labeling

Overall 420 air samples were collected, prepared and scanned in 2006. During the year 325 air samples were selected for manual labeling. (see figure 3.15).

On the selected air samples the segmentation found about 180,000 spherical particles. All of these particles were manually labeled by several pollen experts resulting in about 22,700 labeled pollen grains. (see figure 3.16)

The labeling was done at the computer where the pollen experts had access to the full 3D stack of each particle by using the specially developed programs pfm (“pollen file manager”) which displays all segmented particles (sharpest layer only) in an editable list and polvi (“pollen viewer”) that is started on demand to focus through the whole image stack.

The pollen taxonomy was chosen according to the taxonomy which is used in the daily pollen counts in Germany and Switzerland and which is mainly oriented to the level of visual discernment and to the allergic relevance of the taxa: Most of the pollen are classified at genus level (e. g., *Alnus*^(A.2), *Betula*^(A.4), *Corylus*^(A.5)), some of them can be classified on family level only, e. g. *Poaceae*^(A.6), *Compositae*^(D.4), *Ericaceae* (the names have the suffix “ae”). Very few of them can be classified on species level, e. g. *Alnus viridis*^(A.3), *Ostrya Carpinifolia*^(B.2) or *Fraxinus ornus*. The distinction on species level usually needs more than only the morphology. E. g., *Alnus viridis*^(A.3) can usually be differentiated from the

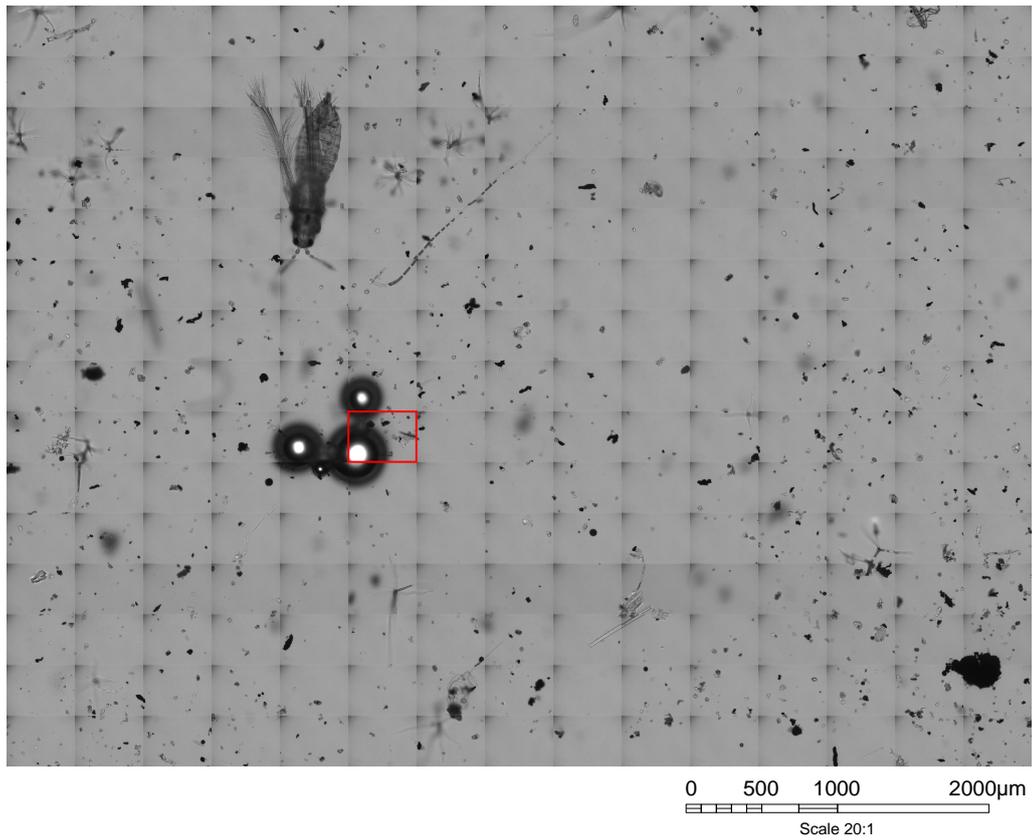


Figure 3.12: Air sample from July 25th, 2006, 16.00h (week 30) taken in Freiburg with the pollen-monitor. At this display scale (20:1) you can see a small insect (about 1mm long) but not the pollen grains. The three black blobs in the center of the image are bubbles from an imperfect preparation. Each of the 15×15 tiles (stacks) contains $1392 \times 1040 \times 70$ voxels, resulting in 22.8 Gigavoxels.

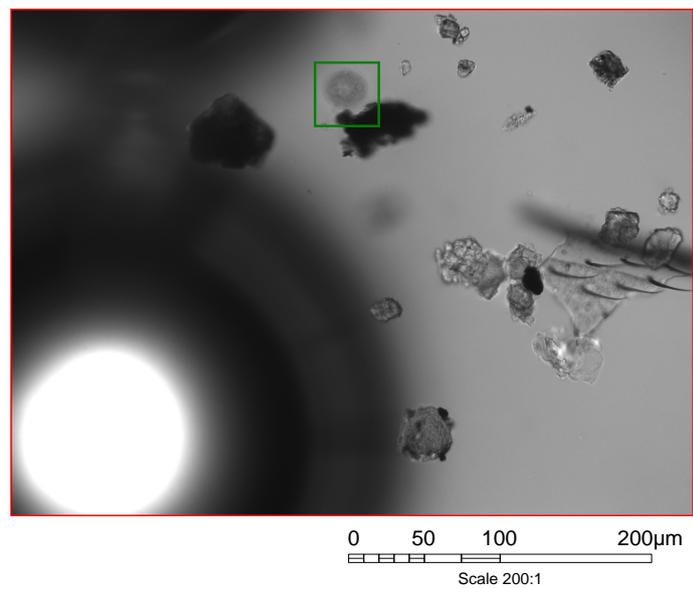


Figure 3.13: Central layer of image stack containing one of the bubbles. The marked pollen grain lies in a slightly different layer.

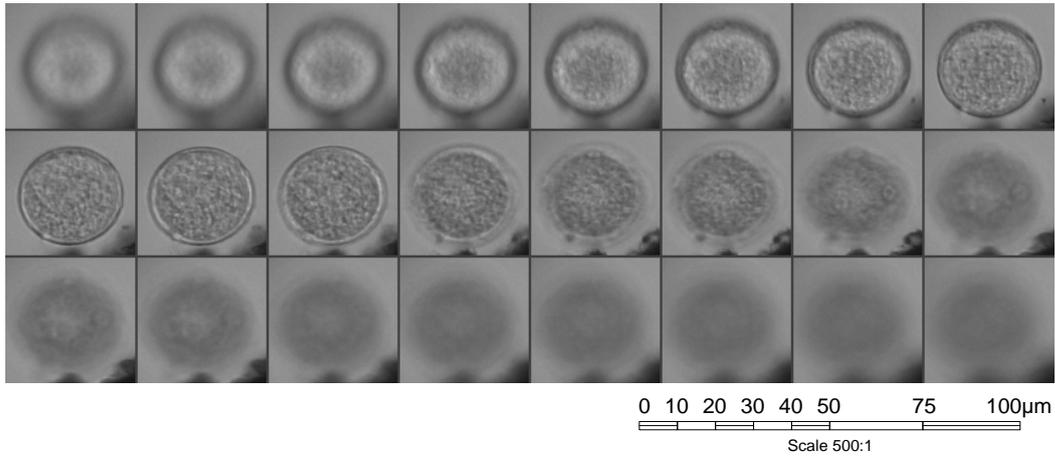


Figure 3.14: Cropped image stack of the pollen grain. By looking at its central layer and some layers above and below, a pollen expert (and our pollen recognition system) can identify it as a *Plantago*^(B.5) pollen.

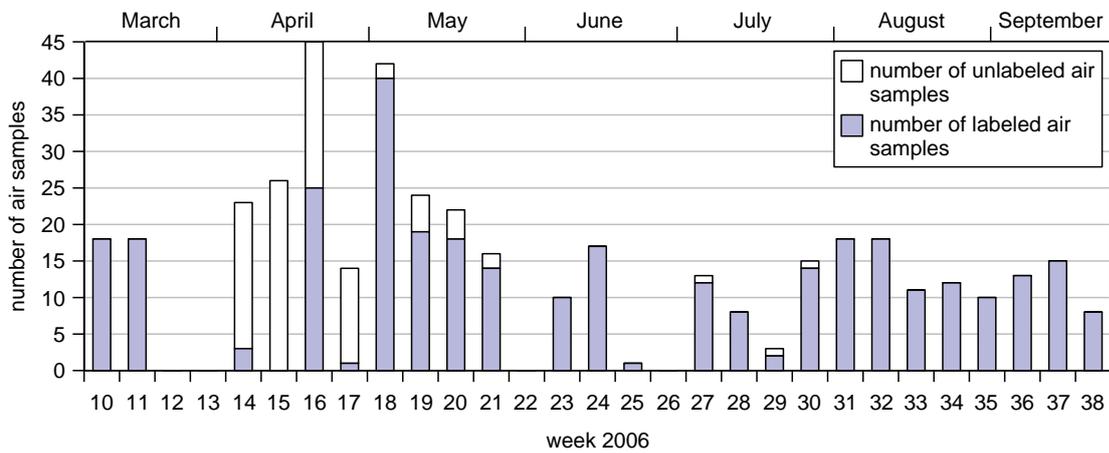


Figure 3.15: Number of air samples taken with the pollenmonitor in 2006. The samples from week 18 are recorded in Zürich, the others are recorded in Freiburg.

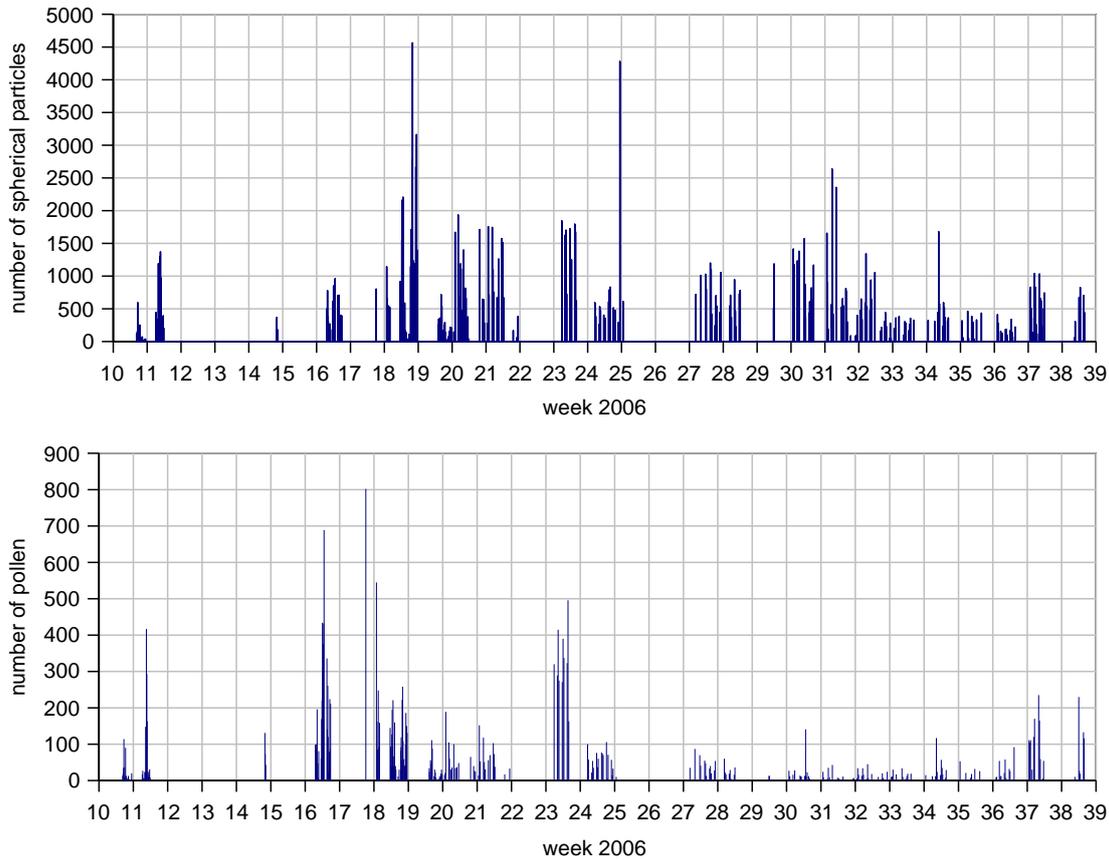


Figure 3.16: Number of spherical particles, and number of pollen grains (manually labeled) that were found in the air samples of the pollenmonitor during 2006

other *Alnus*^(A.2) species, only due to its significantly later flowering period. All pollen that could not be identified unequivocally were labeled as “indeterm.”

The majority of the non-spherical pollen grains (*Pinus*^(D.10) and *Picea*^(D.9)) were not correctly segmented. The fragments of these pollen were labeled as “non-pollen”. Only some of them were spherical enough to be correctly segmented. They were included into the database, even though they do not represent the true characteristics of this pollen type. The total number of pollen grains for each taxon are listed in figure 3.17.

The co-occurrence of the different pollen taxa over the year is shown in figure 3.18

3.2.6 Discussion

Too low Number of Pollen Grains for Several Taxa

Even though the total number of particles and pollen grains in this data set is quite impressive, there are still many pollen taxa where the number of obtained pollen grains during this year is much too low to express all important variants. For the highly allergenic pollen taxa the main reasons for low numbers is the location. E. g., the highly allergenic *Secale*^(A.7) (engl. “rye”, germ. “Roggen”) pollen, are not present in Freiburg, because there are no rye producing areas nearby. The relatively low numbers for *Alnus*^(A.2) and *Corylus*^(A.5) are caused by the late start of the measurements in March, when their main flowering period was already over. So it is relatively easy to obtain large numbers of pollen grains of these taxa just by running the pollenmonitor at the appropriate place in the appropriate season. For

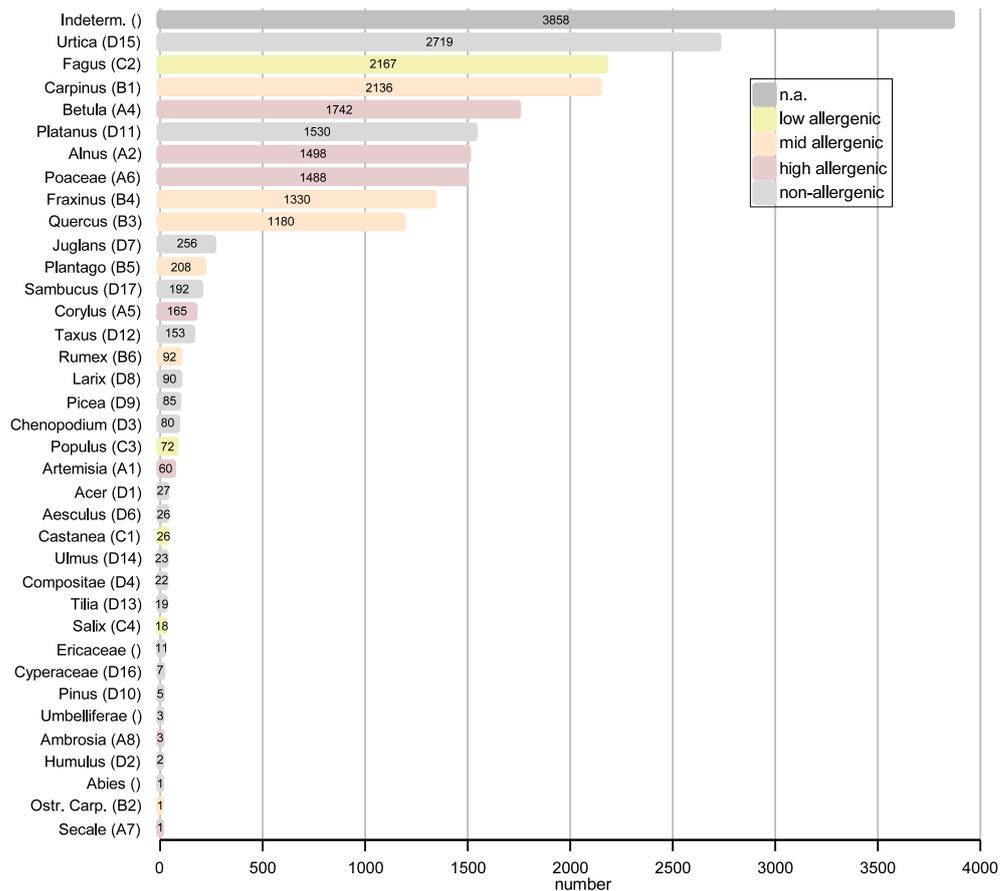


Figure 3.17: Number of pollen grains for each taxon in the pollenmonitor dataset. The numbers for *Pinus*^(D.10) and *Picea*^(D.9) in this dataset do not represent the true number on the air samples (see text)

3 Data Sets

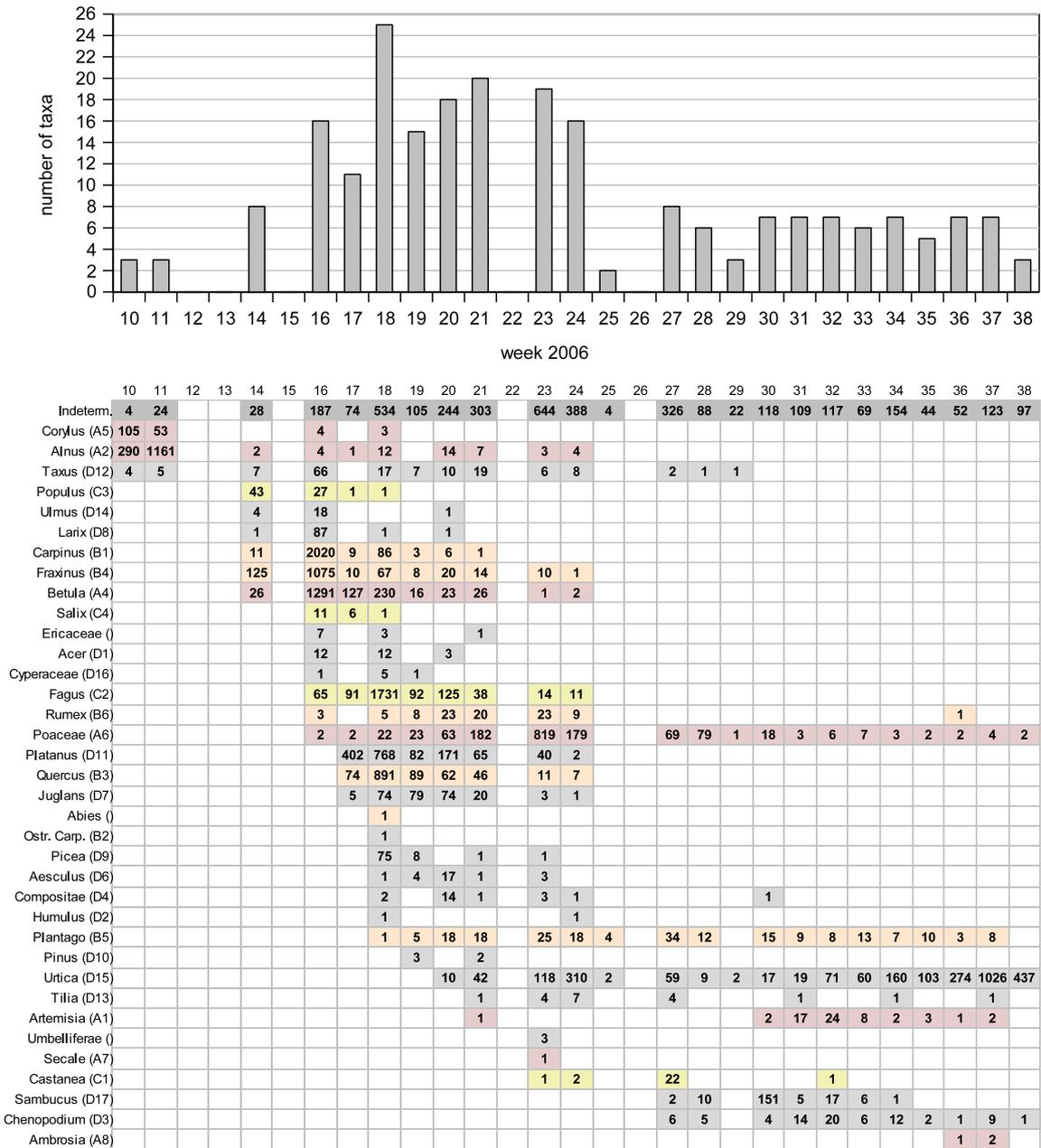


Figure 3.18: Upper figure: The number of different pollen taxa that are found on the samples within each week. Lower figure: Pollen taxa in the pollenmonitor data set, sorted by their appearance in the year. The colors show their allergenic relevance, red: highly allergenic, orange: moderately allergenic; yellow: lowly allergenic, gray: non-allergenic. One can clearly see the main flowering period.

several other pollen taxa the low number really represents their typical concentrations in the air. For those taxa it might be difficult to collect a sufficient number of training pollen with these techniques.

Low Quality of Fluorescence Recordings

There are some other avoidable shortcomings of this data set, which have a significant impact on the final recognition results. The most important one is the very low image quality of the fluorescence image stacks, compared to the resulting images taken with a standard fluorescence microscope from Zeiss, due to some misfortune with the responsible partner. Even though the fluorescence allows to extract significant additional information from the pollen grains, the better choice was to disregard the recorded fluorescence stacks from the pollenmonitor and to use only the stacks from transmitted light microscopy.

Vibrations During Stack Recording

Another problem in this data set are the vibrations that occurred occasionally during the stack recordings. The main sources for this vibration, the pump and the air conditioning, were mechanically decoupled from the system, but other sources (wind, vibrations of the stepper motors themselves) could not be avoided. Here an appropriate re-alignment of the recorded image stacks by a cross correlation analysis of the successive layers was at least able to compensate the displacements in xy-direction.

Non-Uniformity in the Illumination

The strong non-uniformities of the illumination in the transmitted light microscopy can be corrected by an appropriate shading correction (see section 2.5.2), or by the use of features that are invariant to gray scale changes.

Mis-adjusted Optics in Zürich

All stacks that were recorded in Zürich have significantly different image characteristics than those from Freiburg due to mis-adjusted optics (see figure 3.19)

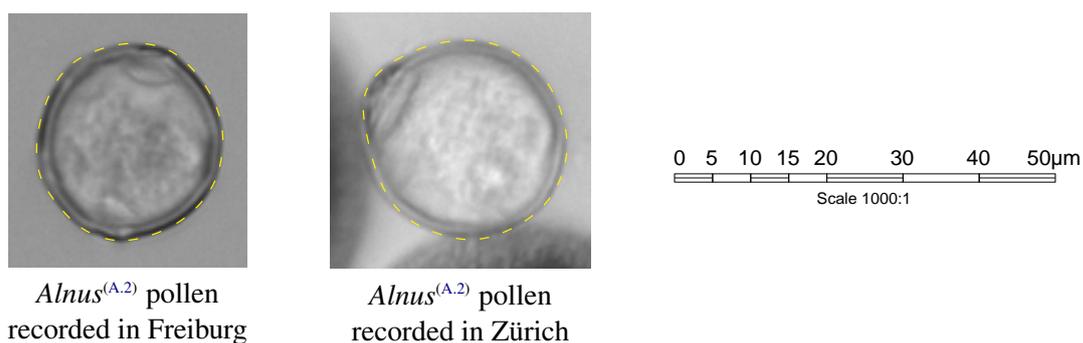


Figure 3.19: Effects of the mis-adjusted optics in Zürich. The images are brighter and show a lower richness of details

To compensate mechanical problems the objective was slightly displaced towards the sample by the addition of a small shim, while the position of the condenser was not accordingly corrected. This resulted in much brighter images with lower contrast and lower richness of details.

Imperfect Embedding of Some Pollen Grains

Another effect that still occurs in this data set is an imperfect embedding for some pollen grains (see figure 3.20 for an example).

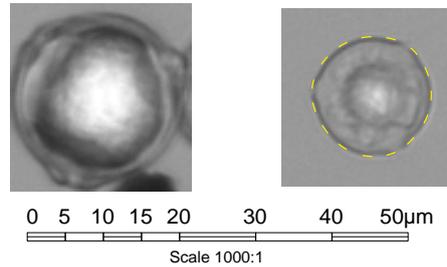


Figure 3.20: Imperfectly embedded *Betula*^(A.4) and *Urtica*^(D.15) pollen grains, resulting in a white spot in the center of the object

While, in the year before, a large fraction of the pollen grains was not fully embedded in the embedding medium, this effect was significantly reduced by a changed composition of the embedding medium and an adapted temperature curve for the preparation process. Nevertheless there are still some pollen grains that did not fully re-hydrate. These dry regions within the pollen grain produce strong reflections that make an identification very difficult even for humans experts.

Avoidable Failures of the Pollenmonitor

During the operation of the prototype of the pollenmonitor several failures of the machine occurred. In most cases these failures have led to strong distortions such that the recorded image stacks were not useful for further evaluations. E. g., a loose contact in the transmitted light illumination caused a randomly distributed dark frames in the image stacks. Another example is the failure of the integrated air conditioning during the preparation and recording of the air sample from June, 18th 16.00h. This resulted in an extreme number of tiny water droplets on the sample (see figure 3.21). These water droplets have the same size and gray value distribution as pollen grains but no corresponding droplets exist in the training data set, which would have led to biased results in some experiments (especially in those where the classifier is forced to assign each particle to a known class). In a final pollenmonitor it is easy to integrate a sensor, that can detect such a failure. As we are interested in the expected performance of the machine during normal operation, this sample was left out in most experiments.

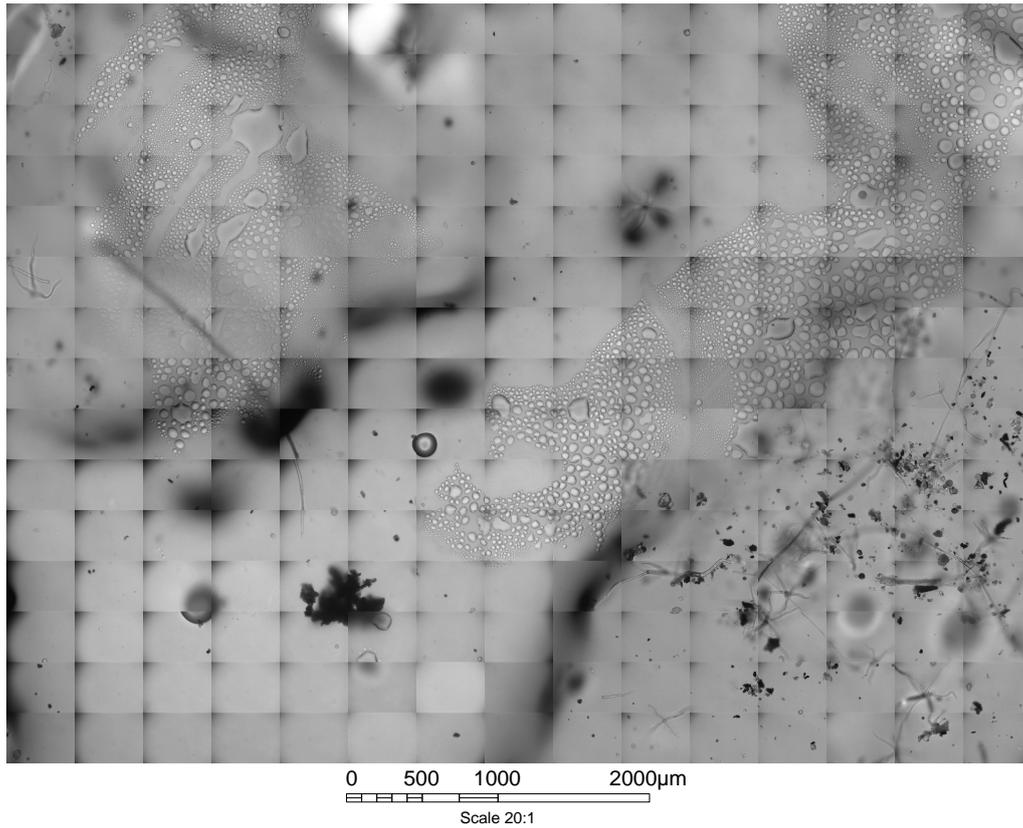


Figure 3.21: Due to a failure of the air conditioning on June, 18th thousands of tiny water droplets cover the air sample.

4 The Haar Integration (HI) Framework

The Haar-Integration framework, introduced by [Schulz-Mirbach \(1995b\)](#) offers a simple but very powerful way of a general feature extraction.

The goal of such a feature extraction is to find a representation of an object that is invariant to a given set of transformations, e. g., the Euclidean motions: The set of all possible 3D volume data sets of one individual pollen grain – scanned in all possible positions and orientations (Euclidean motion) – is an equivalence class. An invariant transformation is able to map all elements of this equivalence class into one point of the feature space and there represents one piece of information on the intrinsic structure of the considered pollen grain, independent of its position and orientation (Figure 4.1).

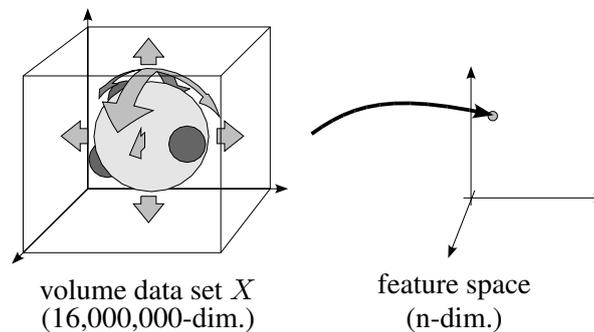


Figure 4.1: Invariant transformation: All representations of an object (here: the object at any orientation or position) are mapped into the same point in the feature space.

4.1 Invariance by Integration over the Transformation Group

The basic recipe to compute these features is to take a non-linear kernel function $f(X)$ in order to relate or combine the gray values of some neighboring pixels or voxels and to integrate the result of this function over all possible representations of the object in the equivalence class ([Schulz-Mirbach, 1995a](#)).

$$T[f](X) := \int_G f(gX) dg \quad (4.1)$$

- G : transformation group
- g : one element of the transformation group
- f : Kernel function
- X : image or volume data
- gX : transformed image or volume data
- dg : (normalized) Haar measure

An illustrative example for the computation of such an invariant is shown in figure 4.2.

$$\begin{aligned}
 \text{a) } T \left(\begin{array}{c} \text{☺} \\ \text{☺} \\ \text{☺} \\ \vdots \\ \text{☺} \end{array} \right) &= \frac{1}{N} \left(f \left(\begin{array}{c} \text{☺} \\ \text{☺} \\ \text{☺} \\ \vdots \\ \text{☺} \end{array} \right) + f \left(\begin{array}{c} \text{☺} \\ \text{☺} \\ \text{☺} \\ \vdots \\ \text{☺} \end{array} \right) + f \left(\begin{array}{c} \text{☺} \\ \text{☺} \\ \text{☺} \\ \vdots \\ \text{☺} \end{array} \right) + \dots + f \left(\begin{array}{c} \text{☺} \\ \text{☺} \\ \text{☺} \\ \vdots \\ \text{☺} \end{array} \right) \right) = 42 \\
 \text{b) } T \left(\begin{array}{c} \text{☹} \\ \text{☹} \\ \text{☹} \\ \vdots \\ \text{☹} \end{array} \right) &= \frac{1}{N} \left(f \left(\begin{array}{c} \text{☹} \\ \text{☹} \\ \text{☹} \\ \vdots \\ \text{☹} \end{array} \right) + f \left(\begin{array}{c} \text{☹} \\ \text{☹} \\ \text{☹} \\ \vdots \\ \text{☹} \end{array} \right) + f \left(\begin{array}{c} \text{☹} \\ \text{☹} \\ \text{☹} \\ \vdots \\ \text{☹} \end{array} \right) + \dots + f \left(\begin{array}{c} \text{☹} \\ \text{☹} \\ \text{☹} \\ \vdots \\ \text{☹} \end{array} \right) \right) = 42
 \end{aligned}$$

Figure 4.2: Simple example for the computation of invariants using the Haar integration framework for the group of rotations on a very friendly pollen species (not yet discovered in the natural environment): $f(\cdot)$ is an arbitrary non-linear function, that computes a scalar value from the given image. N is the number of selected rotation angles. The resulting invariant $T = 42$ is identical for a) and b), independent of the orientation of the test image. Proof: rearranging of the summands.

4.2 Direct Computation Using General N-Point-Kernel Functions

If the kernel function f only depends on a few gray values of the image or volume, i. e., if we can rewrite $f(X)$ as $f(X(\mathbf{x}_1), X(\mathbf{x}_2), \dots, X(\mathbf{x}_k))$ we only need to transform the kernel points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ accordingly instead of the whole data set X (Schulz-Mirbach, 1995a). This transformation of the kernel points will be denoted here as $\mathbf{s}_g(\mathbf{x}_i)$ such that

$$(gX)(\mathbf{x}_i) = X(\mathbf{s}_g(\mathbf{x}_i)) \quad \forall g, \mathbf{x}_i \quad (4.2)$$

With this we can rewrite eq. 4.1 as

$$T[f](X) := \int_G f \left(X(\mathbf{s}_g(\mathbf{x}_1)), X(\mathbf{s}_g(\mathbf{x}_2)), \dots, X(\mathbf{s}_g(\mathbf{x}_k)) \right) dg \quad (4.3)$$

This considerably speeds up the computation, because instead of all pixels/voxels of the image only k points must be transformed for each element of the transformation group.

Furthermore the number of elements of the transformation group, that must be evaluated to correctly compute the integral on discretized data (see next section “correct sampling of invariants”) can also be reduced by the use of sparse kernels. E. g., for the group of Euclidean motions on a 2D image with $N \times N$ pixels, we need $O(N^2)$ translations and $O(N)$ rotations, which results in $O(N^3)$ operations for each pixel. The evaluation of the Haar integral then takes $O(N^5)$ operations just for transforming all pixels according to all transformations.

Using a sparse kernel function we need a fixed number of rotations depending only on the “radius” of the kernel function q , and (as mentioned above) only k points instead of N^2 pixels need to be transformed. This results in a linear complexity $O(k \cdot q \cdot N^2)$ in the number of pixels N^2 of the image (see figure 4.3 for an example of a kernel function that depends only on two points and the Euclidean transformation group).

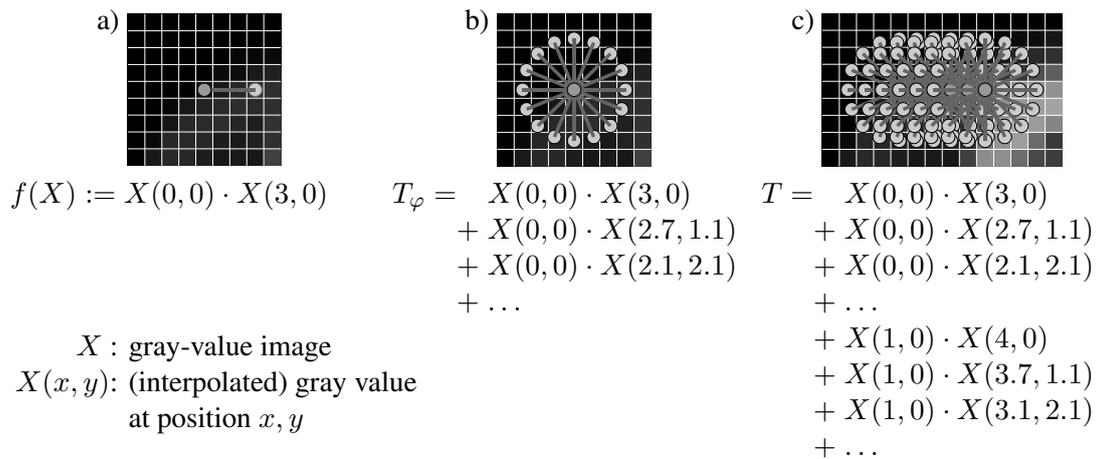


Figure 4.3: Computation of an invariant using a two-point-kernel on a 2D image: (a) Selection of a non-linear kernel function for combining some neighboring pixels: In this example the kernel function $f(X)$ is defined as the multiplication of two gray values at distance 3. (b) This kernel function is evaluated for all angles and the results are summed up, to become invariant to rotations of the object. (c) This set of rotated kernel functions is evaluated at all possible positions of the image and the results are summed up, to become invariant to translations of the object. As a result, identical values of T are obtained independently of the angle and position of the object in the image.

5 Analysis of the Properties of the HI framework

5.1 Correct Sampling of the Transformation Group for the Computation of Invariants

As soon as we want to compute the invariants with the technique described in section 4.2, the question arises, how many sampling steps (angular and translational) are needed for an unbiased result.

While in earlier studies on these invariants the required sampling steps were just heuristically determined (Schulz-Mirbach, 1995b; Siggelkow, 2002; Schael, 2004), for a fast but unbiased computation it is desirable to know the minimal number.

One advantage of these invariants compared to other feature extraction methods is their exact formulation in the continuous space and therefore the independence of the selected parameterization. So for the final computation of the invariants on discretely sampled sensor data, we can use the sampling theorem (Shannon, 1949) to ensure, that the computed features are invariant to any translation and rotation and not only to integer shifts or rotations in 90 degree steps.

To compute the invariants from eq. 4.3 we firstly select a parameterization λ of g so that we can rewrite eq. 4.3 as

$$T[f](X) := \int f\left(X(\mathbf{s}_\lambda(\mathbf{x}_1)), X(\mathbf{s}_\lambda(\mathbf{x}_2)), \dots, X(\mathbf{s}_\lambda(\mathbf{x}_k))\right) d\lambda \quad (5.1)$$

For a given data set X and given kernel points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ we can substitute $X(\mathbf{s}_\lambda(\mathbf{x}_i))$ (which are the gray values, that are touched by the i 'th kernel point \mathbf{x}_i during all transformations described by λ) with $v_i(\lambda)$, resulting in

$$T = \int f\left(v_1(\lambda), v_2(\lambda), \dots, v_k(\lambda)\right) d\lambda \quad (5.2)$$

A simple example for the resulting one-dimensional curves v_1, v_2 and v_3 when using a three-point-kernel on a 2D image and only the transformation group of rotations is given in figure 5.1.

For volumetric data and Euclidean transformations the transformation can be parameterized as $\lambda = (t_1, t_2, t_3, \varphi_1, \varphi_2, \varphi_3)^T$ (see fig 5.2).

In this case $v_i(\lambda)$ describes for each kernel point i a 6-dimensional continuous scalar field. The required sampling intervals can be computed from the positions of the kernel points and the required sampling intervals of the underlying volumetric data set X as follows:

If X is bandwidth limited along the q -axis (with $q \in \{1, 2, \dots, 6\}$) with the Nyquist critical frequency ω_{c, X_q} the sampling interval must be

$$\Delta X_q = \frac{1}{2\omega_{c, X_q}} \quad (5.3)$$

or smaller. For the parameterization $\lambda = (t_1, t_2, t_3, \varphi_1, \varphi_2, \varphi_3)^T$, mentioned above, we find the necessary sampling intervals for each component of λ and each kernel point \mathbf{x}_i by

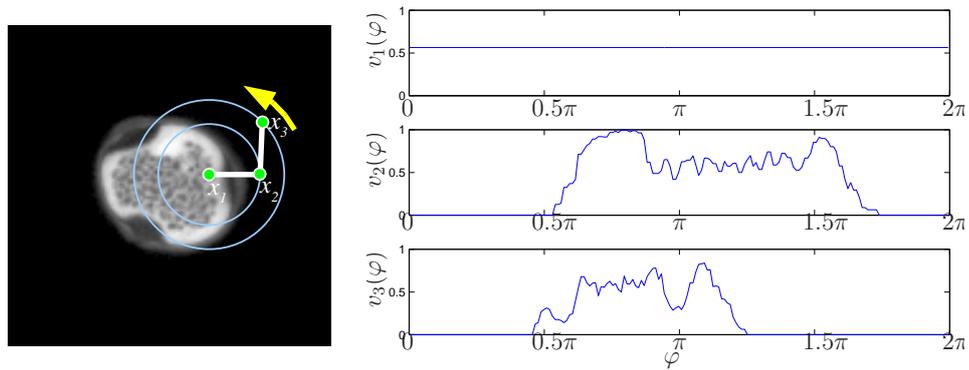


Figure 5.1: When using a three-point kernel on a continuous 2D image and only the transformation group of rotations, the gray values, that are touched by the kernel points \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 , are one-dimensional functions $v_1(\varphi)$, $v_2(\varphi)$ and $v_3(\varphi)$ of the rotation angle. The required number of angular sampling steps can be computed from the spectra of these functions

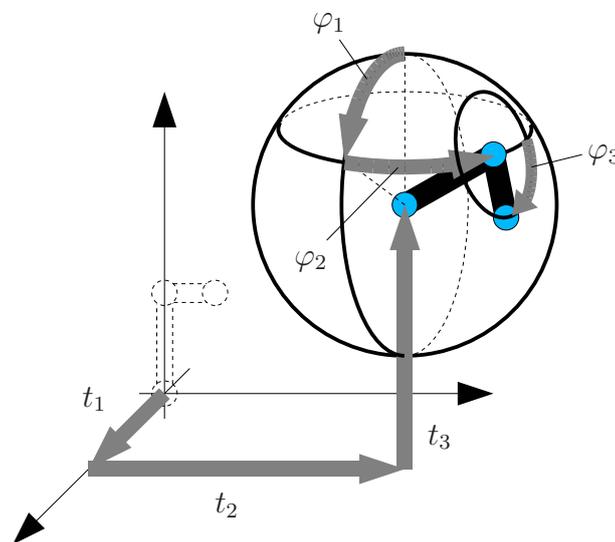


Figure 5.2: Parameterization of Euclidean transformation with $\lambda = (t_1, t_2, t_3, \varphi_1, \varphi_2, \varphi_3)^T$

$$\begin{aligned}
 \Delta t_1 &= \Delta X_1 \\
 \Delta t_2 &= \Delta X_2 \\
 \Delta t_3 &= \Delta X_3 \\
 \Delta \varphi_1 &= \arcsin \frac{\min(\Delta X_1, \Delta X_2, \Delta X_3)}{\|\mathbf{x}_i\|} \\
 \Delta \varphi_2 &= \arcsin \frac{\min(\Delta X_1, \Delta X_2, \Delta X_3)}{\|\mathbf{x}_i\| \cdot \sin(\varphi_1)} \\
 \Delta \varphi_3 &= \arcsin \frac{\min(\Delta X_1, \Delta X_2, \Delta X_3)}{\sqrt{x_{i,1}^2 + x_{i,2}^2}} .
 \end{aligned} \tag{5.4}$$

Now we know the required sampling steps for each $v_i(\boldsymbol{\lambda})$, but for the final integration we need the sampling step for the nonlinear “element-wise” combination of these v_i ’s, which we call $Q(\boldsymbol{\lambda})$,

$$Q(\boldsymbol{\lambda}) := f(v_1(\boldsymbol{\lambda}), v_2(\boldsymbol{\lambda}), \dots, v_k(\boldsymbol{\lambda})) . \tag{5.5}$$

Due to the nonlinear combination, the maximum frequencies in the spectrum of the resulting $Q(\boldsymbol{\lambda})$ may be much higher than in the v_i ’s (see figure 5.3)

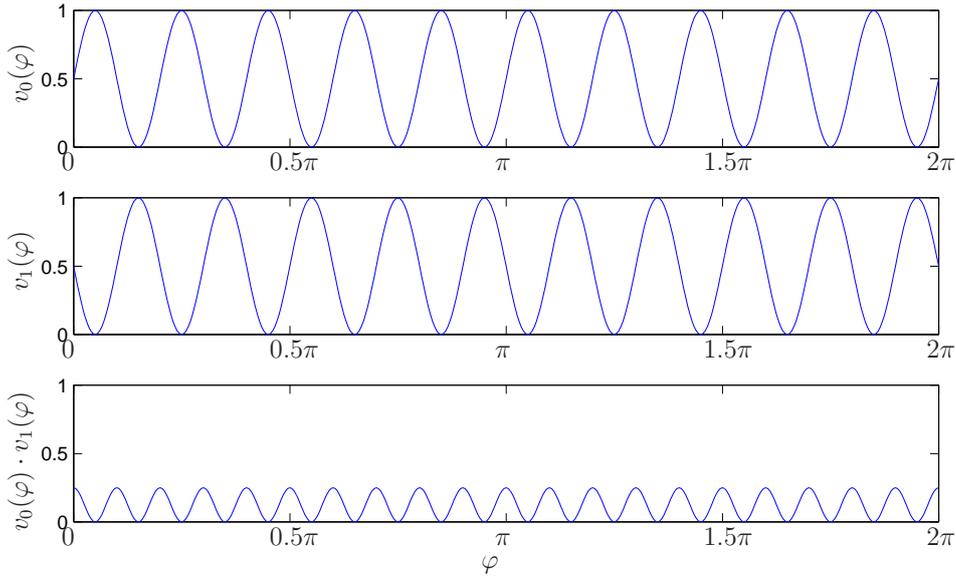


Figure 5.3: Example for nonlinear combination of two v_i ’s. In this case a simple multiplication is done, which corresponds to a kernel function of the type $f(X) = X(\mathbf{x}_1) \cdot X(\mathbf{x}_2)$. The two input functions are band-limited with a maximum frequency of 10, while the nonlinear combination is band-limited with a maximum frequency of 20 and therefore needs twice the number of sampling points to fulfill the sampling theorem

There is no general way to describe the effects in the frequency domain, when applying non-linear operations in the spatial domain. But if we can rewrite or approximate the non-linear function with a polynomial (e. g. by Taylor series expansion and truncation of higher terms), we only need element-wise additions and element-wise multiplications, whose spectral effects can be described with the following rules:

With $\omega_{c,1}$ being the Nyquist critical frequencies of the data set $v_1(\boldsymbol{\lambda})$, where each component of $\omega_{c,1}$ corresponds to one component of the parameterization $\boldsymbol{\lambda}$ we get

$$\begin{aligned}
 v_{\text{res}}(\boldsymbol{\lambda}) &= C + v_1(\boldsymbol{\lambda}) &\Rightarrow \omega_{c,\text{res}} &= \omega_{c,1} &\quad \forall C \in \mathbb{R} \\
 v_{\text{res}}(\boldsymbol{\lambda}) &= C \cdot v_1(\boldsymbol{\lambda}) &\Rightarrow \omega_{c,\text{res}} &= \omega_{c,1} &\quad \forall C \in \mathbb{R} \\
 v_{\text{res}}(\boldsymbol{\lambda}) &= v_1(\boldsymbol{\lambda}) + v_2(\boldsymbol{\lambda}) &\Rightarrow \omega_{c,\text{res}} &= \max(\omega_{c,1}, \omega_{c,2}) \\
 v_{\text{res}}(\boldsymbol{\lambda}) &= v_1(\boldsymbol{\lambda}) \cdot v_2(\boldsymbol{\lambda}) &\Rightarrow \omega_{c,\text{res}} &= \omega_{c,1} + \omega_{c,2}
 \end{aligned} \tag{5.6}$$

where $\max()$ denotes the element-wise maximum. The proof of the first three equations is simple, for the last equation we can use the ‘‘Convolution Theorem’’ (e. g., (Bracewell, 1999)), which allows to interpret the multiplication in the spatial domain as a convolution in the frequency domain:

$$v_{\text{res}}(\boldsymbol{\lambda}) = v_1(\boldsymbol{\lambda}) \cdot v_2(\boldsymbol{\lambda}) \iff \mathcal{F}\{v_{\text{res}}\} = \mathcal{F}\{v_1\} * \mathcal{F}\{v_2\} \tag{5.7}$$

In that convolution it is obvious, that the resulting maximum frequency is the sum of the maximum frequencies of each source function.

The most interesting consequence of this result is, that the sampling interval in t_1, t_2, t_3 direction must be smaller than 1 pixel for nearly all kernel functions, e. g., a monomial of degree 3 like $f(X) = X(0, 0, 0) \cdot X(3, 0, 0) \cdot X(3, 0, 1)$ must be sampled in t_1, t_2, t_3 direction with an interval of $\frac{1}{3}$ pixel to fulfill the sampling theorem!

Of course, this worst-case-estimation of the resulting spectrum by summing the Nyquist critical frequencies for a multiplication in the spatial domain only applies to a spectrum that has a significant part of its energy in the highest frequencies. For a more realistic estimation, we can model the spectra as Gaussian distributions. A convolution of two Gaussian distributions is again a Gaussian distribution with a variance that is the sum of the variances from the source distributions, $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Therefore we can estimate the Nyquist critical frequencies of the resulting signal by

$$v_{\text{res}}(\boldsymbol{\lambda}) = v_1(\boldsymbol{\lambda}) \cdot v_2(\boldsymbol{\lambda}) \Rightarrow \omega_{\sigma,\text{res}} = \sqrt{\omega_{\sigma,1}^2 + \omega_{\sigma,2}^2} \tag{5.8}$$

Here again, the operations on ω are element-wise operations. For the above example this will result in a sampling interval of $\frac{1}{\sqrt{3}} \approx 0.577$ pixel.

5.1.1 Consequences of Under-Sampling

In this section the consequences of under-sampling will be analyzed. Such an under-sampling may be chosen to save computation time.

We assume a continuous signal that is not bandwidth limited to less than the Nyquist critical frequency ω_c . When sampling such a signal any frequency component ω_1 outside the range $(-\omega_c, \omega_c)$ is aliased to that frequency ω_2 into that range, which differs from ω_1 by a multiple of $1/\Delta = 2\omega_c$.

For the Haar integration we are not interested in reconstructing the signal but only in the integral of the signal. Therefore the only relevant frequency component is the zero frequency component¹. All the others do not contribute to the integral and may be disturbed by aliased frequency components (see figure 5.4).

The first frequency outside the Nyquist range, that contributes falsely to the zeroth frequency is $\omega = 2\omega_c$. From that we can draw the following conclusions:

1. **As long as the highest frequency of a signal Q is less than twice the Nyquist critical frequency ω_c , the integral of the reconstructed signal will still be exact**

¹Thanks to Lokesh Setia for this hint

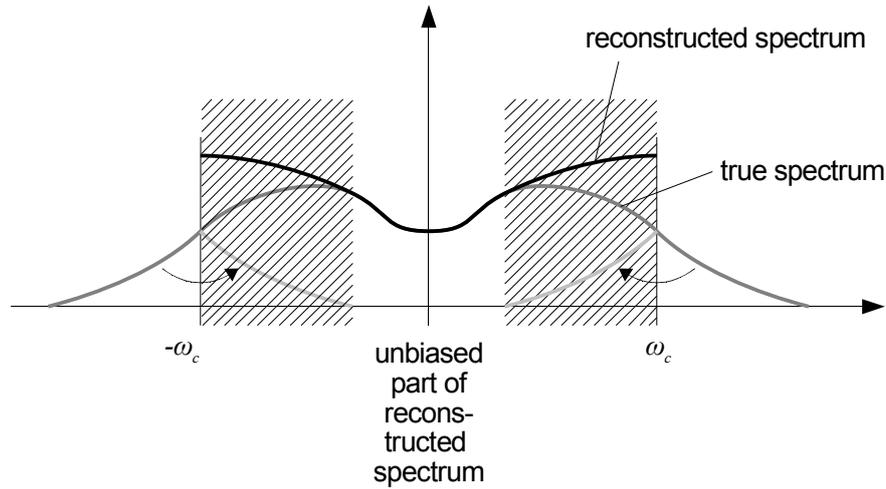


Figure 5.4: When violating the sampling theorem only a part of the reconstructed spectrum is biased.

2. **If the highest frequency of the signal Q is equal or higher than twice the Nyquist critical frequency ω_c , the error of the integral is**

$$Err = \sum_{n \in \mathbb{Z}} \text{Re}(\mathcal{F}\{Q\}(n \cdot 2\omega_c)) \quad (5.9)$$

5.2 Maximal Information that can be Sensed with Two-Point-Kernels and the Euclidean Transformation Group

Kernel functions, that depend only on two gray values will be called “Two-Point-Kernels” in the following.

$$f(X) = f(X(\mathbf{0}), X(\mathbf{q})) \quad (5.10)$$

An example for the direct computation of invariants using a two-point-kernel in 2D is shown in figure 4.3

The big advantage of this class of kernel functions is, that they can be computed efficiently via the Fast Fourier Transformation (see section 7.4) and are therefore well suited for large 3D datasets like confocal recordings of pollen. Due to the arbitrary function that combines these two gray values they are still quite powerful, which will be shown in the following sections.

To analyze the maximal information that can be sensed with any two-point-kernel function, the Haar integration can be split into two parts: The first part just describes the effects of the integration over all angles and all positions (Euclidean motion) and the second part describes the effects of the selected kernel function.

5.2.1 The Gray Value Co-Occurrence Distribution “GVCD”

Each kernel function of type $f(X) = f(X(\mathbf{0}), X(\mathbf{q}))$ can be rewritten as an integral over delta distributions, where each of them only “senses” exactly one gray value,

$$f(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(v_1, v_2) \cdot \delta(v_1 - X(\mathbf{0})) \cdot \delta(v_2 - X(\mathbf{q})) \, dv_2 dv_1 \quad (5.11)$$

When we insert this kernel function into the rewritten Haar integral from eq. 4.3 we can move the kernel function outside of the group integral,

$$\begin{aligned}
 T[f](X) &= \int_G \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(v_1, v_2) \cdot \delta\left(v_1 - X\left(\mathbf{s}_g(\mathbf{0})\right)\right) \cdot \delta\left(v_2 - X\left(\mathbf{s}_g(\mathbf{q})\right)\right) dv_2 dv_1 dg \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(v_1, v_2) \cdot \underbrace{\int_G \delta\left(v_1 - X\left(\mathbf{s}_g(\mathbf{0})\right)\right) \cdot \delta\left(v_2 - X\left(\mathbf{s}_g(\mathbf{q})\right)\right) dg}_{=: \text{gray value co-occurrence distribution "GVCD"}} dv_2 dv_1 ,
 \end{aligned}
 \tag{5.12}$$

and define a gray value co-occurrence distribution “GVCD”. Now, the action of a specific kernel function can be described by the dot product of the kernel function with this GVCD in the Hilbert space,

$$T[f](X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(v_1, v_2) \cdot \text{GVCD}(\mathbf{q}, v_1, v_2) dv_2 dv_1 .
 \tag{5.13}$$

Due to the integration over all rotations within the Haar integral, the orientation of \mathbf{q} does not influence the final feature. So all information that can be extracted by invariants with two-point-kernels, is contained in the 3D gray value cooccurrence distribution (“GVCD”):

$$\text{GVCD}(q, v_1, v_2) = \int_G \delta\left(v_1 - X\left(\mathbf{s}_g(\mathbf{0})\right)\right) \cdot \delta\left(v_2 - X\left(\mathbf{s}_g(\mathbf{q})\right)\right) dg \quad \text{with } q = \|\mathbf{q}\|
 \tag{5.14}$$

A visualization, how the GVCD works is shown in fig 5.5.

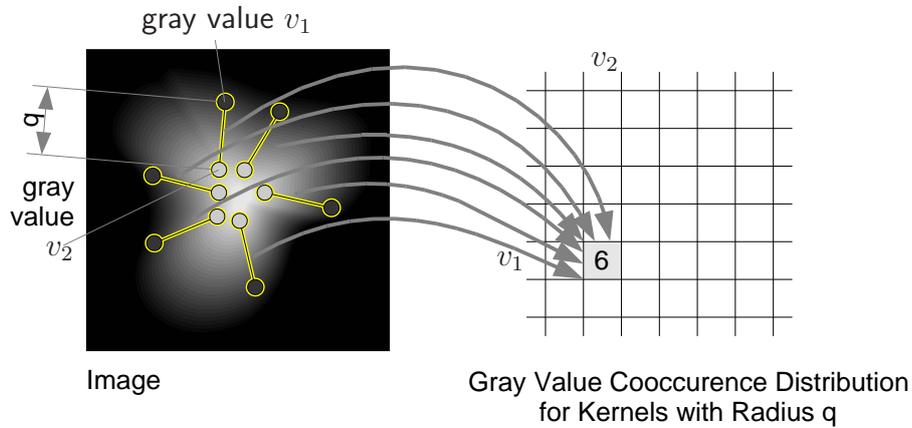


Figure 5.5: The gray value co-occurrence distribution (for a given kernel span q and integration over the group of Euclidean motions) “counts” how often each combination of gray values appears in the image

The quantized version without rotational invariance of the GVCD is well known as GLCM (“Gray Level Co-occurrence Matrix”) and was introduced by Haralick et al. (1973) to compute textural features. The traditional GLCM of Haralick only uses a fixed offset between the two pixels and is therefore only invariant to integer shifts of the image. The introduced GVCD is invariant to any fractional shifts of the object and invariant to rotations (as long as the sampling theorem is fulfilled).

An illustration, how the invariants are computed from an extracted GVCD, is given in figure 5.6 Some additional plots of other typical kernel functions are shown in figure 5.7.

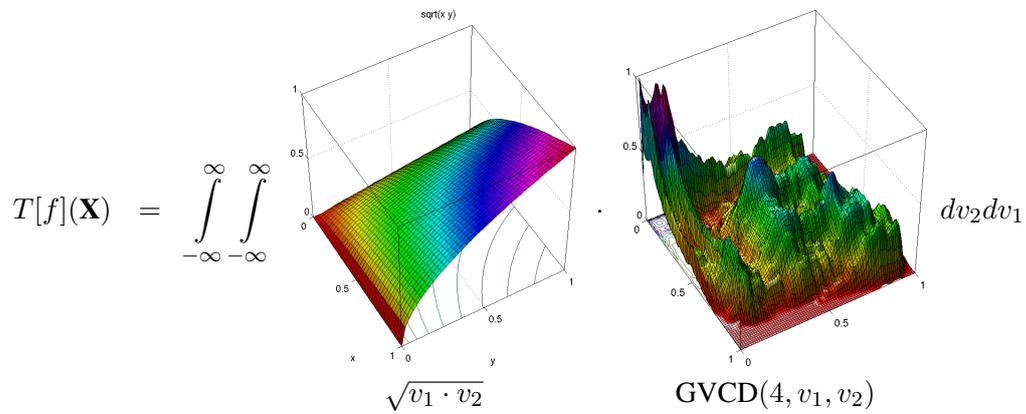


Figure 5.6: Example for the computation of an invariant with a two-point-kernel via the GVCD. The selected kernel function is $f(X) = \sqrt{X(0, 0, 0) \cdot X(0, 0, 4)}$

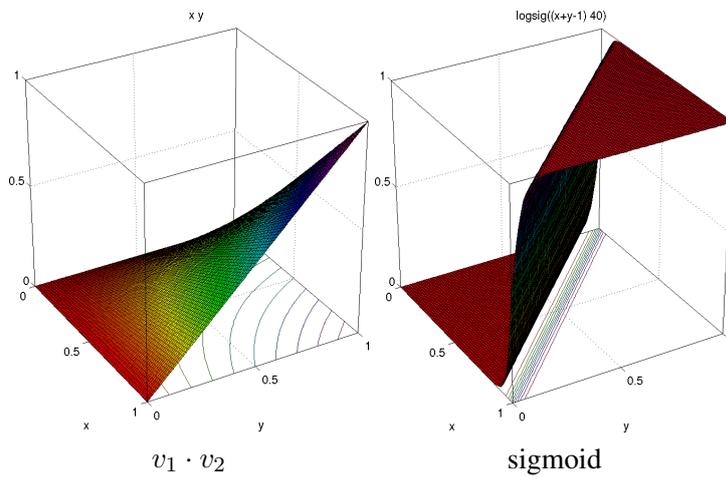


Figure 5.7: Further typical two-point-kernel functions

5.2.2 Using the GVCD to Estimate the Robustness of the Invariants

The splitting of the Haar integral into the GVCD and the kernel function provides a new view and a more intuitive understanding of the processes that happen behind the scenes during the computation of features with the HI framework. E. g. we can see directly the robustness of the kernel functions to variations in the gray values: slight variations in the gray values cause slight shifts of the peaks in the GVCD in v_1 and v_2 direction. When using a “smooth” kernel function like $\sqrt{v_1 \cdot v_2}$ the contribution of each peak to the final integral will change only marginally. This means, the resulting feature will be quite robust to variations in the gray values.

In the case of a sigmoid kernel (as proposed by Schael (2004) for the classification of textures) we can see that the robustness of the resulting feature strongly depends on the distribution of the gray values in the image and correspondingly the location of the peaks in the GVCD. If significant peaks are located near the “step” in the kernel function, already slight variations in the gray values of the image will cause significant variations in the final features. If the majority of the peaks is far away from this “step” the final feature will be very robust to gray value variations.

5.2.3 Using the GVCD to Find Appropriate Kernel Functions

Another possible application of the GVCD is to find the best suited kernel functions for a given problem. By analyzing the GVCD’s for each training pattern one can directly see the inner-class variations and the intra-class differences. Basing on these insights it might be easier to select appropriate kernel functions.

Of course it should also be possible to automatically find the best suited kernel functions with such an analysis. A first feasibility study on this topic was done together with Thorsten Schmidt (Schmidt, 2004). A summary of the results is given in section 10.5.3.

5.2.4 Using the GVCD to Analyze the Completeness of the Invariants

The decomposition of the Haar integration into a dot product of the kernel function with the three-dimensional GVCD allows to answer the question, whether it is possible to reach completeness when using any number of nonlinear two-point-kernel functions of the type $f(X) = f(X(\mathbf{0}), X(\mathbf{q}))$.

A complete set of invariants uniquely describes any manifold in the pattern space, that is spanned by an image and the transformed versions of this image.

So we want to know, if there are two different images (that are not equivalent under the group of Euclidean motions) that have the identical GVCD and therefore have identical features for any possible two-point-kernel function.

To find such an example, we can limit ourselves to a binary image with a fixed number of non-zero pixels, which results in a GVCD, where the “planes” for all q are identical and only those who match a distance between to white pixels are different, e. g. in a two pixel example, the column for $(v_1, v_2) = (1, 1)$ has only one non-zero entry at $q = d$, where d is the distance between the two points, see figure 5.8

For more than two white pixels this column will just contain the distance histogram for all white-pixel distances within the image. The task can therefore be reduced to find two sets of points, that have an identical distance histogram, but that are not identical under the group of Euclidean motions. An example for such two images is shown in figure 5.9². The GVCD’s of these two images are still identical when the points are replaced by small circles and the drawn lines between the points become part of the binary image.

So the result is:

²Thanks to Bernard Haasdonk for finding this example

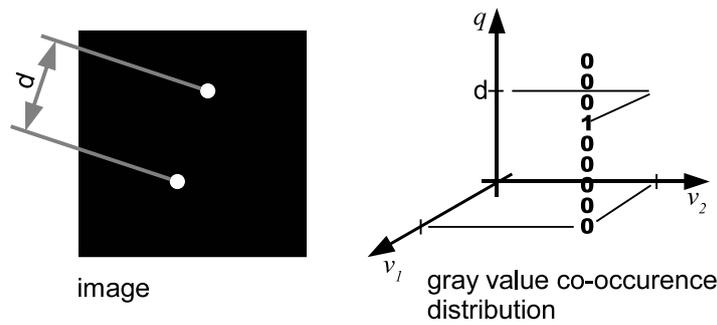


Figure 5.8: gray value co-occurrence distribution for a binary image with only two white pixels

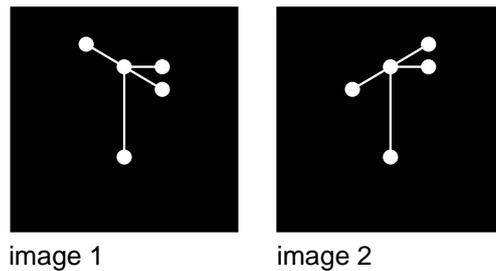


Figure 5.9: Images of two different objects, which have the identical GVCD

It is not possible to describe an image completely with any number of Euclidean invariants that only use two-point-kernels of type $f(X) = f(X(0), X(q))$

To overcome this problem but still benefit from the fast computation algorithms for two-point-kernels, we can use an appropriate pre-processing of the data set, e. g. the low-pass filtering introduced in section 6.4. For the demonstration here, we just take images from figure 5.9, which have an identical GVCD, and convolve them with a disk. For the sake of clearness, we reduce these images to their four points (leaving away the lines) before convolution. If the disk is big enough, such that the disks of two points overlap, the GVCD's of the two images become different (see figure 5.10)

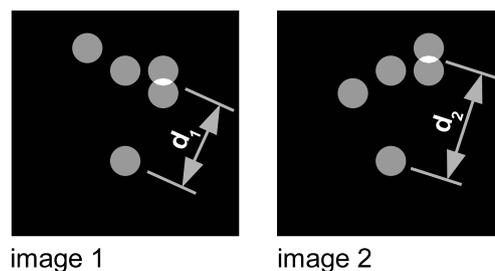


Figure 5.10: When convolving images that have identical GVCD's, with a rotational symmetric function (here a disc), their GVCD's will become unique as soon as the discs overlap

We did not manage to find an example of two different objects with identical GVCD's if this pre-processing step is included. This of course can not be taken as a proof that such objects do not exist, but at least it is a hint that this class of features has a high discrimination power.

6 Extensions of the HI framework

6.1 Constraints for Transformations, Kernel Functions and Data Sets

For the application of the HI framework for the construction of invariants some constraints have to be taken into account. The goal of this section is not to provide extensive analysis of the mathematical properties of Haar integrals. Instead we want to give a more intuitive understanding for its limitations and how to cope with them in the daily use.

Firstly we want to introduce the different parts, that have to operate together:

- **The pattern space** is a high-dimensional space. Each possible image is represented by one single point in this space. e. g., for images of size 100×100 this space is 10,000-dimensional.
- **The transformation group** describes the transformations of the images in this pattern space. Each element of a transformation group is an operator, that maps an image to its transformed version. E. g. if X is an image and g_1 is one element of the transformation group, then $g_1 X$ is the transformed image (see figure 6.1).
- **The kernel function** is a nonlinear function that extracts a scalar value t from a given image X as $t = f(X)$.

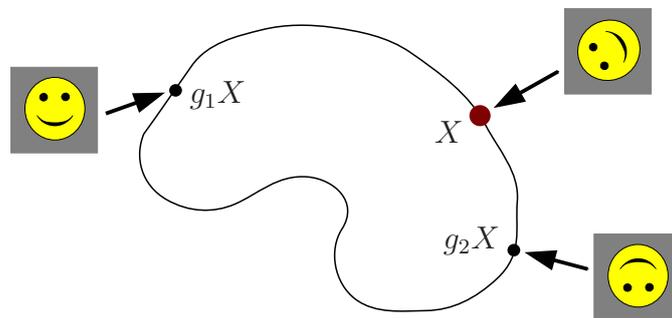


Figure 6.1: Illustration of the pattern space. Each image X corresponds to one point in the pattern space (e. g. for images of size 100×100 this pattern space is 10,000-dimensional). The recorded image X and the transformed versions of it, $g_i X$, are located in a manifold. In the case of a planar rotation this manifold is a one-dimensional trajectory that can be parameterized with the rotation angle φ

Depending on the required invariances in a given problem we have to select an appropriate transformation group.

6.2 Categorization of the Transformation Groups

In the earlier studies on the HI framework (Schulz-Mirbach, 1995b; Siggelkow, 2002; Schael, 2004) only cyclic transformations (rotations and cyclic shifts) were taken into account, which leads to well defined integration limits for the Haar integration (e. g., $[0 \dots 2\pi]$ for rotations or $[1 \dots N_x]$ and $[1 \dots N_y]$)

for the translations, where $N_x \times N_y$ is the size of the image). The Euclidean group was modeled there with a combination of a rotation and a cyclic shift, knowing well, that a rotation in such a cyclic rectangular environment is not well defined. With several additional constraints (e. g., the object must be on a black background, fully visible in the image, the black border around the objects must be larger than the support of the kernel functions), the authors excluded that the undefined regions influence the result.

A closer analysis of these constraints show that they just take care that none of the cyclic shift effects has an influence on the final features (or roughly spoken, only black background pixels are allowed to move from one border to the other). So the only remaining justification for the cyclic shifts is to define a region for the integration in the implementation.

Consequently we replace this combination of rotations and cyclic shifts with the real Euclidean group (where the translations must be integrated from $-\infty$ to $+\infty$) and implicitly limit the integral to a finite result with nearly the same conditions. This allows a clear distinction between the theoretically essential constraints and those which are just necessary for the selected implementation.

As a consequence we find two categories of transformation groups: case 1, if we can find a cyclic parameterization of the transformations, or case 2, if such a parameterization does not exist. These two cases are described in detail in the following sections.

6.2.1 Case 1: Group of Rotations or Cyclic Shifts

The best case for the application of the HI framework is a transformation group that can be described by a cyclic parameterization, like the group of rotations or cyclic shifts. In this case, we can use any reasonable kernel function, the start- and end-point of the integration in the pattern space are just identical, and this start-point can be selected arbitrarily on the manifold. E. g., for a planar rotation, parameterized by the rotation angle φ and the corresponding rotation operator g_φ , the Haar integral can be written as

$$T[f](X) = \int_{-\pi}^{\pi} f(g_\varphi X) d\varphi . \quad (6.1)$$

Alternatively we could set the limits to $[0..2\pi]$, etc. See figure 6.2 for an illustration. Independent of the kernel function, this integral is usually finite.

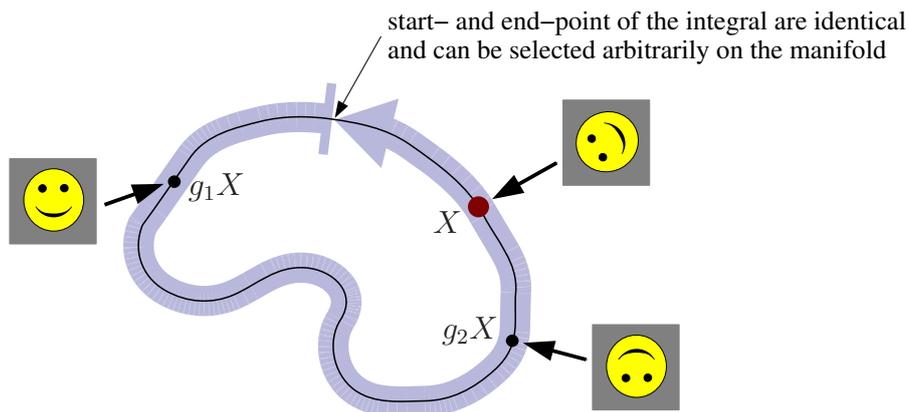


Figure 6.2: **Best case for Haar-integration, (e. g., group of rotations or cyclic translations):** The trajectory is closed, we can arbitrarily select the start- and end-point for the integration, and the integral is usually finite for any reasonable kernel function

6.2.2 Case 2: Group of Euclidean Transformations, Scalings and Similar

If no cyclic parameterization of the transformation group exists, the straight forward application of the HI framework will usually result in an infinite result. E. g., for the group of translations in x -direction, parameterized by the distance t_x and the translation operator g_{t_x} , the integration limits must be set to $[-\infty \dots \infty]$, resulting in

$$T[f](X) = \int_{-\infty}^{\infty} f(g_{t_x} X) dt_x . \quad (6.2)$$

To ensure a finite result, we have to introduce additional constraints to the kernel function and the image X .

6.3 Ensuring a Finite Result for Infinite Transformations

6.3.1 Implicit Limitation of the Integral by Appropriate Kernel Functions

The best possibility to avoid an infinite result, but to keep the full invariance properties is to limit the integral implicitly by pre-processing of the image and an appropriate kernel function. This approach was already used in (Schulz-Mirbach, 1995b) but with a slightly different motivation, to avoid that the cyclic boundary conditions destroy the invariance in Euclidean transformations.

Firstly we select a kernel function with local support. Local support means in this context that the input values for the kernel function are extracted from a small region, i. e. the kernel functions “see” only a small part of the image (see fig 6.3)

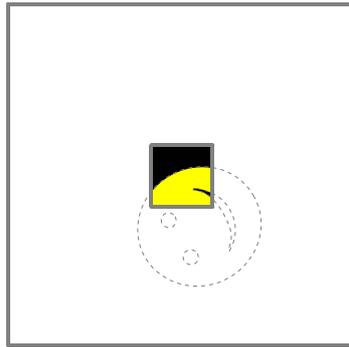


Figure 6.3: Illustration of a kernel function with local support. These kernel functions “look” only to a small part of the image

Now we can try to find a combination of such a kernel function f and a pre-processing of the image X , such that the kernel function returns non-zero values only within a certain area in the pattern space. E. g., this is fulfilled for a segmented object that is placed on a zero background combined with a kernel function that evaluates to zero if it finds only zero values within its “sensing area” (see figure 6.4).

The resulting implicit limits of the integration are bounded to the pattern space, which guarantees always the same result, independent of the initial position of the test object.

6.3.2 Explicit Limitation of the Integral by Partial Haar Integration

If no appropriate combination of kernel function and pre-processing can be found for an implicit limitation of the integral, we can limit the integral explicitly relative to the test pattern, but we will lose the invariance properties of the resulting features. Instead we will get some robustness to the considered transformations.

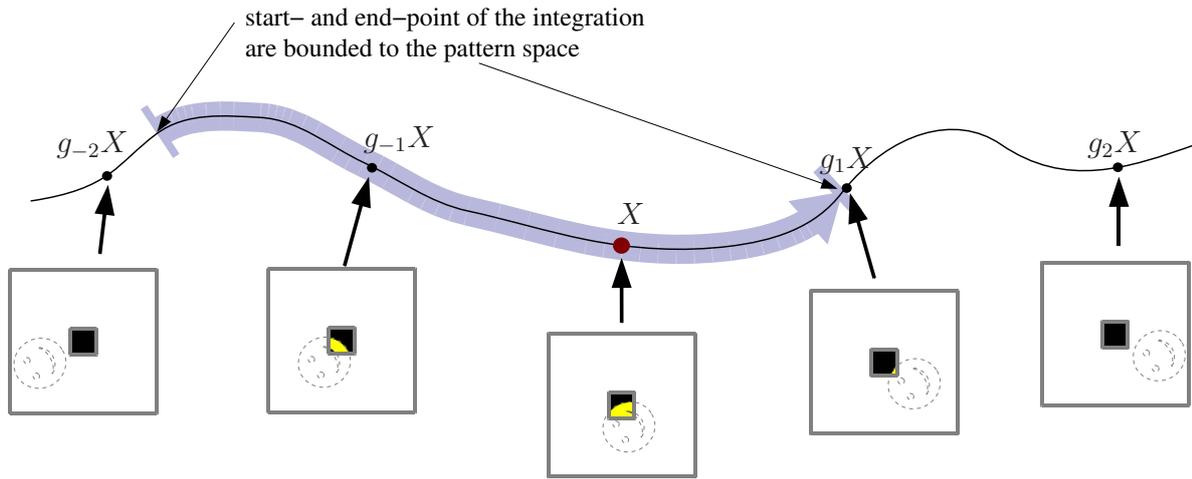


Figure 6.4: **Implicit limitation of the Haar integral.** For transformation groups like the group of translations, the trajectory is infinite and not closed. With an appropriate combination of a kernel function with local support and a pre-processing of the image, the start- and end-point of the integration are bounded to the pattern space, independent of the initial position of the test pattern.

For certain applications this non-invariance is exactly the desired behavior. E. g., in optical character recognition, features that are fully invariant to rotations will return identical values for “M” and “W”. For such applications [Haasdonk et al. \(2004\)](#) introduced such an explicit limitation of the integral as “partial Haar-integration features”, where only a subset of the transformation group is used. The resulting limits of the Haar integral are bounded to the test object. An example for small rotations is

$$T[f](X) = \int_{-0.1\pi}^{0.1\pi} f(g_\varphi X) d\varphi , \quad (6.3)$$

(see figure 6.5).

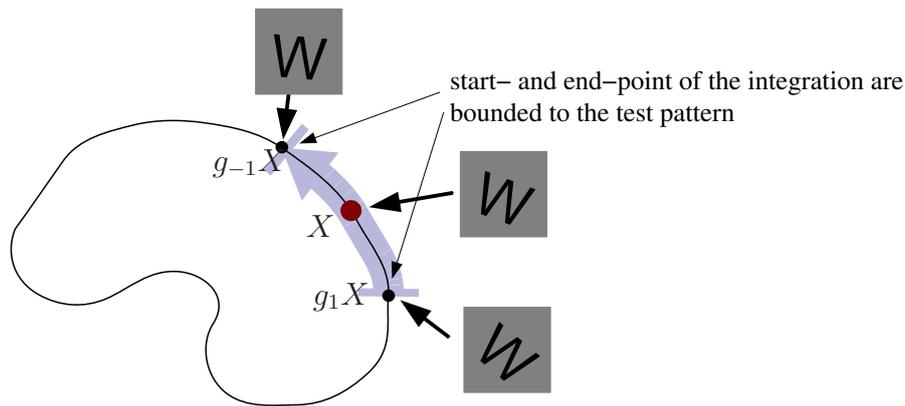


Figure 6.5: The partial Haar-integration features (introduced by [Haasdonk et al. \(2004\)](#) for optical character recognition) only use a subset of the transformation group. The images (marked in blue) that contribute to the final integral depend on the given test object. The resulting features are robust, but not invariant to the given transformations.

6.3.3 Explicit Limitation of the Integral by Probability-Weighted Haar Integration

For highly structured objects or a pre-processing of the images that generates strong discontinuities, the partial Haar-integration from the previous section might show big changes for small variations of the test-pattern. For this we introduce here a generalization of this idea to a probability-weighted Haar integration, that replaces the hard limits in (6.3) with “soft” limits,

$$T[f](X) = \int_{-\infty}^{\infty} f(g_q X) p(q) dq , \quad (6.4)$$

where $p(q)$ is the probability for the occurrence of the transformation q on the test image (see figure 6.6).

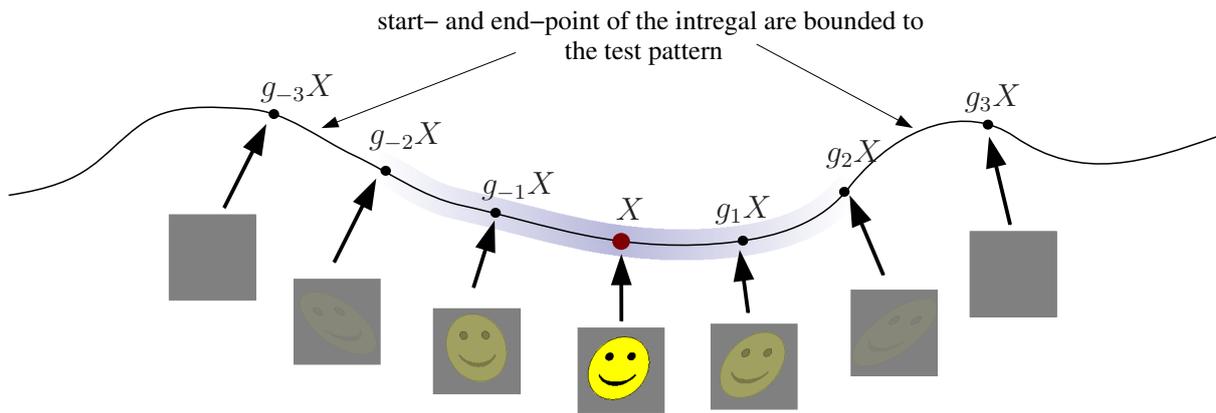


Figure 6.6: **Probability-weighted Haar integration:** For several important transformations no techniques for an implicit limitation of the integral can be found. Here we can use the same “trick” as in the partial Haar-integration to ensure the finiteness of the Haar integral. For more robustness, the hard limits are replaced by “soft” limits by introducing a probability-weighted Haar-integration.

6.3.4 Implicit Limitation of the Integral by the Use of Synthetic Channels

In the following we present a new idea for an implicit limitation of the integral for cases, where we cannot apply the implicit limitation from section 6.3.2 to ensure a finite result and where just a robustness by an explicit limitation (section 6.3.2 and 6.3.3) is not sufficient.

If we have further information about the object (e. g. a mask from a prior segmentation process), we can incorporate this prior knowledge by adding a synthetic channel to the input data – resulting in a multi-channel data set. This synthetic channel is computed from the raw data using the prior knowledge. By using an appropriate multi-channel kernel function with local support, we can reach the desired implicit limitation of the integral which ensures a finite result and keeps the full invariance properties.

For an illustrative example we take the group of scalings. The further assumptions are, that the object under consideration has a well defined border, and that an algorithm exists, that extracts this border, independent of the actual scaling of the object. Then we can render this border into a synthetic channel and set the rest of this channel to zero (see figure 6.7)

The selected multi-channel kernel function “senses” a certain area in the raw-data channel and in the synthetic channel and evaluates to zero if it finds only zero values within the synthetic channel (see

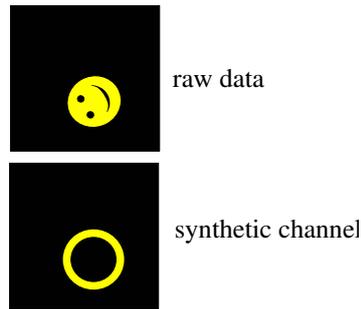


Figure 6.7: Adding of a synthetic channel that contains the border of the object

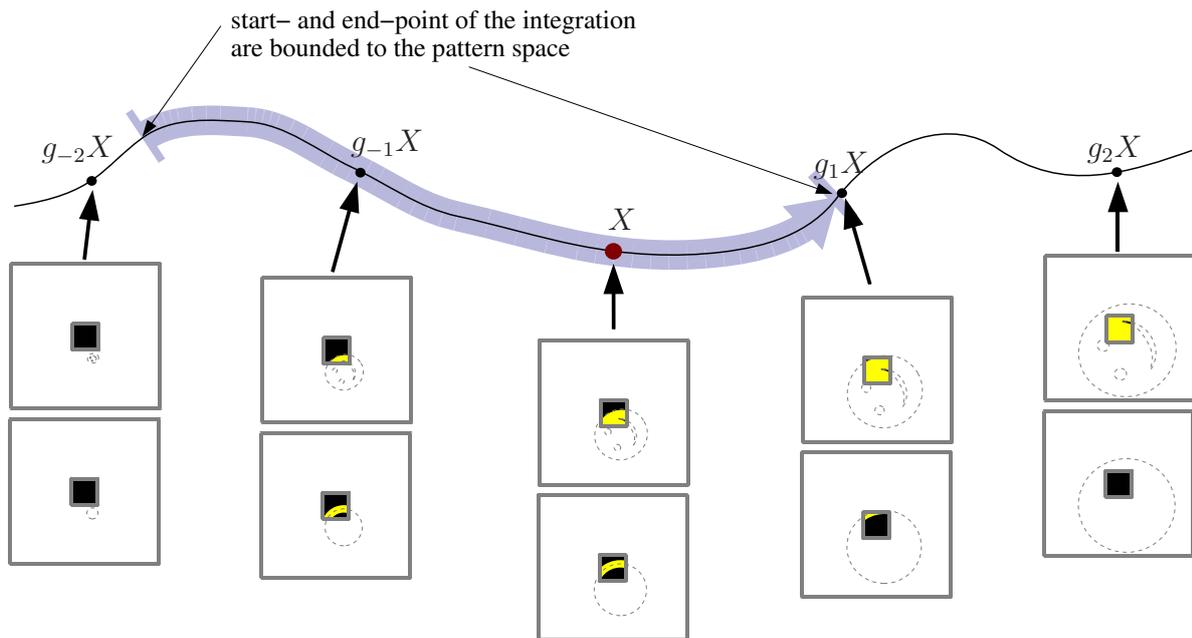


Figure 6.8: **Implicit limitation of the integral with a synthetic channel:** Extension of the dataset to a multi-channel-data set by adding a synthetic channel. In combination with a multi-channel kernel the required implicit limitation of the Haar integral, and therefore the full invariance properties are reached.

figure 6.8 for an illustration). As we can see, the kernel function returns a nonzero result only for that scalings of the object, where the border is within its “sensing area”. So independent of the scale of the test object, the points in the pattern space that contribute to the Haar integral are always the same.

Of course in a real application a scale invariance can only be reached down to that object size, where the sensor can resolve the desired structures and up to that size, where the full object is within the range of the sensor.

A more detailed discussion of the synthetic channels and their application to radial deformations is found in section 9.2.2.

6.4 Multi-Scale Approach to Enhance Richness of Sparse Kernel Functions

In the general formulation of the invariants, eq. (4.1), appropriate kernel functions can be used in order to sense any features of the structures at any scales. Computable kernel functions (4.3) depend only on a few gray values at certain points. To use these sparse kernels for sensing also large-scale structures it is usually not sufficient to just increase the distance between the kernel points. Only in an application where the test image is a rotated and translated exact copy of the reference image this results in reliable features. But in real applications with much more variations between the test and the reference image, it is not a good idea to base the features on the relation of two single gray values, that are, e. g., 200 pixels away from each other. It is obviously much better in this case to compare the mean gray value of a certain region with the mean gray value of another region 200 pixels away (see figure 6.9)

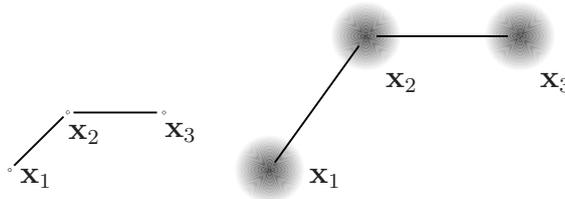


Figure 6.9: Computable kernels rely on a small number of sampling points. To sense information at multiple scales, the sampling points are “enlarged” with Gaussians of multiple size.

To integrate this desired behavior into the HI framework, but still use the fast-computable sparse kernels, we can apply a low-pass filtering (e. g. convolution with a Gaussian) to the data set before the evaluation of the kernel functions.

These techniques were proposed and used already since (Ronneberger et al., 2002a) and were based just on the above mentioned considerations. In section 9.1.3ff. we will show, that a description of the possible transformations by a local deformation model and probability-weighted integration over these deformations leads to very similar results and confirms the considerations above.

6.5 Integration of Prior Knowledge into the HI framework

For nearly all applications a significant amount of prior knowledge about the objects, their possible variations, the kind of distortions or noise in the data set, the background, etc. is available. For a reliable recognition we should integrate as much as possible of this prior knowledge into the recognition framework.

There are several possibilities to integrate prior knowledge into the HI framework. They can be categorized into the following classes:

1. adaption of the transformation group
2. adaption of the kernel functions
3. pre-processing of the raw data
4. creation of multi-channel datasets by computation of synthetic channels
5. post-processing of the resulting invariants

There is a smooth transition between these categories, and often the identical formulas can be derived, e.g, by incorporating the previous knowledge via a pre-processing or an adapted kernel function. This identity can also be used to speed up the computation of certain invariants.

6.5.1 Constraints of the Different Techniques

For the application of the above mentioned techniques some constraints have to be taken into account, to guarantee the desired invariance of the resulting features.

Adaption of the Transformation Group and the Kernel Function

The conditions for a specially adapted transformation group are the same that apply for any transformation group that shall be used within the HI framework and was discussed in detail in section 6.3.

The same can be stated for the selection of the kernel function. While here no direct limitations exist, it has to be selected such that the Haar integral is finite.

Pre-Processing of the Data and Generation of Synthetic Channels

The mapping, that is used for the pre-processing of the raw data or the generation of synthetic channels must be commutable with the action of the transformation group. I. e., for the pre-processing (mapping) ω it must be guaranteed that

$$g\omega X = \omega gX \quad \forall g \in G, \quad (6.5)$$

where G is the selected transformation group. Schulz-Mirbach (1995b) further weakened this condition and called it “weak-commutative mapping”: A mapping $w : V \rightarrow V$, that maps the space V onto itself is called weak-commutative, if for each $X \in V$ and $g_1 \in G$ you can find a $g_2 \in G$ such that

$$wg_1X = g_2wX. \quad (6.6)$$

For our purposes, the general commutation was sufficient. A simple example of a mapping that is commutable with the group of Euclidean motions is a convolution of the image with a rotationally symmetric Gaussian. The result is identical, no matter, if you first apply a rotation and second the convolution or if you first apply the convolution and second the rotation, see figure 6.10 for an example.

The same conditions apply for the creation of synthetic channels. Even though for the extraction of the outer surfaces it is quite difficult to prove their invariance (see chapter 8 for details). An example for the creation of a synthetic channel is given in fig 6.11.

Post-Processing of the Invariants

Once the invariants are computed, you can apply any post-processing to them without destroying the invariance properties. The motivation for such a post-processing might be to deform the feature space, such that it matches better to the needs of the classifier or to decompose the features in such a way, that certain variations in the raw data are removed or only affect a few features.

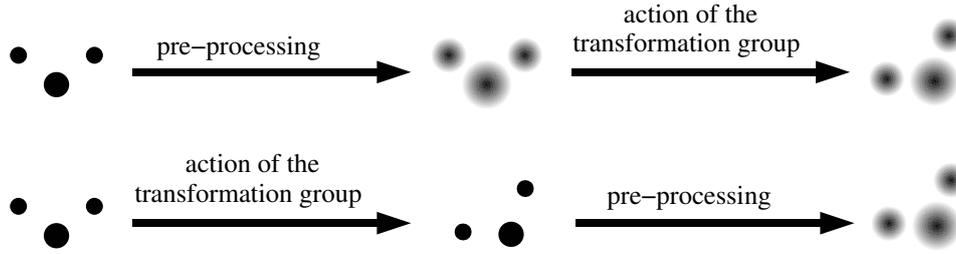


Figure 6.10: A pre-processing of the data must be commutable with the action of the transformation group, to ensure the desired invariance properties.

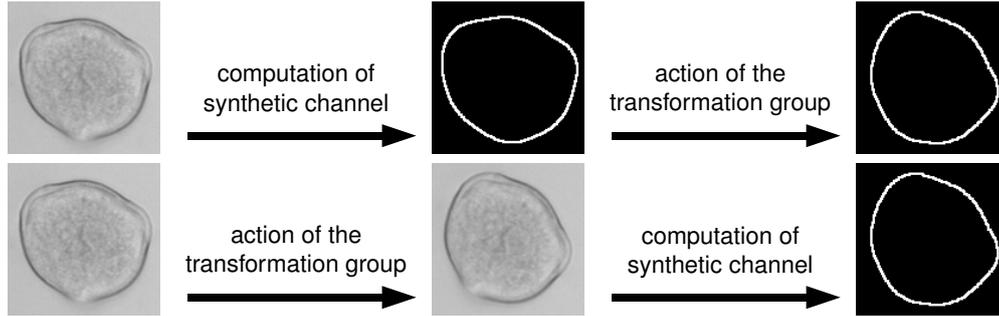


Figure 6.11: The extraction of a synthetic channel must be commutable with the action of the transformation group.

E. g., if we construct multiple invariants, that are proportional to a scaling of the gray values α of the raw image,

$$T[f_i](\alpha \cdot X) = \alpha \cdot T[f_i](X) \quad (6.7)$$

we can create features, that are invariant to such gray value scalings, by normalization with one of the invariants,

$$T_{\text{norm}}[f_i](\alpha \cdot X) = \frac{\alpha \cdot T[f_i](X)}{\alpha \cdot T[f_0](X)} = \frac{T[f_i](X)}{T[f_0](X)} = T_{\text{norm}}[f_i](X) \quad , \quad (6.8)$$

or (for more robustness) with the sum of all these invariants,

$$T_{\text{norm}}[f_i](X) = \frac{T[f_i](X)}{\sum_j T[f_j](X)} \quad . \quad (6.9)$$

6.6 Parameterized Kernel Functions

In most real applications one needs more than just a few invariants that result from a few manually selected kernel functions. A solution for this problem is to insert one or more freely selectable parameters p_1, p_2, \dots into the kernel-function. For a clearer notation we'll combine them into the parameter vector $\mathbf{p} = (p_1, p_2, \dots)$:

$$f(X) = f(X, \mathbf{p}) \quad (6.10)$$

When this kernel is used within the HI framework, we will get an invariant feature for each selected \mathbf{p} :

$$T[f](X, \mathbf{p}) := \int_G f(gX, \mathbf{p}) dg \quad (6.11)$$

For multiple kernel parameters, we now have a multidimensional invariant feature array describing the object. Each single feature in this array shares the same invariance and robustness properties, which are automatically generated by the Haar integration.

An obvious example for such a parameterized kernel is a two-point-kernel with a variable span p_1 :

$$f(X, \mathbf{p}) = f\left(X(0, 0, 0), X(0, 0, p_1)\right) \tag{6.12}$$

A less obvious example is to express the computation of the GVCD from section 5.2 as an invariant with a parameterized “co-occurrence kernel” with the first gray value p_1 , the second gray value p_2 and the span p_3 :

$$f(X, \mathbf{p}) = \delta\left(p_1 - X(0, 0, 0)\right) \cdot \delta\left(p_2 - X(0, 0, p_3)\right) \tag{6.13}$$

Especially in combination with synthetic channels these parameterized kernels allow to integrate several classical feature extraction techniques into the framework, while the framework guarantees the invariance to the selected transformation group.

6.7 Voxel-Wise Invariants for Simultaneous Segmentation and Recognition

For the recognition of objects or structures in a cluttered surrounding, we are often confronted with the following problem. For a reliable classification of the object, the object must be segmented from the background, but for the reliable segmentation we need to know to which class the object belongs.

A possible solution of this dilemma is to extract local features, that describe the structure around each voxel independent of a segmentation. To compute voxel-wise invariants using the HI framework, the transformation group is just a rotation, where the origin of the coordinate system is set to a certain voxel in advance. The resulting features from different kernel functions are collected in a feature vector, which then describes the surrounding of this voxel in a unique and rotation invariant way. This is done for all voxels in a volume. In a following step the feature vector of each voxel can be classified by a support vector machine (see figure 6.12 for an illustration).

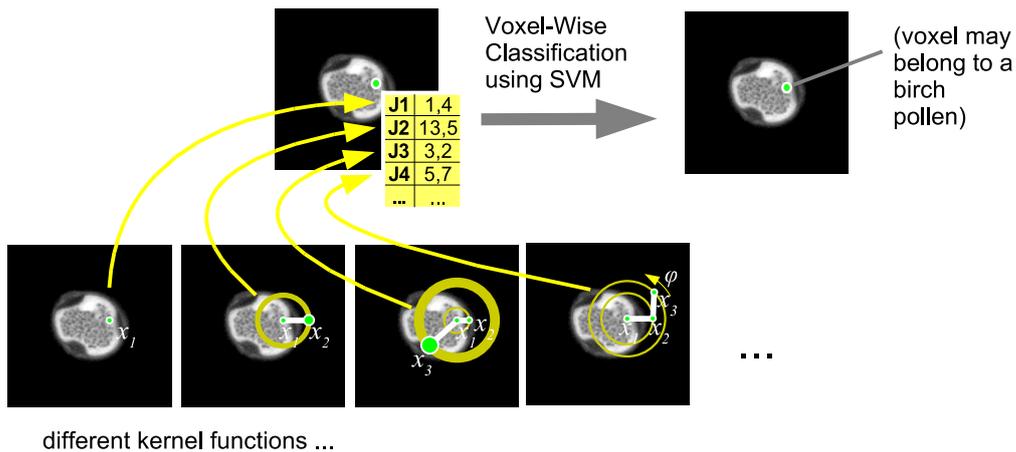


Figure 6.12: Example usage of voxel-wise invariants for simultaneous segmentation and recognition of objects.

6.7.1 Limitations of Centered Two-Point-Kernels for Voxel-Wise Invariants

Centered Two-point-kernel functions, i. e. kernels, whose first point is located at the rotation center¹, perform quite well in Haar integrals over Euclidean motions. For voxel-wise invariants they are somewhat limited in their discrimination power, because the resulting features are not only invariant to rotation of the surrounding but also to any random permutation of the gray values at the same radius (see figure 6.13).

For this application, centered three-point-kernels are much better suited, because they are able to sense the order of the gray values along the rotation path (see figure 6.14).

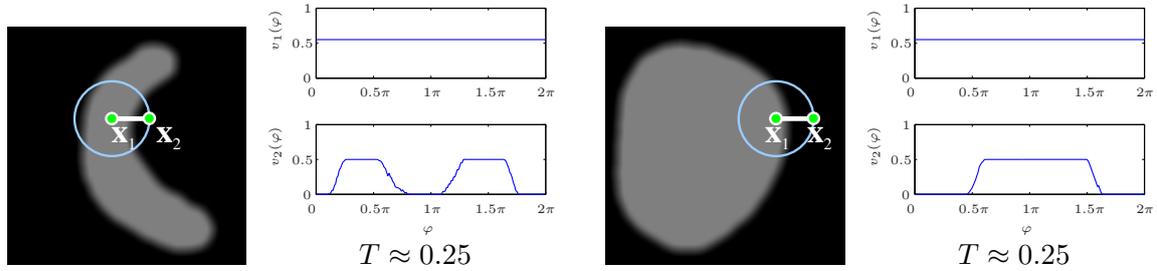


Figure 6.13: Voxel-wise invariants using the centered two-point kernel function $T = \int_0^{2\pi} v_1(\varphi) \cdot v_2(\varphi) d\varphi$ are not only invariant to rotation but also to any random permutation of the gray values in $v_2(\varphi)$. In this toy-example this results in nearly identical features for very different surroundings.

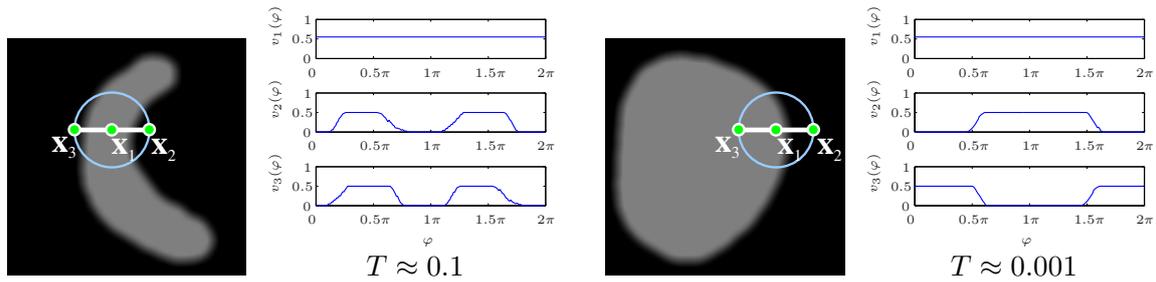


Figure 6.14: Voxel-wise invariants using centered three-point kernel functions $T = \int_0^{2\pi} v_1(\varphi) \cdot v_2(\varphi) \cdot v_3(\varphi) d\varphi$ are sensitive to such permutations and allow to distinguish between these surroundings.

6.8 Voxel-Wise Vectorial Invariants to Detect Spherical Structures

In cooperation with Janina Schulz (Schulz, 2005) another important class of kernel functions for voxel-wise invariants was developed. They are predestined for the detection of spherical structures in very cluttered environments. Besides their high robustness to even nonlinear gray value changes, a fast computation scheme with linear complexity (and a very small pre-factor) exists, which makes them suitable for even very large 3D volumetric data sets (see section 7.10)

¹Due to the close relation of the voxel-wise invariants to the full invariants over the group of Euclidean motions, the nomenclature for the number of kernel points was kept. This introduces a further restriction to these kernels, i. e. the first kernel point must be located in the rotation center. For full Euclidean motions this restriction does not apply, because for this transformation group the result is identical for kernels with the first kernel point located in the rotation center and for translated versions of it (see section 7.2).

These kernel functions operate on the gradient image ∇X and contain only one kernel point \mathbf{q} ,

$$f(X) = \|\nabla X(\mathbf{q})\| \cdot \delta \left(1 - \frac{\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\nabla X(\mathbf{q})}{\|\nabla X(\mathbf{q})\|} \right) . \quad (6.14)$$

the δ -function in the second term effects that this kernel returns a non-zero value only for certain conditions – Due to this we call it a “sparse vectorial kernel” function. By inserting this kernel into the HI framework,

$$T = \int_G f(gX) dg = \int_G \left\| (\nabla(gX))(\mathbf{q}) \right\| \cdot \delta \left(1 - \frac{\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{(\nabla(gX))(\mathbf{q})}{\|(\nabla(gX))(\mathbf{q})\|} \right) dg , \quad (6.15)$$

we see that the gradient must be computed on the transformed image, such that the replacement of the image transformation with the transformation of the kernel points (Schulz-Mirbach, 1995b) cannot be applied directly. Our definition for the image transformation is (4.2)

$$(gX)(\mathbf{x}) = X(\mathbf{s}_g(\mathbf{x})) \quad \forall g, \mathbf{x} . \quad (6.16)$$

By inserting the transformed image into the computation of a gradient,

$$(\nabla X)(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \begin{pmatrix} X(\mathbf{x} + \epsilon \cdot \mathbf{e}_1) - X(\mathbf{x}) \\ X(\mathbf{x} + \epsilon \cdot \mathbf{e}_2) - X(\mathbf{x}) \\ X(\mathbf{x} + \epsilon \cdot \mathbf{e}_3) - X(\mathbf{x}) \end{pmatrix} \quad (6.17)$$

with $\mathbf{e}_1 = (1, 0, 0)^T$, $\mathbf{e}_2 = (0, 1, 0)^T$, and $\mathbf{e}_3 = (0, 0, 1)^T$, we get

$$(\nabla(gX))(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \begin{pmatrix} X(\mathbf{s}_g(\mathbf{x} + \epsilon \cdot \mathbf{e}_1)) - X(\mathbf{s}_g(\mathbf{x})) \\ X(\mathbf{s}_g(\mathbf{x} + \epsilon \cdot \mathbf{e}_2)) - X(\mathbf{s}_g(\mathbf{x})) \\ X(\mathbf{s}_g(\mathbf{x} + \epsilon \cdot \mathbf{e}_3)) - X(\mathbf{s}_g(\mathbf{x})) \end{pmatrix} . \quad (6.18)$$

If the transformation group is the Euclidean motion, parameterized by

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t} , \quad (6.19)$$

we can rewrite (6.18) to

$$\begin{aligned} (\nabla(gX))(\mathbf{x}) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \begin{pmatrix} X(\mathbf{R} \cdot (\mathbf{x} + \epsilon \cdot \mathbf{e}_1) + \mathbf{t}) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \\ X(\mathbf{R} \cdot (\mathbf{x} + \epsilon \cdot \mathbf{e}_2) + \mathbf{t}) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \\ X(\mathbf{R} \cdot (\mathbf{x} + \epsilon \cdot \mathbf{e}_3) + \mathbf{t}) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \end{pmatrix} \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \begin{pmatrix} X(\mathbf{R}\mathbf{x} + \mathbf{t} + \epsilon \cdot \mathbf{R}\mathbf{e}_1) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \\ X(\mathbf{R}\mathbf{x} + \mathbf{t} + \epsilon \cdot \mathbf{R}\mathbf{e}_2) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \\ X(\mathbf{R}\mathbf{x} + \mathbf{t} + \epsilon \cdot \mathbf{R}\mathbf{e}_3) - X(\mathbf{R}\mathbf{x} + \mathbf{t}) \end{pmatrix} \\ &= \mathbf{R}^{-1} ((\nabla X)(\mathbf{R}\mathbf{x} + \mathbf{t})) . \end{aligned} \quad (6.20)$$

We can insert this result into the Haar integral (6.15), resulting in

$$T(\mathbf{t}) = \int_{\mathcal{O}_3} \left\| \mathbf{R}^{-1} ((\nabla X)(\mathbf{R}\mathbf{q} + \mathbf{t})) \right\| \cdot \delta \left(1 - \frac{\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\mathbf{R}^{-1} ((\nabla X)(\mathbf{R}\mathbf{q} + \mathbf{t}))}{\| \mathbf{R}^{-1} ((\nabla X)(\mathbf{R}\mathbf{q} + \mathbf{t})) \|} \right) d\mathbf{R} , \quad (6.21)$$

where O_3 denotes the group of all rotation matrices. As a rotation does not change the norm of a vector, this can be simplified to

$$T(\mathbf{t}) = \int_{O_3} \|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t}) \cdot \delta \left(1 - \frac{\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\mathbf{R}^{-1} \left(\frac{\nabla X}{\|\nabla X\|}(\mathbf{R}\mathbf{q} + \mathbf{t}) \right)}{\|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t})} \right) d\mathbf{R} . \quad (6.22)$$

The dot product of two vectors is identical, if both vectors are transformed with the same transformation, $\mathbf{x} \cdot \mathbf{R}^{-1}\mathbf{y} = \mathbf{R}\mathbf{x} \cdot \mathbf{R}\mathbf{R}^{-1}\mathbf{y} = \mathbf{R}\mathbf{x} \cdot \mathbf{y}$. This allows to further simplify the second term of the integral, yielding

$$\begin{aligned} T(\mathbf{t}) &= \int_{O_3} \|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t}) \cdot \delta \left(1 - \frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\frac{\nabla X}{\|\nabla X\|}(\mathbf{R}\mathbf{q} + \mathbf{t})}{\|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t})} \right) d\mathbf{R} \\ &= \int_{O_3} \|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t}) \cdot \delta \left(1 - \frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{R}\mathbf{q} + \mathbf{t}) \right) d\mathbf{R} . \end{aligned} \quad (6.23)$$

Now we can clearly see the properties of the Haar integral with this kernel function. For a given \mathbf{t} every point on a sphere (or circle for planar rotations) with the radius $\|\mathbf{q}\|$ is examined during the integration. Only if the normalized gradient $\frac{\nabla X}{\|\nabla X\|}$ (which is a unit vector that points into the gradient direction) is co-linear to the radial direction $\frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|}$, then the dot product will become 1 and the argument of the δ -function becomes 0 such that the gradient magnitude $\|\nabla X\|$ at this position contributes to the integral (see figure 6.15 for an illustration). For a perfect black circle on white background with sharp

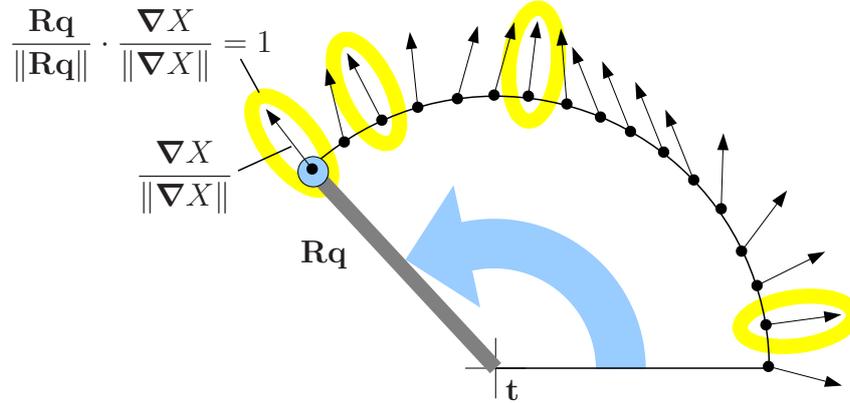


Figure 6.15: Voxel-wise Invariants for the detection of spherical structures. Only those gradients (marked in yellow) whose direction matches the direction of the rotated kernel contribute to the integral.

edges (high gradient magnitude) this integral returns a high value. For smoother edges or a deformed border a lower value is returned. For a real application the δ -function should be replaced with a less strict function (e. g. a sharp Gaussian) to also detect noisy and slightly deformed circles.

7 Fast Computation of the Invariants

For nearly all applications, the direct computation of the invariants (described in section 4.2) is too expensive. Therefore [Siggelkow and Schael \(1999\)](#) proposed a fast estimation of the integrals by a Monte-Carlo approach. While this technique is universally applicable to any kernel function and transformation group, the resulting features change from experiment to experiment. Furthermore this technique requires to specify the required accuracy for each feature. From this accuracy the needed number of evaluation points is computed. Especially for highly sophisticated classifiers like support vector machines, there is no general way to compute the required accuracy for each feature in advance, and even for a given setup we can just repeat the same experiment multiple times and try to find the needed number of evaluation points for each feature by an analysis of the statistics.

Due to these shortcomings we developed a framework for the fast and *exact* computation of the Haar integrals for certain kernel functions that makes use of convolutions based on the Fast Fourier Transform (FFT). For more complicated kernels, an approximative method based on Fourier- and Spherical-Harmonic-series expansion will be introduced later in this chapter. In contrast to the Monte-Carlo-estimation this approximation concerns only the given kernel function but not the result of the integral, such that the exact invariance properties are preserved.

7.1 Separable N-Point-Kernel Functions

This framework can only be used directly for the class of separable N-point-kernel functions,

$$f(X) = f_1(X(\mathbf{q}_1)) \cdot f_2(X(\mathbf{q}_2)) \cdot \dots \cdot f_N(X(\mathbf{q}_N)) . \quad (7.1)$$

For the class of general N-point-kernel functions,

$$f(X) = f(X(\mathbf{q}_1), X(\mathbf{q}_2), \dots, X(\mathbf{q}_N)) , \quad (7.2)$$

it can be applied, if the kernel function can be decomposed into a sum of separable kernel functions. E. g., the kernel

$$f(X) = (X(\mathbf{q}_1) - X(\mathbf{q}_2))^2 \quad (7.3)$$

can be decomposed into

$$f(X) = X^2(\mathbf{q}_1) - 2 \cdot X(\mathbf{q}_1) \cdot X(\mathbf{q}_2) + X^2(\mathbf{q}_2) , \quad (7.4)$$

where each summand is a separable kernel function and can be computed separately by the framework. In cases, where no such decomposition can be found, we can approximate the function with a (truncated) Taylor series expansion and subsequently evaluate each term of this series with the framework.

7.2 Using the Equivalence of Kernels

One central property of the HI framework is the existence of equivalent kernels for which the Haar integral returns the identical results. For a given kernel, the equivalent kernels can be constructed by applying any element of the transformation group g_k to the kernel points:

7 Fast Computation of the Invariants

If f_2 was constructed by the application of $g_k \in G$ from the kernel function f_1 such that $f_2(X) = f_1(g_k X) \quad \forall X$, then f_1 and f_2 are equivalent kernels and the Haar integral over these two kernels will be identical.

Proof by using the group properties of G :

$$T[f_2](X) = \int_G f_2(gX) dg = \int_G f_1(g_k g X) dg = \int_G f_1(g_2 X) dg_2 = T[f_1](X) . \quad (7.5)$$

E. g., for any two-point kernel we can find an Euclidean transformation, that shifts the first kernel point to the origin and places the second kernel point on the x -axis (see fig 7.1).

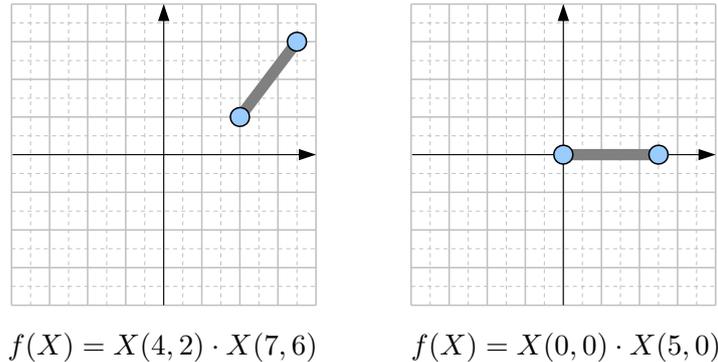


Figure 7.1: In the HI framework it is always possible to construct equivalent kernels by the application of any element of the transformation group to the kernel points. E. g., a Haar integration over the group of Euclidean transformations returns the identical results for the left and the right kernel, because the right kernel is just a shifted and rotated version of the left one.

We make use of this property for the fast computation of the Haar integral by replacing the kernel with an equivalent kernel such that certain kernel points become invariant to parts of the transformations, e. g. the point $(0, 0)$ is invariant to rotations.

7.3 Categorization of the Methods

Before going into detail we give a short overview of the possible applications of this framework, depending on the number of kernel points and the given transformation group (see table 7.1). In this table, the boldface entries describe, whether the result is exactly the same as the direct implementation, or if it is approximated (as mentioned above, this approximation only concerns the kernel, not the invariance properties of the final feature). Below the boldface entries the needed main operations are listed. The details are found in the following sections.

7.4 Two-Point-Kernels with 3D Euclidean Transformations

The first step in the fast computation of the Haar integral over Euclidean transformations with a two-point-kernel is to find the kernel equivalent to the given kernel(see section 7.2), where the first kernel point is located in the origin and therefore is invariant to rotations.

$$f(X) = f_a\left((X)(\mathbf{0})\right) \cdot f_b\left((X)(\mathbf{q})\right) \quad \begin{array}{l} f_a, f_b : \text{any nonlinear functions that} \\ \text{transform the gray values} \\ \mathbf{q} : \text{span of the kernel function} \end{array} \quad (7.6)$$

Table 7.1: Possible application of the proposed fast computation schemes depending on the kernel function and the transformation group. The term ‘‘sphere’’ is used here in its mathematical definition, which denotes the ‘‘spherical surface’’, not a ‘‘ball’’

kernel function	transformation group			
	3D translation + planar rotation	3D translation + 3D rotation	no translation planar rotation	no translation 3D rotation
$f_1(X(\mathbf{q}_1))$	exact (direct mean of full dataset)	exact (direct mean of full dataset)	exact (direct mean on circle)	exact (direct mean on sphere)
$f_1(X(\mathbf{q}_1)) \cdot f_2(X(\mathbf{q}_2))$	exact (convolution with circle)	exact (convolution with sphere)	approx. (Fourier trafo’s on circles)	approx. (SH-trafo’s on spheres)
$\prod_{i=1}^N f_i(X(\mathbf{q}_i))$ ($N = 3$)	approx. (convolutions with $e^{-ik\varphi}$ -weighted circles)	approx. (convolutions with SH-weighted spheres)	approx. (Fourier trafo’s on circles)	–
$\prod_{i=1}^N f_i(X(\mathbf{q}_i))$ ($N > 3$)	approx. (convolutions with $e^{-ik\varphi}$ -weighted circles)	–	approx. (Fourier trafo’s on circles)	–

by splitting the Euclidean group into the group of translations G_t and the group of rotations G_r the Haar integral (5.1) becomes

$$T[f](X) = \int_{G_t} \int_{G_r} f_a((g_t g_r X)(\mathbf{0})) \cdot f_b((g_t g_r X)(\mathbf{q})) dg_r dg_t . \quad (7.7)$$

Obviously the rotation g_r of the image before extracting the value at position $\mathbf{0}$ has no effect and can be omitted, which allows to move this term in front of the integral

$$T[f](X) = \int_{G_t} f_a((g_t X)(\mathbf{0})) \cdot \int_{G_r} f_b((g_t g_r X)(\mathbf{q})) dg_r dg_t . \quad (7.8)$$

By defining $A := f_a(X)$ and $B := f_b(X)$ and the parameterization for the transformed kernel point $\mathbf{q}' = \mathbf{R}\mathbf{q} + \mathbf{x}$, where \mathbf{R} is a rotation matrix and \mathbf{x} the translation vector, the Haar integral becomes

$$T[f](X) = \int_{\mathbb{R}^3} A(\mathbf{x}) \cdot \int_{O_3} B(\mathbf{R}\mathbf{q} + \mathbf{x}) d\mathbf{R} d\mathbf{x} . \quad (7.9)$$

where O_3 is the group of all rotation matrices \mathbf{R} . The integration over all rotations just describes the integral over a sphere (or a circle for planar rotation) with radius $\|\mathbf{q}\|$ and the center \mathbf{x} . We can therefore substitute the second integral with an integration over \mathbb{R}^3 and an appropriate δ -function,

$$T[f](X) = \int_{\mathbb{R}^3} A(\mathbf{x}) \cdot \int_{\mathbb{R}^3} B(\mathbf{t}) \cdot \delta(\|\mathbf{t} - \mathbf{x}\| - \|\mathbf{q}\|) d\mathbf{t} d\mathbf{x} , \quad (7.10)$$

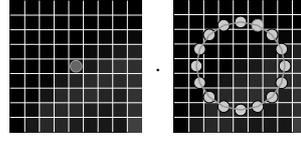
which allows to rewrite it as

$$\boxed{T[f](X) = \int_{\mathbb{R}^3} A(\mathbf{x}) \cdot (B * S)(\mathbf{x}) d\mathbf{x}} \quad \text{with } S(\mathbf{t}) := \delta(\|\mathbf{t}\| - \|\mathbf{q}\|) , \quad (7.11)$$

7 Fast Computation of the Invariants

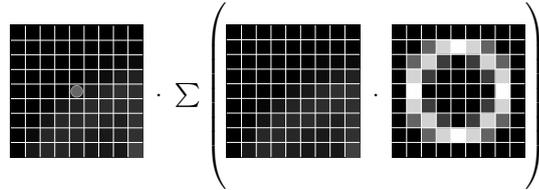
which is the convolution of B with the surface of a sphere (for 3D rotations) or with a circle (for planar rotations) and the subsequent point-wise multiplication with A . An intuitive derivation for this formula is shown in figure 7.2

(a) The sequential evaluation of the rotated kernel functions (as shown in figure 4.3b) is split into two steps: step 1: the gray values touched by the second kernel point within the rotation are summed up. step 2: the result is multiplied with the gray value of the first kernel point.



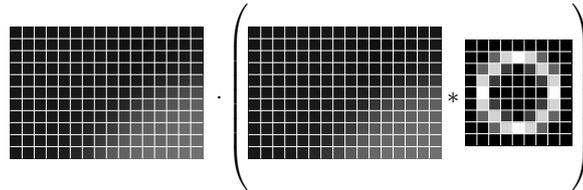
$$T_{\varphi}(0,0) = X(0,0) \cdot (\begin{array}{l} X(3,0) \\ + X(2.7,1.1) \\ + X(2.1,2.1) \\ + \dots \end{array})$$

(b) step 1 could now be replaced by a pixel-wise multiplication with an image of a circle and the Integration of the results.



$$T_{\varphi}(0,0) = X(0,0) \cdot \sum_{\mathbf{x}} X(\mathbf{x}) \cdot C(\mathbf{x})$$

(c) The evaluation of step 1 for all positions in the image is a simple convolution which could efficiently be computed by means of the Fast Fourier Transform (FFT).



$$T = \sum_{\mathbf{x}} X(\mathbf{x}) \cdot (X * C)(\mathbf{x})$$

$$T = \sum_{\mathbf{x}} X(\mathbf{x}) \cdot \left(\text{FFT}^{-1} \left(\text{FFT}(X) \cdot \text{FFT}^*(C_{\text{padded}}) \right) \right) (\mathbf{x})$$

Figure 7.2: Fast computation of Euclidean invariants with separable two-point kernels

The Fourier transformed circle or spherical surface can be computed in advance either directly in Fourier space or via an FFT. So only 2 FFT's are needed to compute one feature.

This method is preferred, when the integration over translations should be omitted (e. g. for voxel-wise features) or if the integration is replaced e. g. by computing a histogram over the resulting values.

7.4.1 Alternative Approach

In the second method¹ the integration over all translations is done first and can be expressed as a cross-correlation. For this we rewrite eq. 7.9 using $q = \|\mathbf{q}\|$ as

$$T[f](X) = \int_{\|\mathbf{t}\|=q} \int_{\mathbb{R}^3} A(\mathbf{x}) \cdot B(\mathbf{x} + \mathbf{t}) \, d\mathbf{x} \, d\mathbf{t} = \int_{\|\mathbf{t}\|=q} \text{ccf}(A, B)(\mathbf{t}) \, d\mathbf{t} \quad (7.12)$$

¹proposed by Qing Wang

which is the integration of the values on a sphere in the cross correlation function ccf. This integration of the cross correlation function can be done directly in the Fourier space

$$T[f](X) = \int_{\|\mathbf{t}\|=q} \left(\mathcal{F}^{-1} \left\{ \mathcal{F}\{A\}^* \cdot \mathcal{F}\{B\} \right\} \right) (\mathbf{t}) dt \quad (7.13)$$

$$= \int_{\|\mathbf{t}\|=q} \int_{\mathbb{R}^3} \hat{A}^*(\mathbf{k}) \cdot \hat{B}(\mathbf{k}) \cdot e^{i\mathbf{k}\mathbf{t}} d\mathbf{k} dt \quad (7.14)$$

$$= \int_{\mathbb{R}^3} \hat{A}^*(\mathbf{k}) \cdot \hat{B}(\mathbf{k}) \cdot \underbrace{\int_{\|\mathbf{t}\|=q} e^{i\mathbf{k}\mathbf{t}} dt}_{=: C_q(\mathbf{k})} d\mathbf{k} \quad (7.15)$$

$$\boxed{T[f](X) = \int_{\mathbb{R}^3} \hat{A}^*(\mathbf{k}) \cdot \hat{B}(\mathbf{k}) \cdot C_q(\mathbf{k}) d\mathbf{k}} \quad (7.16)$$

In 2D images $C_q(\mathbf{k})$ is a Fourier-transformed circle (which is the Bessel function $C_q(\mathbf{k}) = J_0(\|\mathbf{k}\| q)$). In 3D images $C_q(\mathbf{k})$ is a Fourier-transformed sphere surface, which is the sinc-function $C_q(\mathbf{k}) = \text{sinc}(\|\mathbf{k}\| q)$. This can be directly computed in Fourier space, such that only the two forward Fourier transforms of A and B are needed.

7.4.2 Comparison of the Computational Costs of the Two Approaches

The differences in computational costs show up, when computing more than one invariant on the same data set. Depending on the variation of the kernel functions, the intermediate results of the first computation can be reused, see table 7.2.

Table 7.2: Number of FFT's needed for the fast computation of invariants using two-point-kernels and Euclidean transformations

	Method A (7.11)	Method B (7.16)
first run on a new image	2 FFT's	2 FFT's
Variation of f_a	0	1 FFT
Variation of f_b	2 FFT's	1 FFT
Variation of kernel span $\ \mathbf{q}\ $	1 FFT	0

Depending on the selected kernel functions Method A or Method B or a mixture of both has to be selected to get the maximum speedup in the computation.

7.5 Three-Point-Kernels with 3D Translations and Planar Rotations

The first step here is again the replacement of the kernel function by an equivalent kernel, with its first point located at the rotation center,

$$f(X) = f_1(X(\mathbf{0})) \cdot f_2(X(\mathbf{q}_2)) \cdot f_3(X(\mathbf{q}_3)) , \quad (7.17)$$

7 Fast Computation of the Invariants

which allows (analogous to (7.8)) to move the first term out of the “rotation integral”,

$$T[f](X) = \int_{G_t} f_1\left((g_t X)(\mathbf{0})\right) \cdot \int_{G_r} f_2\left((g_t g_r X)(\mathbf{q}_2)\right) f_3\left((g_t g_r X)(\mathbf{q}_3)\right) dg_r dg_t . \quad (7.18)$$

but now two terms remain in the “rotation integral” such that it cannot be replaced by a single convolution as in the previous section.

In the following sections we present an expansion into a series of convolutions with the favorable property, that every truncated evaluation of this series still fulfills the invariance criterion.

For an easier understanding, we describe this solution first for planar rotations, because the required Fourier series expansion is more intuitive. The generalization for full 3D rotations (using spherical harmonic series expansion) will be given later.

The transformed kernel points are $\mathbf{q}'_i = \mathbf{R}_\varphi \mathbf{q}_i + \mathbf{x}$, where \mathbf{R}_φ is a rotation matrix that describes a planar rotation with angle φ around a given axis. To keep the further analysis clear, we choose w.l.o.g. a coordinate system, where the rotation axis is the z -axis. Now we can write the Haar integral as

$$T = \int_{\mathbb{R}^3} f_1\left(X(\mathbf{x})\right) \cdot \int_{-\pi}^{\pi} f_2\left(X(\mathbf{R}_\varphi \mathbf{q}_2 + \mathbf{x})\right) f_3\left(X(\mathbf{R}_\varphi \mathbf{q}_3 + \mathbf{x})\right) d\varphi d\mathbf{x} . \quad (7.19)$$

Defining the scalar values that are sensed by each kernel point during the integration as

$$v_1(\mathbf{x}) := f_1\left(X(\mathbf{x})\right) \quad (7.20)$$

$$v_2(\mathbf{x}, \varphi) := f_2\left(X(\mathbf{R}_\varphi \mathbf{q}_2 + \mathbf{x})\right) \quad (7.21)$$

$$v_3(\mathbf{x}, \varphi) := f_3\left(X(\mathbf{R}_\varphi \mathbf{q}_3 + \mathbf{x})\right) , \quad (7.22)$$

allows to rewrite the integral as

$$T = \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \int_{-\pi}^{\pi} v_2(\mathbf{x}, \varphi) v_3(\mathbf{x}, \varphi) d\varphi d\mathbf{x} . \quad (7.23)$$

For each given \mathbf{x} , $v_2(\mathbf{x}, \varphi)$ and $v_3(\mathbf{x}, \varphi)$ are periodic in φ and square-integrable, such that we can express them as Fourier series. By defining $V_{j,k}(\mathbf{x})$ as the k 'th Fourier coefficient of $v_j(\mathbf{x})$,

$$V_{j,k}(\mathbf{x}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} v_j(\mathbf{x}, \varphi) e^{-ik\varphi} d\varphi \quad (7.24)$$

we can replace $v_2(\mathbf{x}, \varphi)$ and $v_3(\mathbf{x}, \varphi)$ in (7.23) by their Fourier series expansions,

$$T = \int_{\mathbb{R}^3} v_1(\mathbf{x}) \int_{-\pi}^{\pi} \left(\sum_{k_2=-\infty}^{\infty} V_{2,k_2}(\mathbf{x}) \cdot e^{ik_2\varphi} \right) \cdot \left(\sum_{k_3=-\infty}^{\infty} V_{3,k_3}(\mathbf{x}) \cdot e^{ik_3\varphi} \right) d\varphi d\mathbf{x} . \quad (7.25)$$

Utilizing the orthogonality relationships between the basis functions

$$\int_{-\pi}^{\pi} e^{ik_2\varphi} e^{ik_3\varphi} d\varphi = 0 \quad \forall k_2 \neq -k_3 \quad (7.26)$$

allows to reduce this integral to a simple summation of products of these Fourier coefficients

$$\begin{aligned}
 T &= \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{k=-\infty}^{\infty} V_{2,k}(\mathbf{x}) V_{3,-k}(\mathbf{x}) \cdot \int_{-\pi}^{\pi} e^{ik\varphi} e^{-ik\varphi} d\varphi d\mathbf{x} \\
 &= \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{k=-\infty}^{\infty} V_{2,k}(\mathbf{x}) V_{3,-k}(\mathbf{x}) \cdot 2\pi d\mathbf{x} .
 \end{aligned} \tag{7.27}$$

If f_1 and f_2 are non-complex functions operating on non-complex gray values (which is the usual definition), we can use the symmetries in the Fourier transform of real-valued functions for a further reduction of the complexity.

$$T = 2\pi \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{k=0}^{\infty} \Re \left(V_{2,k}(\mathbf{x}) V_{3,k}^*(\mathbf{x}) \right) d\mathbf{x} \tag{7.28}$$

(where \Re denotes the real part of a complex variable). Furthermore we can truncate the series after the N 'th Fourier coefficient to compute the integral approximately

$$T = 2\pi \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{k=0}^N \Re \left(V_{2,k}(\mathbf{x}) V_{3,k}^*(\mathbf{x}) \right) d\mathbf{x} , \tag{7.29}$$

which is still perfectly invariant to rotation. This invariance can be proven by interpreting the truncated series as prior low-pass filtering of $v_2(\mathbf{x}, \varphi)$ and $v_3(\mathbf{x}, \varphi)$ in φ -direction (which is rotation invariant) and the subsequent evaluation of the full series.

To compute the Fourier coefficients $V_{2,k}(\mathbf{x})$ and $V_{3,k}(\mathbf{x})$ efficiently for each voxel, we use a framework similar to (7.11).

First the nonlinear gray value transformations f_1 , f_2 , and f_3 are applied to each voxel of the image X

$$A_1(\mathbf{x}) := f_1(X(\mathbf{x})) \tag{7.30}$$

$$A_2(\mathbf{x}) := f_2(X(\mathbf{x})) \tag{7.31}$$

$$A_3(\mathbf{x}) := f_3(X(\mathbf{x})) . \tag{7.32}$$

Then we combine for each \mathbf{x} the extraction of $v_2(\mathbf{x}, \varphi)$ (which are the gray values along a circle with center \mathbf{x} in image A_2) and the integration over the products $v_2(\mathbf{x}, \varphi) \cdot e^{-ik\varphi}$ from (7.24) (see figure 7.3) to a pixel-wise multiplication of image A_2 with a pre-computed image of a circle, that contains

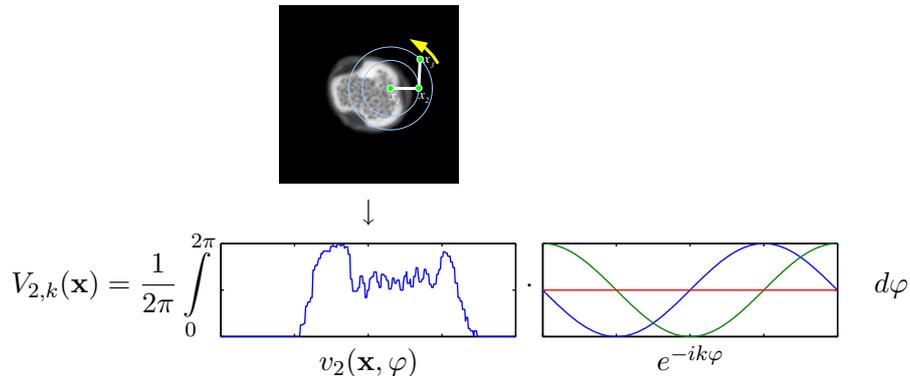


Figure 7.3: Direct computation of the Fourier coefficient V_{bk} ($k=1$) for the second kernel point x_2 the values $e^{-ik\varphi}$ along its circumference $S_{q_i,k}$ (see figure 7.4).

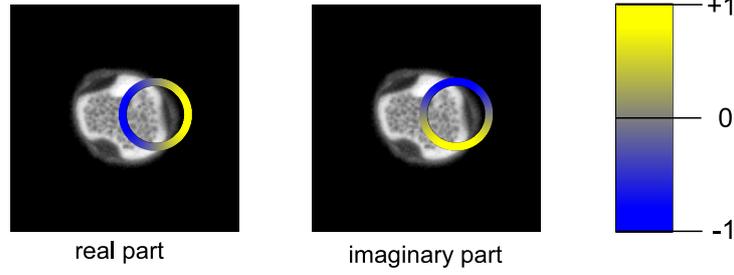


Figure 7.4: The extraction of the gray values along the circle $v_2(\mathbf{x}, \varphi)$ and the integration over the point-wise products of $v_2(\mathbf{x}, \varphi) \cdot e^{-ik\varphi}$, is replaced by a pixel-wise multiplication with a pre-computed image of a circle $S_{\mathbf{q}_2, k}$, that contains the values $e^{-ik\varphi}$ along its circumference.

This $S_{\mathbf{q}_i, k}$ can be best described in cylinder coordinates (x_r, x_φ, x_z) . The i 'th kernel point in cylinder coordinates is $\mathbf{q}_i = (q_{r_i}, q_{\varphi_i}, q_{z_i})$. Then

$$S_{\mathbf{q}_i, k}(x_r, x_\varphi, x_z) = \delta(x_z - q_{z_i}) \cdot \delta(x_r - q_{r_i}) \cdot e^{-ik(x_\varphi + q_{\varphi_i})} \quad (7.33)$$

This allows to compute the Fourier coefficients for each position of the kernel function simultaneously by a convolution of image A_2 with $S_{\mathbf{q}_2, k}$ (see figure 7.5):

$$V_{2, k}(\mathbf{x}) = \int_{\mathbb{R}^3} A_2(\mathbf{x} + \mathbf{t}) \cdot S_{\mathbf{q}_2, k}(\mathbf{t}) d\mathbf{t} \quad (7.34)$$

$$= (A_2 * S_{\mathbf{q}_2, k})(\mathbf{x}) \quad (7.35)$$

Analogously the Fourier coefficients $V_{3, k}$ are computed as:

$$V_{3, k}(\mathbf{x}) = (A_3 * S_{\mathbf{q}_3, k})(\mathbf{x}) \quad (7.36)$$

Inserting these results into (7.29), results in

$$T[f](X) = 2\pi \int_{\mathbb{R}^3} A_1(\mathbf{x}) \cdot \sum_{k=0}^N \Re \left(\left((A_2 * S_{\mathbf{q}_2, k})(\mathbf{x}) \cdot (A_3 * S_{\mathbf{q}_3, k})^*(\mathbf{x}) \right) \right) d\mathbf{x} , \quad (7.37)$$

where the convolutions can efficiently be computed via the FFT.

7.6 N-Point-Kernels with 3D Translations and Planar Rotations

For N-point kernel functions and planar rotations (with or without translations) the fast computation scheme described in the previous section 7.5 can be directly applied. For a clearer notation we describe it here without the translations. For the N-point kernel function

$$f(X) = f_1(X(\mathbf{q}_1)) \cdot f_2(X(\mathbf{q}_2)) \cdot \dots \cdot f_n(X(\mathbf{q}_n)) , \quad (7.38)$$

the transformed gray values, that are sensed by each kernel point during rotation are defined as

$$v_j(\varphi) := f_j(X(\mathbf{R}_\varphi \mathbf{q}_j)) . \quad (7.39)$$

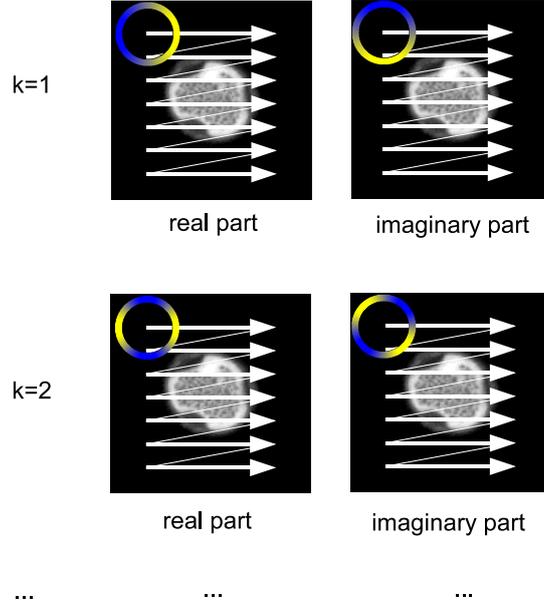


Figure 7.5: Effective computation of the Fourier coefficients for all translations of the kernel function by convolutions (via FFT)

The Haar integral then becomes

$$T = \int_{-\pi}^{\pi} v_1(\varphi) \cdot v_2(\varphi) \cdot \dots \cdot v_n(\varphi) d\varphi . \quad (7.40)$$

Using the Fourier series expansion (7.24), we get

$$T = \int_{-\pi}^{\pi} \left(\sum_{k_1=-\infty}^{\infty} V_{1,k_1} e^{ik_1\varphi} \right) \cdot \left(\sum_{k_2=-\infty}^{\infty} V_{2,k_2} e^{ik_2\varphi} \right) \cdot \dots \cdot \left(\sum_{k_n=-\infty}^{\infty} V_{n,k_n} e^{ik_n\varphi} \right) d\varphi . \quad (7.41)$$

Moving the Fourier coefficients before the integral

$$T = \sum_{(k_1, \dots, k_n) \in \mathbb{Z}^n} V_{1,k_1} \cdot V_{2,k_2} \cdot \dots \cdot V_{n,k_n} \cdot \underbrace{\int_{-\pi}^{\pi} e^{ik_1\varphi} \cdot e^{ik_2\varphi} \cdot \dots \cdot e^{ik_n\varphi} d\varphi}_{\neq 0 \text{ iff } (k_1 + \dots + k_n) = 0} , \quad (7.42)$$

allows to use the orthogonality relationships, such that

$$T = \sum_{(k_1, \dots, k_n) \in Y_{\infty}} V_{1,k_1} \cdot V_{2,k_2} \cdot \dots \cdot V_{n,k_n} \cdot 2\pi \quad (7.43)$$

where $Y_{\infty} = \left\{ (k_1, \dots, k_n) \mid (k_1, \dots, k_n) \in \mathbb{Z}^n \wedge (k_1 + \dots + k_n) = 0 \right\} .$

By truncating the Fourier series after the m 'th Fourier coefficient, we get

$$T = \sum_{(k_1, \dots, k_n) \in Y_m} V_{1,k_1} \cdot V_{2,k_2} \cdot \dots \cdot V_{n,k_n} \cdot 2\pi \quad (7.44)$$

where $Y_m = \left\{ (k_1, \dots, k_n) \mid (k_1, \dots, k_n) \in \{-m, \dots, m\}^n \wedge (k_1 + \dots + k_n) = 0 \right\} .$

If this scheme is used together with the group of translations, the integral becomes

$$T = 2\pi \cdot \int_{\mathbb{R}^3} \sum_{(k_1, \dots, k_n) \in Y_m} V_{1,k_1}(\mathbf{x}) \cdot V_{2,k_2}(\mathbf{x}) \cdot \dots \cdot V_{n,k_n}(\mathbf{x}) d\mathbf{x} \quad (7.45)$$

Analogously to (7.37) we can compute the voxel-wise Fourier coefficients with a convolution. By pre-computing the images with the nonlinearly transformed gray values as

$$A_i(\mathbf{x}) = f_i(X(\mathbf{x})) \quad (7.46)$$

and the images of the $e^{-ik\varphi}$ -weighted circles corresponding to each kernel-point as

$$S_{\mathbf{q}_i,k}(x_r, x_\varphi, x_z) = \delta(x_z - q_{z_i}) \cdot \delta(x_r - q_{r_i}) \cdot e^{-ik(x_\varphi + q_{\varphi_i})} \quad (7.47)$$

the final integral can be written as

$$T[f](X) = 2\pi \cdot \int_{\mathbb{R}^3} \sum_{(k_1, \dots, k_n) \in Y_m} (A_1 * S_{\mathbf{q}_1,k_1})(\mathbf{x}) \cdot (A_2 * S_{\mathbf{q}_2,k_2})(\mathbf{x}) \cdot \dots \cdot (A_N * S_{\mathbf{q}_n,k_n})(\mathbf{x}) d\mathbf{x} \quad (7.48)$$

7.7 Three-Point-Kernels with 3D Euclidean Transformations

The same techniques as in section 7.5 can be applied for 3D rotations, with the only difference, that rotations must be parameterized by three angles $\lambda = (x_1, x_2, x_3, \varphi_1, \varphi_2, \varphi_3)^T$ (see figure 7.6).

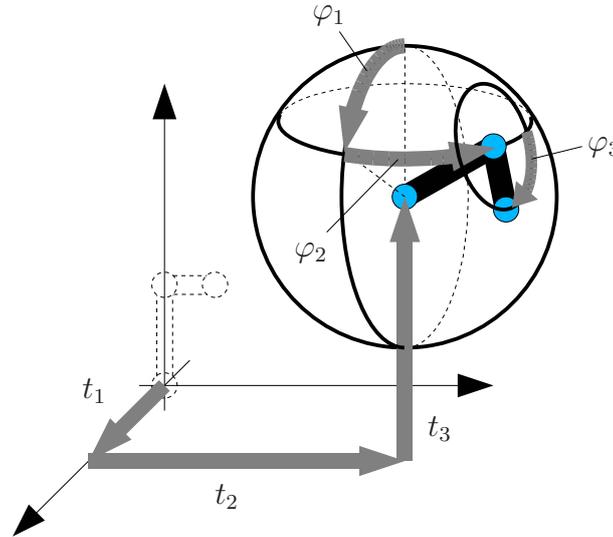


Figure 7.6: Parameterization of the 3D rotation with $\lambda = (x_1, x_2, x_3, \varphi_1, \varphi_2, \varphi_3)^T$

As in the previous sections the first step is to replace the given kernel by an equivalent kernel, such that the first kernel point is located at the origin. Additionally we rotate the second kernel point to the the z-axis,

$$\begin{aligned} \mathbf{q}_1 &= (0, 0, 0) \\ \mathbf{q}_2 &= (q_{21}, q_{22}, 0) \\ \mathbf{q}_3 &= (q_{31}, q_{32}, p_{33}) \end{aligned} \quad (7.49)$$

The grayvalues that are sensed by the kernel-points as function of the transformation parameters are

$$\begin{aligned} v_1(\mathbf{x}) &= f_1(X(\mathbf{x})) \\ v_2(\mathbf{x}, \varphi_1, \varphi_2) &= f_2\left(X\left(\mathbf{R}_{(\varphi_1, \varphi_2, 0)} \cdot \mathbf{q}_2 + \mathbf{x}\right)\right) \\ v_3(\mathbf{x}, \varphi_1, \varphi_2, \varphi_3) &= f_3\left(X\left(\mathbf{R}_{(\varphi_1, \varphi_2, \varphi_3)} \cdot \mathbf{q}_3 + \mathbf{x}\right)\right) . \end{aligned} \quad (7.50)$$

Therefore $v_1(\boldsymbol{\lambda})$ is invariant to all three rotations, $v_2(\boldsymbol{\lambda})$ invariant to φ_3 -rotations and only $v_3(\boldsymbol{\lambda})$ depends on all rotations. With this parameterization the Haar integral becomes

$$T = \int_{\mathbb{R}^3} v_1(\mathbf{x}) \int_0^\pi \int_{-\pi}^\pi v_2(\mathbf{x}, \varphi_1, \varphi_2) \int_{-\pi}^\pi v_3(\mathbf{x}, \varphi_1, \varphi_2, \varphi_3) d\varphi_3 d\varphi_2 \sin(\varphi_1) d\varphi_1 d\mathbf{x} . \quad (7.51)$$

The integration over φ_3 in the last term can be interpreted as a convolution of the gray values on the spherical surface with a circle on this spherical surface. The result only depends on the first two rotation angles, such that we define

$$v_{3\text{conv}}(\mathbf{x}, \varphi_1, \varphi_2) := \int_{-\pi}^\pi v_3(\mathbf{x}, \varphi_1, \varphi_2, \varphi_3) d\varphi_3 \quad (7.52)$$

This reduces the evaluation of the Haar integral to an element-wise multiplication and subsequent integration of two scalar fields defined on the 2-sphere. (see figure 7.7).

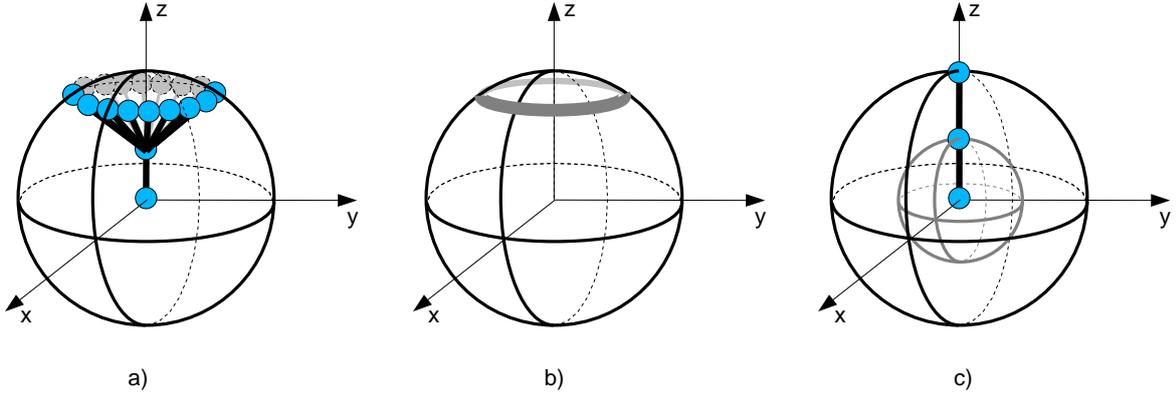


Figure 7.7: The integration of the kernel function over the third rotation angle φ_3 (a) can be replaced by a convolution (in spherical coordinates) with a circle (b) on the sphere and a subsequent multiplication of the corresponding gray values on the inner and the outer sphere (c).

The Haar integral becomes

$$T = \int_{\mathbb{R}^3} v_1(\mathbf{x}) \int_0^\pi \int_{-\pi}^\pi v_2(\mathbf{x}, \varphi_1, \varphi_2) \cdot v_{3\text{conv}}(\mathbf{x}, \varphi_1, \varphi_2) d\varphi_2 \sin(\varphi_1) d\varphi_1 d\mathbf{x} . \quad (7.53)$$

Analogous to Fourier series in 2D, the second term can be approximated with spherical harmonics as basis functions,

$$Y_m^\ell(\varphi_1, \varphi_2) = \sqrt{\frac{(2\ell + 1)(\ell - m)!}{4\pi(\ell + m)!}} P_m^\ell(\cos \varphi_1) e^{im\varphi_2} \quad (7.54)$$

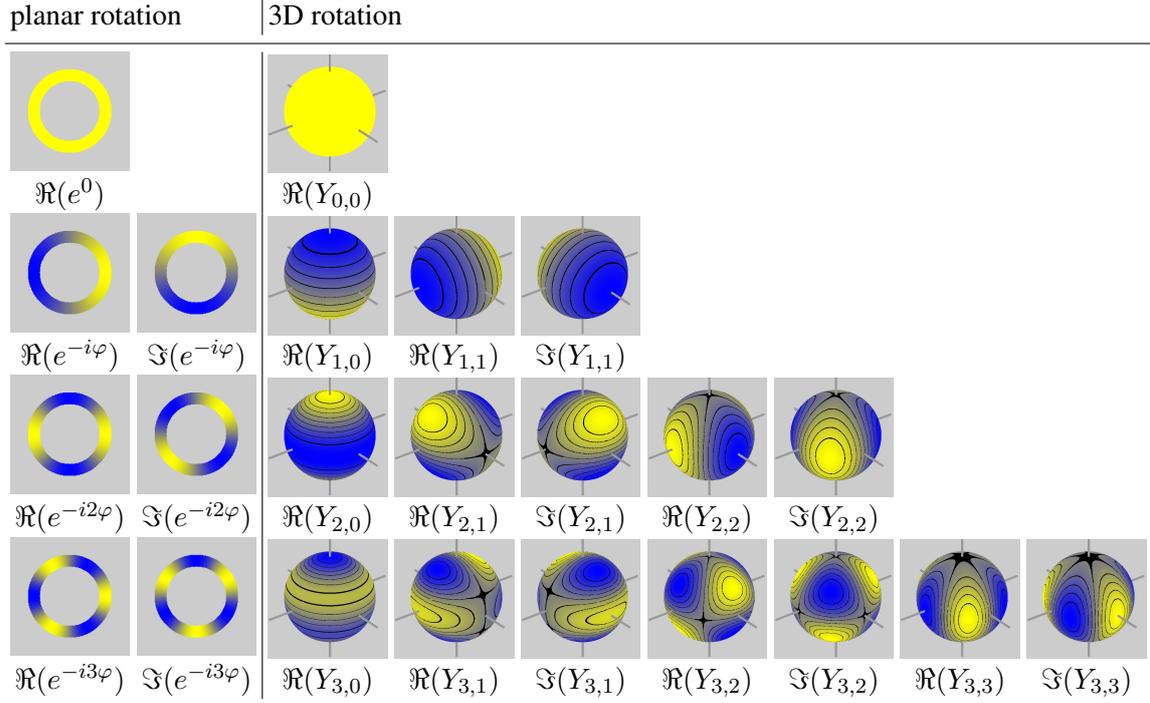


Figure 7.8: Fast Computation of voxel-wise invariants for 3D rotation. Instead of $e^{-ik\varphi}$ functions on circles, we need spherical harmonics on spherical surfaces

where P_m^ℓ are the associated Legendre polynomials. In figure 7.8 the basis functions are depicted. The SH-coefficients for the values sensed by the second kernel point are computed by

$$W_{2,\ell,m}(\mathbf{x}) = \int_0^\pi \int_{-\pi}^\pi v_2(\mathbf{x}, \varphi_1, \varphi_2) Y_m^{\ell*}(\varphi_1, \varphi_2) d\varphi_2 \sin(\varphi_1) d\varphi_1, \quad (7.55)$$

where $Y_m^{\ell*}(\varphi_1, \varphi_2) = (-1)^m Y_{-m}^\ell(\varphi_1, \varphi_2)$. The SH-coefficients of the values sensed by the third kernel point (already integrated around φ_3) are

$$W_{3\text{conv},\ell,m}(\mathbf{x}) = \int_0^\pi \int_{-\pi}^\pi v_{3\text{conv}}(\mathbf{x}, \varphi_1, \varphi_2) Y_m^{\ell*}(\varphi_1, \varphi_2) d\varphi_2 \sin(\varphi_1) d\varphi_1, \quad (7.56)$$

This allows to write the Haar integral as

$$T = \int_{\mathbb{R}^3} v_1(\mathbf{x}) \int_0^\pi \int_{-\pi}^\pi \left(\sum_{\ell_1=0}^\infty \sum_{m_1=-\ell_1}^{\ell_1} W_{2,\ell_1,m_1}(\mathbf{x}) \cdot Y_{m_1}^{\ell_1}(\varphi_1, \varphi_2) \right) \cdot \left(\sum_{\ell_2=0}^\infty \sum_{m_2=-\ell_2}^{\ell_2} W_{3\text{conv},\ell_2,m_2}(\mathbf{x}) \cdot Y_{m_2}^{\ell_2}(\varphi_1, \varphi_2) \right) d\varphi_2 \sin(\varphi_1) d\varphi_1 d\mathbf{x} \quad (7.57)$$

Using the orthogonality relationships between the basis functions Y_m^ℓ and the precondition, that our data is real-valued, allows to reduce this integral to a simple summation of products of the spherical harmonics coefficients

$$T \sim \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{l=0}^N \sum_{m=0}^l \Re \left(W_{2,\ell,m}(\mathbf{x}) \cdot W_{3\text{conv},\ell,m}^*(\mathbf{x}) \right) d\mathbf{x}, \quad (7.58)$$

where the series may be truncated after the N 'th coefficient without violating the rotation invariance. Here again we can compute the spherical harmonics coefficients for the surrounding of each voxel via a fast FFT-based convolution. So for each kernel function we need 2 forward FFT's and $2(N + 1)^2$ backward FFT's of the whole data set.

In the case of a first order approximation ($N = 1$) these are 10 FFT's. For a volume of 128^3 voxels a conventional 3GHz Pentium can compute such a real-to-complex FFT in about 60ms (using the "Fastest Fourier Transform of the West" FFTW (Frigo and Johnson, 2005)). So in this case the whole feature extraction could be done in less than a second for each kernel function.

7.7.1 Convolution on the 2-Sphere

For the computation of $v_{3\text{conv}}(\mathbf{x}, \varphi_1, \varphi_2)$ in (7.52) we need the convolution of a signal defined on the 2-sphere with a circle on the 2-sphere. Similarly to the convolution theorem for Fourier transforms, there is a convolution theorem for spherical harmonics transforms (Driscoll and Healy, 1994): Let $f, h \in L^2(S^2)$. Then

$$(\widehat{f * h})(\ell, m) = 2\pi \sqrt{\frac{4\pi}{2\ell + 1}} \hat{f}(\ell, m) \hat{h}(\ell, 0) . \quad (7.59)$$

As we anyhow need the spherical harmonics coefficients only of $v_{3\text{conv}}(\mathbf{x}, \varphi_1, \varphi_2)$, we can first develop the signal into a spherical harmonics series, were the kernel point \mathbf{q}_3 is rotated to the z-axis (as depicted in 7.7c),

$$\mathbf{q}_3' = (0, 0, \|\mathbf{q}_3\|) , \quad (7.60)$$

such that the sensed signal of this kernel point only depends on two rotation angles φ_1, φ_2 :

$$v_3'(\mathbf{x}, \varphi_1, \varphi_2) = f_3 \left(X \left(\mathbf{R}_{(\varphi_1, \varphi_2, 0)} \cdot \mathbf{q}_3' + \mathbf{x} \right) \right) \quad (7.61)$$

The extracted SH coefficients are then

$$W_{3,\ell,m}'(\mathbf{x}) = \int_0^\pi \int_{-\pi}^\pi v_3'(\mathbf{x}, \varphi_1, \varphi_2) Y_m^{\ell*}(\varphi_1, \varphi_2) d\varphi_2 \sin(\varphi_1) d\varphi_1 . \quad (7.62)$$

Now we can compute the required coefficients $W_{3\text{conv},\ell,m}(\mathbf{x})$ just by a band-wise multiplication with the coefficients of a SH-transformed circle around the north pole $W_{\text{circ},\ell,0}$:

$$W_{3\text{conv},\ell,m}(\mathbf{x}) = 2\pi \sqrt{\frac{4\pi}{2\ell + 1}} W_{3,\ell,m}'(\mathbf{x}) \cdot W_{\text{circ},\ell,0} . \quad (7.63)$$

Due to the fact, that the circle is just a delta pulse in φ_1 direction with a unit integral in φ_2 direction,

$$C(\varphi_1, \varphi_2) = \frac{1}{\sin \varphi_{\text{circ}}} \delta(\varphi_1 - \varphi_{\text{circ}}) , \quad (7.64)$$

where φ_{circ} is the colatitude of the circle, its SH-coefficients $W_{\text{circ},\ell,0}$ are

$$\begin{aligned} W_{\text{circ},\ell,0} &= \int_0^\pi \int_{-\pi}^\pi \frac{1}{\sin \varphi_{\text{circ}}} \delta(\varphi_1 - \varphi_{\text{circ}}) \cdot Y_0^{\ell*}(\varphi_1, \varphi_2) d\varphi_2 \sin(\varphi_1) d\varphi_1 \\ &= \int_{-\pi}^\pi Y_0^{\ell*}(\varphi_{\text{circ}}, \varphi_2) d\varphi_2 \\ &= \sqrt{\frac{2\ell + 1}{4\pi}} P_0^\ell(\cos \varphi_{\text{circ}}) \cdot \int_{-\pi}^\pi e^{i \cdot 0 \cdot \varphi_2} d\varphi_2 \\ &= \sqrt{\frac{2\ell + 1}{4\pi}} P_0^\ell(\cos \varphi_{\text{circ}}) \cdot 2\pi . \end{aligned} \quad (7.65)$$

So they can be directly computed from the associated Legendre polynomial. By inserting this result in (7.63), we get

$$W_{3\text{conv},\ell,m}(\mathbf{x}) = 4\pi^2 P_0^\ell(\cos \varphi_{\text{circ}}) \cdot W'_{3,\ell,m}(\mathbf{x}) . \quad (7.66)$$

This can be inserted into equation (7.58), and the $P_0^\ell(\cos \varphi_{\text{circ}})$ (which has no imaginary component and does not depend on m) can be moved before the sum, such that

$$T \sim \int_{\mathbb{R}^3} v_1(\mathbf{x}) \cdot \sum_{l=0}^N P_0^\ell(\cos \varphi_{\text{circ}}) \cdot \sum_{m=0}^l \Re \left(W_{2,\ell,m}(\mathbf{x}) \cdot W_{3,\ell,m}^{*\prime}(\mathbf{x}) \right) d\mathbf{x} . \quad (7.67)$$

7.8 Variations of the Kernel Angles

One important advantage of the fast algorithms described above is, that we can reuse the computed Fourier- or SH-coefficients for further kernels, if these kernels differ from the first just by the angles between their “arms” (see figure 7.9 for an illustration).

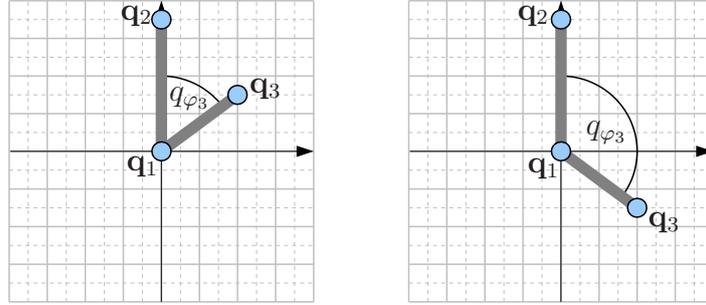


Figure 7.9: Two 3-point-kernels that differ only by the angle between their “arms”

If two N -point-kernels

$$f_a(X) = \prod_{i=1}^N f_{a,i}(X(\mathbf{q}_{a,i})) \quad \text{and} \quad f_b(X) = \prod_{i=1}^N f_{b,i}(X(\mathbf{q}_{b,i})) \quad (7.68)$$

have identical sub-functions and the kernel points have the identical distances to the rotation center,

$$\begin{aligned} f_{a,i}(x) &= f_{b,i}(x) & \forall i \in \{1, \dots, N\}, \forall x \in \mathbb{R} \\ \|\mathbf{q}_{a,i}\| &= \|\mathbf{q}_{b,i}\| & \forall i \in \{1, \dots, N\} , \end{aligned} \quad (7.69)$$

then the required Fourier- or SH- coefficients for a fast computation of the Haar integral with the second kernel can be computed directly from the Fourier- or SH- coefficients of the first kernel. For the Fourier coefficients this is a simple phase shift, that can be performed by a complex multiplication:

$$\begin{aligned} V_{b,j,k}(\mathbf{x}) &= e^{-ik\Delta\varphi_j} \cdot V_{a,j,k}(\mathbf{x}) \\ \text{with } \Delta\varphi_j &:= q_{\varphi_{b,j}} - q_{\varphi_{a,j}} , \end{aligned} \quad (7.70)$$

where $q_{\varphi_{a,i}}$ and $q_{\varphi_{b,i}}$ are the angular components of the kernel points in cylinder coordinates.

For the 3-point-kernels that we can compute via the spherical harmonic expansion, the computation for further kernels with different angles is even simpler: the only parameter that changes, is the colatitude φ_{circ} of the circle (depicted in figure 7.7c) that is required for the convolution on the 2-sphere. $W_{2,\ell,m}(\mathbf{x})$ and $W_{3,\ell,m}^{*\prime}(\mathbf{x})$ keep identical, such that we can directly compute T using equation (7.67) with the modified φ_{circ} .

7.9 Fast Computation of Voxel-Wise Invariants with Separable Kernels

For the fast computation of the voxel-wise invariants for simultaneous segmentation and recognition that were introduced in section 6.7, the above described techniques can be directly applied (except for the alternative approach of section 7.4.1). The only difference is that the final integration over all translations is left out. The different steps, that need to be applied for such a computation of voxel-wise invariants using a 3-Point-Kernel and full 3D rotations is depicted in figure 7.10.

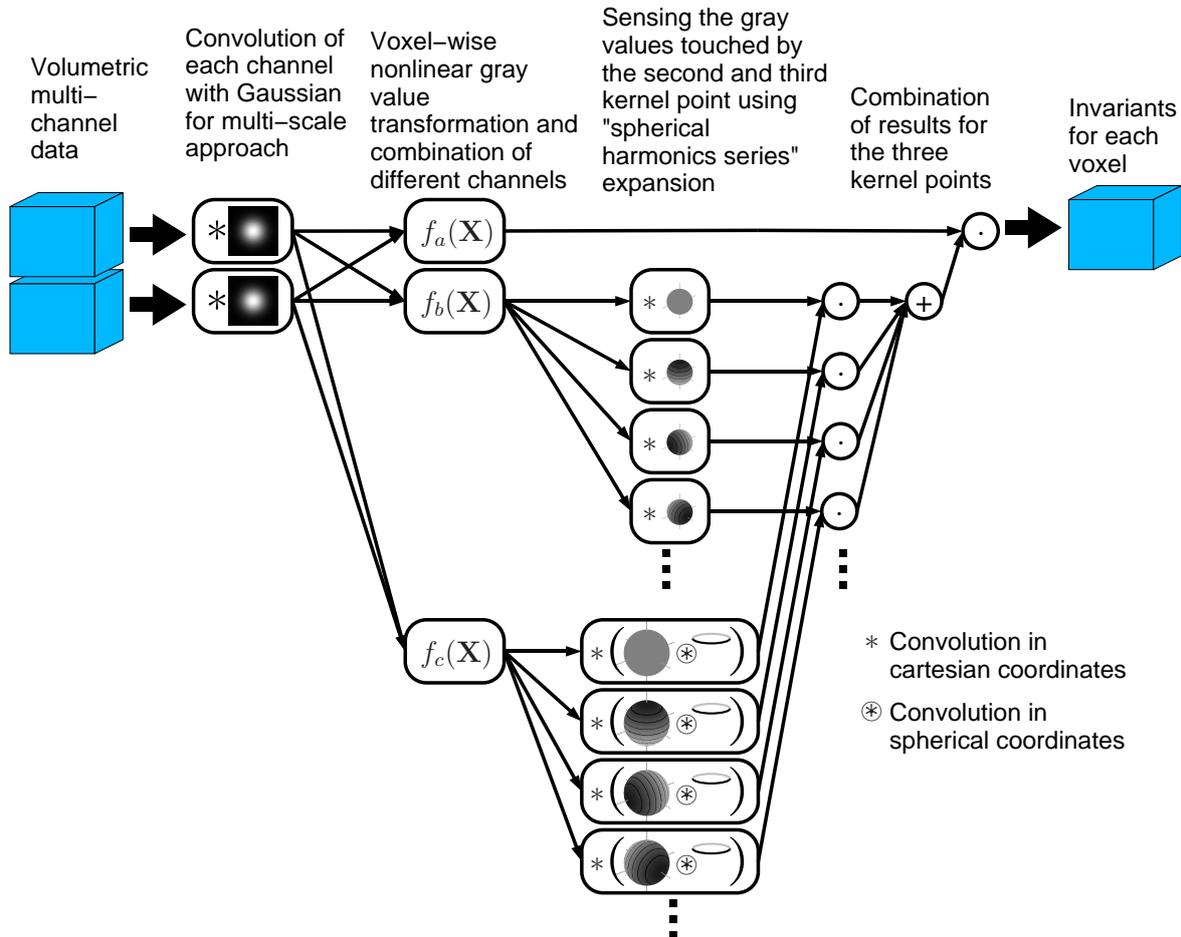


Figure 7.10: Computation of voxel-wise invariants using three-point-kernels $f(X) = f_a(X(\mathbf{0})) \cdot f_b(X(\mathbf{q}_2)) \cdot f_c(X(\mathbf{q}_3))$ on multi-channel volumetric data. For each kernel function this scheme simultaneously calculates the features for all voxels.

7.10 Fast Computation of Voxel-Wise Invariants with Sparse Vectorial Kernels

For the fast computation of voxel-wise invariants with sparse vectorial kernels (introduced in section 6.8)

$$T(\mathbf{t}) = \int_{O_3} \|\nabla X\|(\mathbf{R}\mathbf{q} + \mathbf{t}) \cdot \delta\left(1 - \frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{R}\mathbf{q} + \mathbf{t})\right) d\mathbf{R} \quad (7.71)$$

a different scheme is used, that does not base on FFT's or spherical harmonics. In this case the sparseness of the kernel function (the kernel returns a non-zero-result only in very sparse cases) allows a significant speedup by a simultaneously “inverse” computation of the kernel for multiple positions.

The direct computation would be to translate the coordinate origin to position \mathbf{t} and then to iterate over all rotation matrices to evaluate the kernel function at the resulting position $\mathbf{R}\mathbf{q} + \mathbf{t}$ and integrate the results. After that, this procedure is repeated for the next position \mathbf{t} . As mentioned above in most cases, the result of the kernel function will be just zero, but we can not know this in advance until we have checked the gradient at position $\mathbf{R}\mathbf{q} + \mathbf{t}$.

The idea is now, to evaluate all integrals simultaneously by setting up an individual accumulator for each position \mathbf{t} . Then we iterate over all positions \mathbf{x} and check, to which integral this voxel will contribute, i. e. for which \mathbf{R} and \mathbf{t} the position \mathbf{x} will be reached.

$$\mathbf{x} \stackrel{!}{=} \mathbf{R}\mathbf{q} + \mathbf{t} \quad (7.72)$$

For a non-sparse kernel function this would result in an iteration over all rotation matrices \mathbf{R} . For each given \mathbf{R} we could compute the according \mathbf{t} and add the kernel result to the attached accumulator. This would result in exactly the same complexity like the direct evaluation.

For the sparse kernels, we can compute in advance, for which \mathbf{R} the kernel returns a non-zero result. For this the argument of the δ -function must be zero,

$$\begin{aligned} 1 - \frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{x}) &\stackrel{!}{=} 0 \\ \Rightarrow \frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{x}) &\stackrel{!}{=} 1 . \end{aligned} \quad (7.73)$$

The dot product of two normalized vectors equals only to 1, if the are identical, i. e.

$$\frac{\mathbf{R}\mathbf{q}}{\|\mathbf{q}\|} = \frac{\nabla X}{\|\nabla X\|}(\mathbf{x}) . \quad (7.74)$$

From this we can compute $\mathbf{R}\mathbf{q}$ as

$$\mathbf{R}\mathbf{q} = \|\mathbf{q}\| \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{x}) . \quad (7.75)$$

Without the need to compute the concrete rotation \mathbf{R} , we can directly insert $\mathbf{R}\mathbf{q}$ into (7.72) and find the position of the accumulator as

$$\mathbf{t} = \mathbf{x} - \|\mathbf{q}\| \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{x}) . \quad (7.76)$$

All other accumulators need not to be touched, which reduces the computational costs to only a few multiplications and one accumulator increment per voxel.

7.11 Computational Costs

At the end of this chapter we give a summary of the computational costs for the proposed techniques compared to the direct implementation described in section 4.2. There are several parameters that

significantly influence the final computation time. None of these parameters varies in multiple orders of magnitude, such that the real computational costs are significantly influenced by all of them. Due to this all relevant factors and some pre-factors are listed in table 7.3, such that a column-wise comparison of the values is possible. For the parameter variation in the two-point-kernels the cheaper of both proposed methods is given (see section 7.4.2 for a detailed comparison).

Table 7.3: Computational costs.

kernel	transl.	rotation	direct impl. per feature	fast implementation variation of per feature		initialization
2-point	3D	planar	$\mathcal{O}(N_x^3 \cdot N_\sigma^2 \cdot N_q)$	\mathbf{q}	$\mathcal{O}(N_x^3)$	$2 \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
				f_a	$\mathcal{O}(N_x^3)$	$3 \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
				f_b	$\mathcal{O}(N_x^3 + N_x \cdot \text{FFT}(N_x^2))$	$1 \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
2-point	3D	3D	$\mathcal{O}(N_x^3 \cdot N_\sigma^3 \cdot N_q^2)$	\mathbf{q}	$\mathcal{O}(N_x^3)$	$2 \cdot \mathcal{O}(\text{FFT}(N_x^3))$
				f_a	$\mathcal{O}(N_x^3)$	$3 \cdot \mathcal{O}(\text{FFT}(N_x^3))$
				f_b	$\mathcal{O}(N_x^3 + \text{FFT}(N_x^3))$	$1 \cdot \mathcal{O}(\text{FFT}(N_x^3))$
3-point	3D	planar	$\mathcal{O}(N_x^3 \cdot N_\sigma^2 \cdot N_q)$	f_1	$\mathcal{O}(N_x^3)$	$(2N_b + 2) \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
				$\Delta\varphi$	$N_b \cdot \mathcal{O}(N_x^3)$	$(2N_b + 2) \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
				other	$N_b \cdot \mathcal{O}(N_x^3 + 2N_x \cdot \text{FFT}(N_x^2))$	$2 \cdot \mathcal{O}(N_x \cdot \text{FFT}(N_x^2))$
3-point	3D	3D	$\mathcal{O}(N_x^3 \cdot N_\sigma^3 \cdot N_q^3)$	f_1	$\mathcal{O}(N_x^3)$	$(2N_b + 2) \cdot \mathcal{O}(\text{FFT}(N_x^3))$
				$\Delta\varphi$	$N_b^2 \cdot \mathcal{O}(N_x^3)$	$(2N_b + 2) \cdot \mathcal{O}(\text{FFT}(N_x^3))$
				other	$N_b^2 \cdot \mathcal{O}(N_x^3 + 2 \cdot \text{FFT}(N_x^3))$	$2 \cdot \mathcal{O}(\text{FFT}(N_x^3))$
sparse vect.	3D	3D	$\mathcal{O}(N_x^3 \cdot N_q^2)$	–	$\mathcal{O}(N_x^3)$	–

N_F	number of features, that share certain kernel parameters	(typically 1-50)
N_x	extent of data set in one direction	(typically 100 - 500)
N_σ	smoothing radius	(typically 2 - 50)
N_q	radius of kernel function	(typically 1 - 100)
N_b	number of bands for approximation	(typically 1-5)
other	any of the parameters $f_2, f_3, \mathbf{q}_2, \mathbf{q}_3$	

In this presentation we see that the costs for the direct implementation are always linear in the number of pixels (or voxels) N_x^3 as stated by Schulz-Mirbach (1995b), but they are also up to cubic in the kernel radius N_q and up to cubic in the smoothing radius N_σ , which results usually in an enormous pre-factor.

In contrast to this the costs for the proposed methods are independent of the kernel radius N_q and the smoothing radius N_σ , and are in many cases linear in the number of voxels with a very low pre-factor. On the other hand, they introduce some fix costs for the initialization, and for certain variations of the kernel parameters a certain number of 2D- or 3D- FFT's is needed per feature.

By looking at the typical ranges for the parameters (listed below the table), we see that in nearly all cases the proposed methods will outperform the direct implementation. Only for very small kernel radii and very small smoothing the direct implementation might be faster.

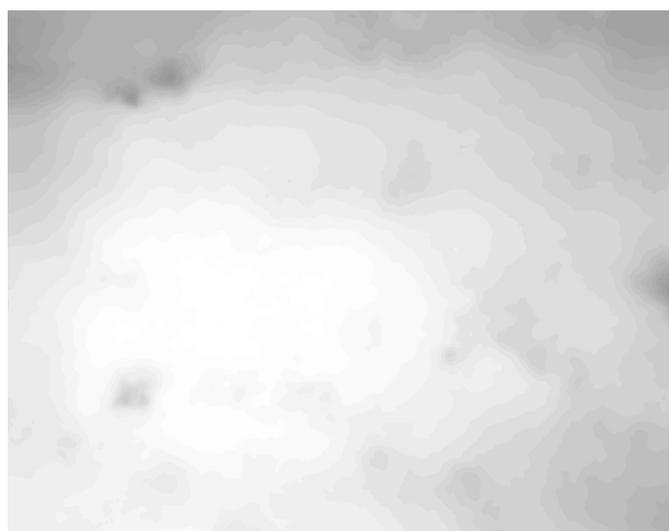
8 Detection and Segmentation

8.1 2D Segmentation using Morphological Techniques

For air samples with simple conditions (clear background, not too many particles in the image, no agglomerations of pollen, no occlusions, etc.) the segmentation can be done with traditional morphological techniques. In such cases we can assume that the majority of the image pixels belong to the background. Furthermore we have to know in advance the diameter of the smallest pollen and the largest pollen that shall be segmented. Under these conditions the method described in the following will return satisfactory results. It can be applied either to the fluorescence image (searching particles that are brighter than the background) or in transmitted light images by searching particles that are darker than the background. Depending on the age and the current bleaching state of the pollen in the sample the first or the latter is more advantageous.

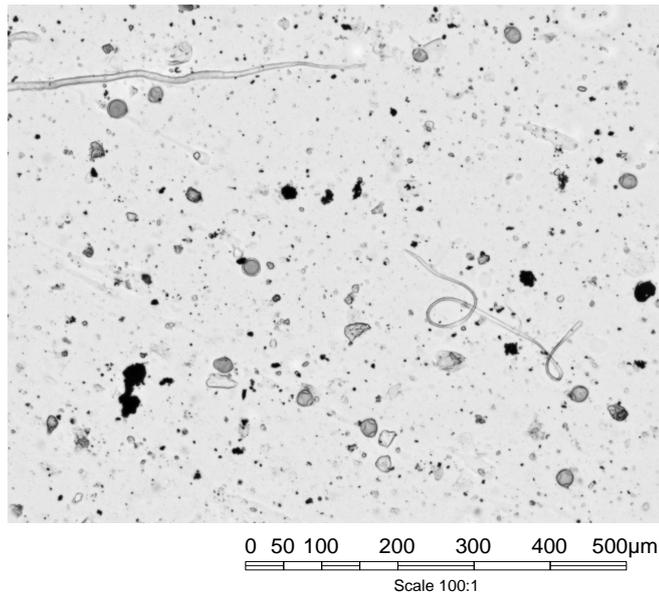
The illustrating images for each step show an air sample that was taken with the traditional Burkard spore trap, manually prepared and recorded with a 10x objective with a 12bit camera at 1280x1024 pixels. The resulting size of each image-pixel is $0.67\mu\text{m} \times 0.67\mu\text{m}$

Step 1: Estimation of the background intensity: The background intensity at each pixel is estimated by a running median filter using a disc as structural element with twice the diameter of the largest expected particle. If the particle density is not too high, this diameter guarantees that the filter response is based on at least 3 times more background pixels than foreground pixels.

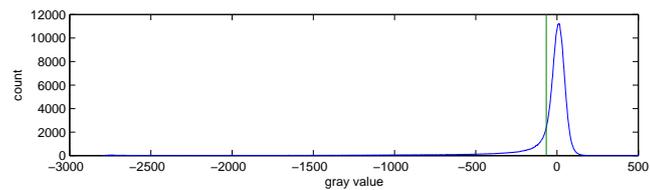


0 50 100 200 300 400 500 μm
Scale 100:1

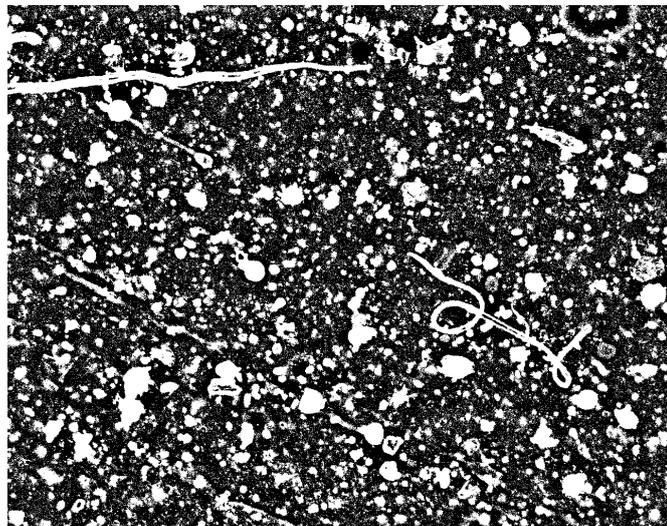
Step 2: Shading correction: The estimated background is subtracted. After this operation the mean background has everywhere the gray value 0 and dark particles have negative gray values



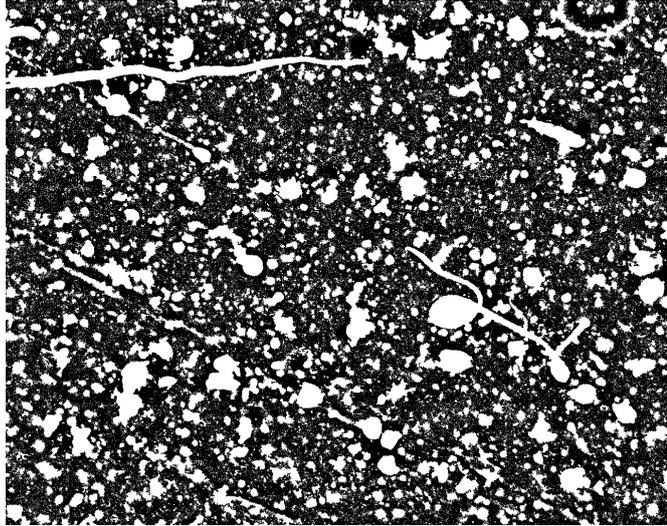
Step 3: Automated determination of the best threshold: The histogram of a typical air sample image is always dominated by one peak belonging to the background pixels (which is centered here around zero due to the prior shading correction). Its width depends on the background noise of the image. A robust estimation for the threshold to separate low contrast particles from the background noise was heuristically found to be the maximum of the second derivative of the histogram curve on the left side of the main peak.



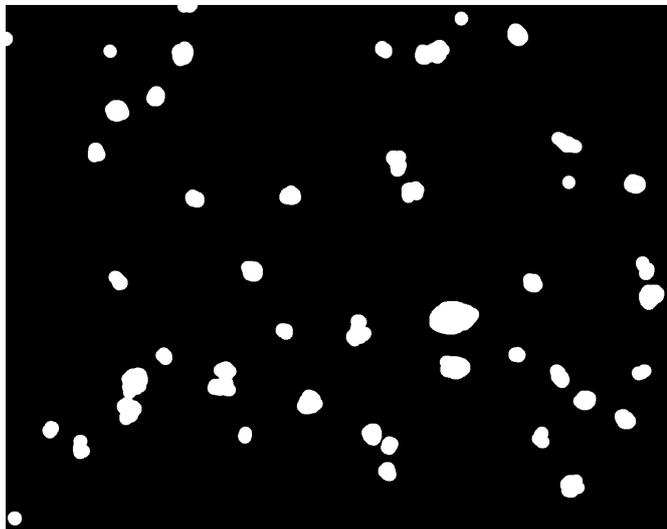
Step 4: Binarization of the shading corrected image with the determined threshold



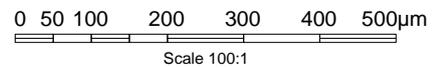
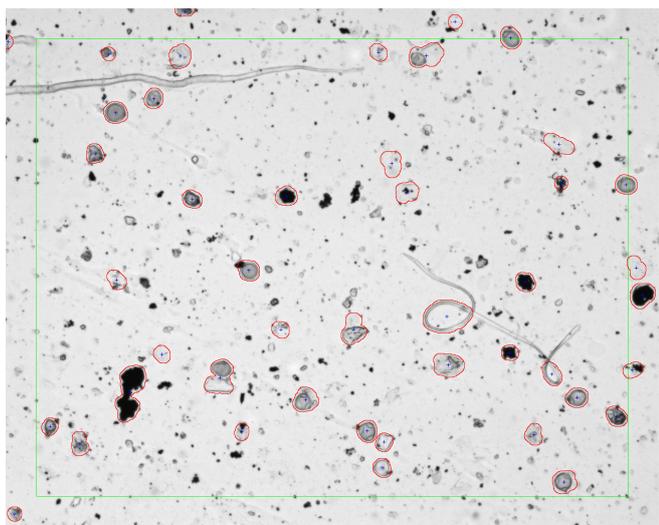
Step 5: Morphological fill-holes operation to completely mark pollen with transparent inner structures (e.g. *Taxus*^(D.12))



Step 6: Morphological Erosion and Dilation operation using a disc as structural element with the diameter of the smallest required particle. The erosion operation removes small dust and the rests of background noise, the subsequent dilation enlarges the masks back to their original size with the advantageous side effect, that the final masks have smooth borders.



Step 7: Connected component labeling to find the individual particles. The particle masks are displayed as red outlines. the blue crosses mark the center of each mask. To avoid cropped particles and to use the same reference area for small and big particles, only those particles are taken into account whose center is located within the green marked evaluation area. The size of the border area is determined by the radius of the largest particle.



8.1.1 Preprocessing of Data Sets with Particles in Different Planes

The 2D method described in the previous section worked reliably on manually prepared samples that are recorded with a 10x objective. Due to the cover slip most particles on these samples are located in a well defined plane. The depth of sharpness of the 10x objective is high enough such that it is nearly always possible to find one focal position where all pollen are well-focused.

In contrast to this the automatic preparation used in the pollenmonitor leaves the pollen grains in slightly different layers. Furthermore the used 20x objective has a much higher numerical aperture, such that the depth of sharpness is limited to a few micrometers. As a result it is quite unlikely to find all pollen well focused within one plane of the recorded stack (see figure 8.1a). To use the 2D method from the last section also in this environment, a preprocessing was developed in cooperation with Qing Wang, that searches a curved 2D plane in the 3D stack, which tries to cut all particles in their sharpest layer (see figure 8.1b).

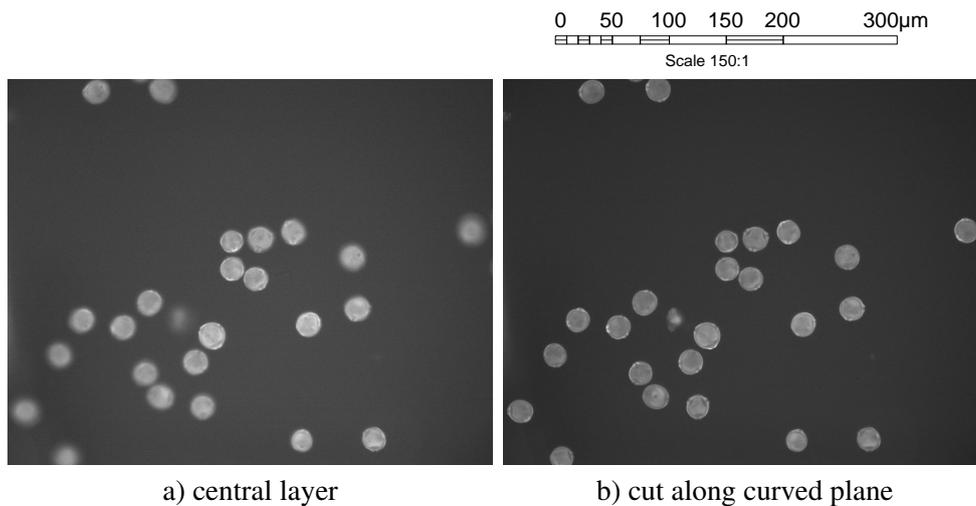


Figure 8.1: a) Central layer of a fluorescence image stack of a sample that was prepared and recorded analogously to the techniques used in the pollenmonitor. b) Cut along a curved plane, that was defined to best match the sharpest regions in the 3D stack

8.2 Detection with Voxel-Wise Invariants (MiSP Detector)

Due to the unsatisfactory results of the threshold-based morphological segmentation techniques, especially on highly cluttered samples or for agglomerated particles, a more robust alternative was required.

By ignoring some non-allergenic pollen taxa that have no spherical shape (like *Pinus*^(D.10) and *Picea*^(D.9)), we can robustly detect the remaining pollen taxa with the voxel-wise invariants using the sparse vectorial kernel function introduced in section 6.8.

To meet the timing requirements of the pollenmonitor in real-time operation (One air sample should be processed within one hour) we operate on down-scaled data and compute the invariants for all radii at once. The output is a 4D accumulator (radius, x, y, z) that contains a “probability” for each possible sphere within the given data set. The implementation is depicted in pseudo code in algorithm 8.1

The search for the most probable non-overlapping spheres was done in a first implementation just by a linear search for the global maximum in the accumulator, masking out the region around it (that belongs to overlapping spheres) and repeat this procedure until the global maximum is below a certain threshold (see algorithm 8.2).

Algorithm 8.1 MiSP detector sphere probs. = Detector(X)

Input: 3D array X **Output:** probabilities for all spheres

```

1: Down-scale  $X$  for faster processing
2: Initialize 4D accu to zero    // dimensions are radius, x, y, z
3: for all positions in  $X$  do
4:    compute gradient for current position
5:    if gradient magnitude > thresh then
6:     for all radii do
7:       center1 := current position + radius · gradient direction
8:       center2 := current position – radius · gradient direction
9:       accu( radius, center1) += gradient magnitude
10:      accu( radius, center2) += gradient magnitude
11:     end for
12:    end if
13: end for
14: for all radii do
15:    smooth accu(radius) in x,y,z direction by factor · radius
16: end for

```

Algorithm 8.2 Find Spheres 1 list of spheres = FindSpheres(accu4D)

Input: 4D accu (radius, x, y, z) with probabilities for each sphere**Output:** list of most probable non-overlapping spheres

```

1: repeat
2:    Find global maximum in 4D accu
3:    append radius and center to output sphere list
4:    for all positions in accu that belong to overlapping spheres do
5:      accu(position) := 0
6:    end for
7: until value of maximum < thresh

```

Due to the very large extents of the data sets (and correspondingly the even larger extents of the accumulator) this technique is quite slow. For the low number of spheres that are searched in the present application, a significant speedup can be reached by a two pass-algorithm, that first extracts a list of all candidates for a sphere by a local maximum criterion and the threshold criterion and then extracts the non-overlapping spheres from this list (see algorithm 8.3).

For the application of the MiSP detector on the transmitted light image stacks from the pollenmonitor data set, the only change is to set the z-component of the gradient to zero. This limits the voting for the sphere center to the same plane. This procedure is necessary, because the borders in z-direction of the pollen grain are hidden by the structures, that originate from the diffraction of the light (see figure 8.2)

Algorithm 8.3 Find Spheres 2 list of spheres = FindSpheres(accu4D)

Input: 4D accu (radius, x, y, z) with probabilities for each sphere

Output: list of most probable non-overlapping spheres

// Find all possible spheres as local maxima

1: Initialize local maxima list

2: **for all** positions in accu **do**

3: **if** accu(position) > thresh AND accu(position) is local maximum **then**

4: append value, radius and center to local maxima list

5: **end if**

6: **end for**

// Select most probable non-overlapping spheres

7: Sort local maxima list by value of maximum (highest to the begin)

8: **for all** entries in local maxima list **do**

9: **if** sphere with current radius and center does not overlap with any sphere in output sphere list **then**

10: append radius and center to output sphere list

11: **end if**

12: **end for**

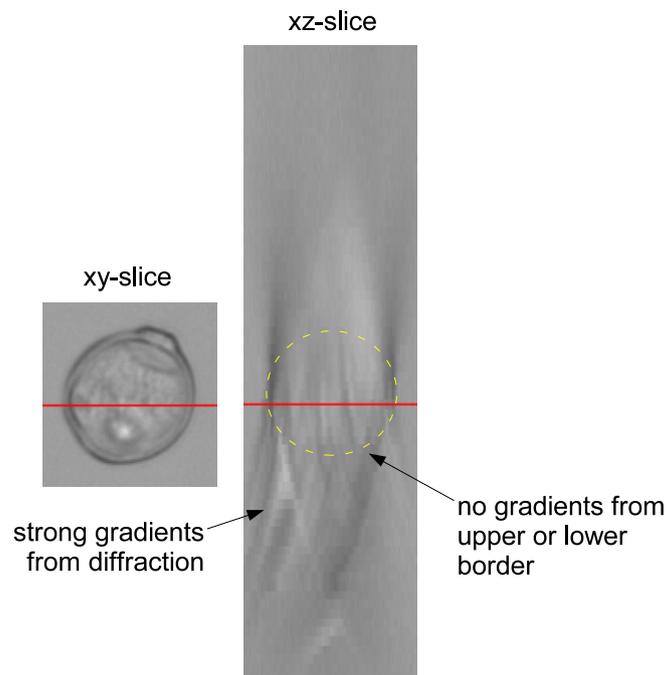
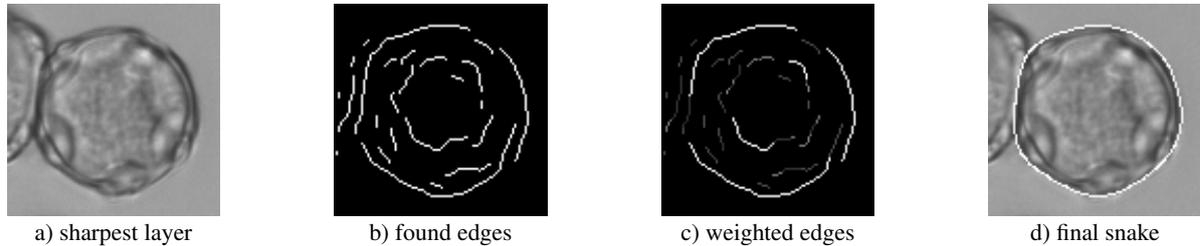


Figure 8.2: Orthogonal slices of a transmitted light stack from a *Betula*^(A.4) pollen grain. Only the gradients in x- and y- direction are used to find the object center. In z-direction the gradients are dominated by the diffraction patterns.

8.3 2D Boundary Extraction with Snakes

After the object centers and the radius of the corresponding sphere are determined, the precise border in the sharpest layer is searched. This part of the segmentation was mainly developed by Qing Wang, so only a short outline is presented here:

As parts of the object border are often missing or not clear, we use snakes to find a smooth and complete border. To avoid the common problem of snakes being attracted to undesired edges (if plain gradient magnitude is used as force field), we take the steps depicted in fig 8.3.



1. Applying modified Canny edge detection.

As pollen grains have a nearly round shape, the edges that are approximately perpendicular to the radial direction are more relevant. We replace the gradient with its radial component in the original Canny edge detection algorithm.

2. Model-based weighting of the edges.

The curvatures and relative locations of the edges are analyzed and each edge is given a different weight. Some edges are even eliminated. As a result, a much clearer weighted edge image is obtained.

3. Employing snakes to find the final border.

The initial contour is chosen to be the circle found in the detection step. The external force field is the so-called “gradient vector flow” (Xu and Prince, 1998) computed from the weighted edge image.

Figure 8.3: Segmentation of transmitted light microscopic images

8.4 3D Segmentation with Graph-Cuts

To find the 3D surface of the pollen grains in the confocal data set, we use the graph cut algorithm described in (Boykov and Kolmogorov, 2004). This technique needs two types of information as input:

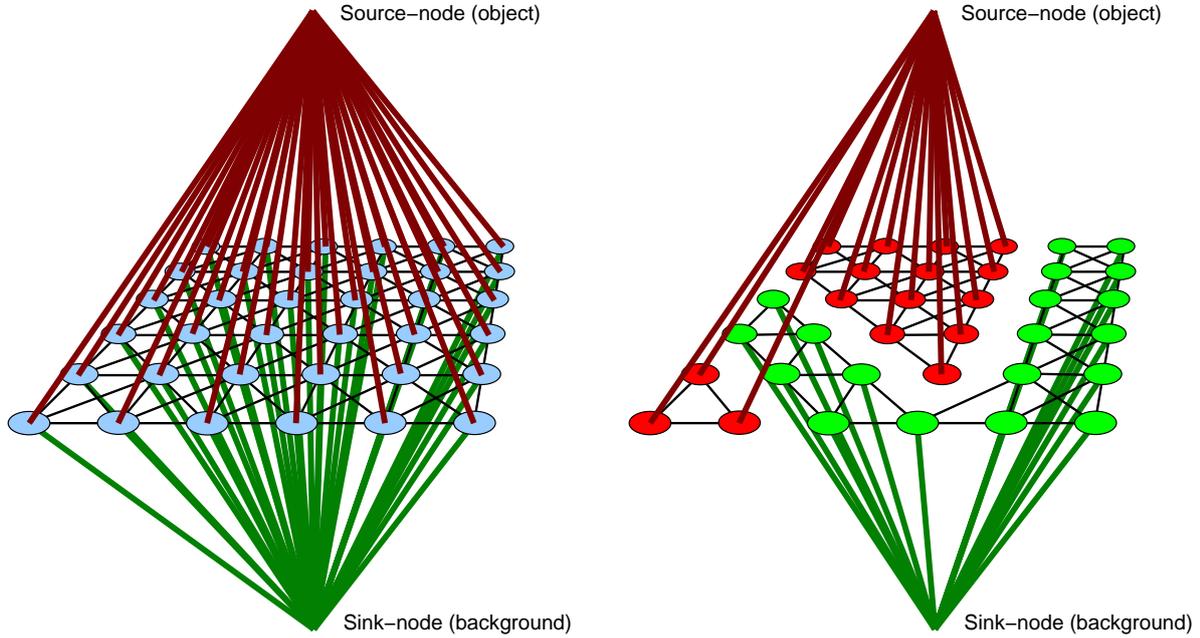
- For each voxel \mathbf{p} in the volume: A penalty $D_{\mathbf{p}}(L_{\mathbf{p}})$, if this voxel is assigned to object ($L_{\mathbf{p}} = 1$) or background ($L_{\mathbf{p}} = 0$). This corresponds to the probability that this voxel belongs to the object or not.
- For each pair of neighboring voxels (\mathbf{p}, \mathbf{q}) : A penalty $K_{(\mathbf{p}, \mathbf{q})}$ if they are assigned to the same label or to different labels. This corresponds to the probability that the object border lies between these two voxels.

For a given labeling L all these penalties in the volume are summed up to an “Energy”

$$E(L) = \sum_{\mathbf{p} \in P} D_{\mathbf{p}}(L_{\mathbf{p}}) + \sum_{(\mathbf{p}, \mathbf{q}) \in P \times P} K_{(\mathbf{p}, \mathbf{q})} \cdot \delta(L_{\mathbf{p}} \neq L_{\mathbf{q}}) . \quad (8.1)$$

The task of object segmentation can now be formulated as the search for the labeling, that minimizes this energy. While there are several heuristic approaches, like simulated annealing that may converge to a local minimum, the formulation of this problem as a min-cut of the corresponding graph allows to find the global optimum. For this, the voxels build the nodes in the graph. The object/background

penalties D_p are described by a weighted edge from each voxel to the “source”-node and a weighted edge to the “sink”-node of the graph. The voxel-to-voxel penalties are described directly with weighted edges between these voxels (see figure 8.4a). The graph-cut algorithm now searches for the cheapest



a) Initial graph for the graph-cut segmentation. The pixels of the image build the nodes. Each edge has a certain weight, corresponding to the probability that the pixel belongs to the object (red edges), to the background (green edges), or to the same class as its neighboring pixel (black edges).

b) Graph after graph-cut segmentation. The graph-cut algorithm has found the “cheapest” possibility to cut this graph into two parts. The pixels, connected to the source will be labeled as object and the pixels connected to the sink will be labeled as background.

Figure 8.4: Segmentation using the graph-cut algorithm

cut of this graph between the source-node and the sink node (see figure 8.4b). The pixels connected to the source will be labeled as object, and the pixels connected to the sink will be labeled as background. In contrast to the snake-based segmentation, this technique can also deal with disjoint objects, and arbitrary dimension of the data set (i. e., images and volumes).

Boykov and Kolmogorov (2004) used the graph-cut in an interactive environment, where the user has to label regions within the object and outside the object. From the histograms of these labeled regions the voxel-wise object/background penalties are computed.

For the application on the confocal data set a non-interactive segmentation is required: The probabilities for each gray value v to belong to background $p_{bg}(v)$ or to foreground $p_{fg}(v)$ were modelled with halved Gaussian distributions. The mean and standard deviation of the background distribution is defined as μ_{bg} and σ_{bg} , and for the foreground distribution as μ_{fg} and σ_{fg} . The values below μ_{bg} and above μ_{fg} were set to 1, resulting in

$$p_{bg}(v) = \begin{cases} 1 & \text{if } v < \mu_{bg} \\ \exp\left(-\frac{(v-\mu_{bg})^2}{2\sigma_{bg}^2}\right) & \text{if } v \geq \mu_{bg} \end{cases} \quad p_{fg}(v) = \begin{cases} \exp\left(-\frac{(v-\mu_{fg})^2}{2\sigma_{fg}^2}\right) & \text{if } v \leq \mu_{fg} \\ 1 & \text{if } v > \mu_{fg} \end{cases} \quad (8.2)$$

8 Detection and Segmentation

The probabilities that two neighboring voxels belong to the same class are also modeled with a Gaussian as

$$p(v_1, v_2) = \frac{1}{d} \exp\left(-\frac{(v_1 - v_2)^2}{2\sigma_v^2}\right), \quad (8.3)$$

where v_1, v_2 are the gray values of the two voxels and d is their spatial distance.

9 Specialized Invariants for Microscopical Spherical Particles (MiSP)

In our first approaches we used the HI framework only to make the features invariant to rotation and translation of the structures, and left the remaining variations to the classifier.

In real applications it is usually possible to identify several additional variations that do not change the meaning of the structures. For the recognition of objects in microscopical recordings, we can use the theoretical considerations about the microscopic imaging in chapter 2, e. g. aging effects of the fluorescence lamp, or variations of the light intensity due to absorption of out-of-focus dust changes the fluorescence intensities. So the variation of the mean fluorescence intensity does not give any valuable information about the pollen species, and it would be advantageous, if the invariants are insensitive to such variations.

In the same way we should take effects of variations within the preparation and the collection (e. g. the possible destruction of the natural fluorophores in the pollen grains due to sunlight or oxidation processes, or different absorption due to variations of the embedding media, etc.) into account, when we want to recognize pollen grains.

The most important prior knowledge that we use for pollen recognition, is their approximately spherical shape and their well-defined outer border. This is an important advantage, e. g., compared to recognition of “border-less” structures in biomedical tissue samples. Furthermore fresh pollen grains are quite stiff, such that they keep their spherical shape even under mechanical stress. This knowledge is used for the robust detection of pollen grains in cluttered samples and for the extraction of the precise outer contour in the sharpest of the transmitted light microscopic recordings, or of the surface in confocal recordings (see chapter 8).

In contrast to the general Euclidean invariants that sense only the global statistics of the structures, we can now use this outer shape to restrict the invariants to certain regions relative to this border.

This idea gave the greatest performance boost during our research on pollen recognition (especially in the avoidance of false positives) and can be considered as the most important step towards a reliable system for real-world samples.

9.1 The Transformation Group for MiSP Invariants

One of the most important challenges in biological object recognition is that we never find two identical objects. So even if we have a large reference data base with hundreds of example objects for each class, the most similar object to the test object will never be just a rotated and translated version of it.

In such environments an elastic registration of the test object with the most similar reference object seems to be the best way for a meaningful comparison. The similarity measure could then be based on the pixel-wise similarities of the test object with the deformed reference object and the “magnitude” of the deformation that was needed to match the two objects (see figure 9.1). The computational costs of the typical elastic registration approaches, the problems of local optima and the problems, when trying to register two objects of different classes make such an approach impossible for pollen recognition in a routine environment.

So our goal is to extract features from the objects such that the distance between two objects in the feature space best matches the above described similarity measure, i. e. the distance should be small, if object 2 is just a deformed version of object 1, and it should be large, if no deformation exists, that transforms object 1 into object 2.

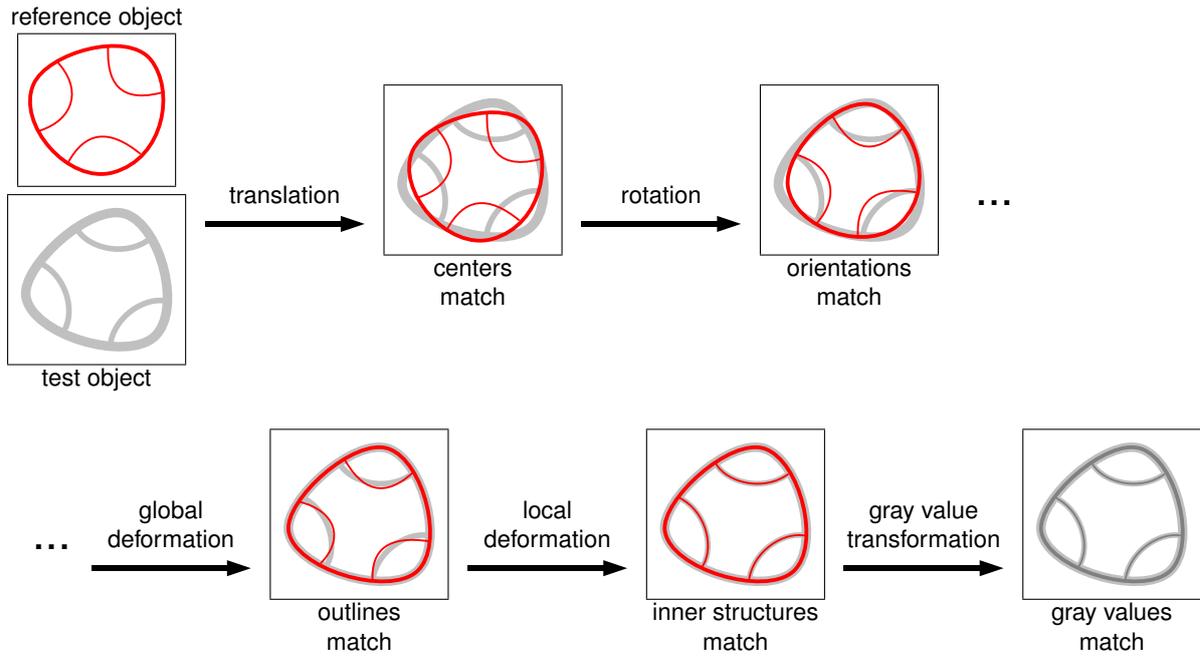


Figure 9.1: Hypothetical elastic registration of two pollen grains for a meaningful comparison. Instead of trying to find all needed transformation parameters, we construct features that are invariant or at least robust to all depicted transformations.

These requirements can be fulfilled by defining a transformation group that describes such deformations and inserting it into the HI framework. By the use of synthetic channels and the probability-weighted Haar-integration introduced in section 6.1 we can construct such features.

9.1.1 Invariance to Translation by Normalization

Invariance to translations is achieved by moving the center of mass of the segmentation mask to the origin. The final features are quite insensitive to errors in this normalization step, because they are computed “far” away from this center and only the direction to the center (not the distance) is used.

9.1.2 Invariance to Rotation.

Invariance to rotation around the object center with

$$\mathbf{x}' = \mathbf{R}\mathbf{x} , \tag{9.1}$$

where \mathbf{R} is a rotation matrix, is achieved by integration over the rotation group. In the confocal data set the recording properties in xy - and z -direction are similar enough, such that we can model a 3D rotation of a real-world object by a 3D rotation of the recorded volumetric data set (see figure 9.2b).

The transmitted light microscopic image stacks show very different characteristics in xy - and z -direction, (see figure 9.2c). A rotation around the x - or y -axis of the real-world object results in so different gray value distributions, that it is more reasonable to model only the rotation around the z -axis for the construction of the invariants and “learn” the effects of the other rotations by providing enough training objects at different orientations.

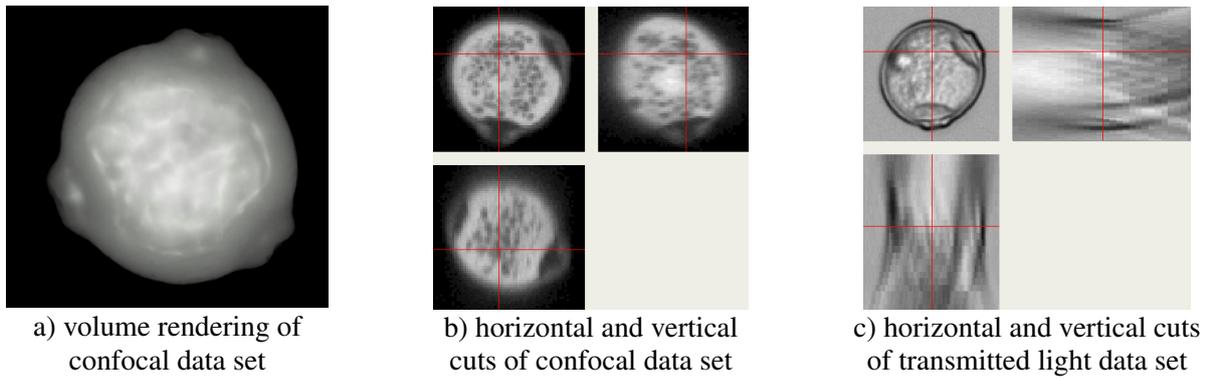


Figure 9.2: 3D recordings of a *Betula*^(A.4) pollen. In transmitted light microscopy the recording properties in z-direction (the direction of the optical axis) are significantly different from those in the xy-direction, because the effects of diffraction, refraction and absorption depend on the direction of the transmitted light. Furthermore there is a significant loss of information in z-direction due to the low-pass property of the optical transfer function

9.1.3 Selection of an Appropriate Model for the Deformations

The integration of deformation models into the feature extraction was one of the most important steps during our research on automated pollen recognition. However, there is no straightforward way for this implementation, because (compared to rotation or translation) no unique physically motivated definition of possible deformations exist. Of course, for the observation of one individual object under different deformations, one might find physically motivated models (e. g. by preserving the volume, defining different elasticities for the different structures, etc.), but when it comes to the comparison between two different individuals, none of these conditions applies any more.

The first goal of a deformation model in the feature extraction is therefore not to define a model, that is valid for the deformation of one individual, but to define a model that allows to deform the different objects in such a way, that corresponding structures are sensed by the same features.

For the pollen recognition, we want to use the outer surface of the object to define the initial correspondences between two individual objects. This outer surface is well defined for all pollen grains, and it has already been extracted in the segmentation step. Because we deal here with spherical, star-shaped particles, and because we have already defined a rotation around the center as a possible transformation, it is reasonable to define a first global deformation model, that

- is perpendicular to the action of the rotation operator,
- is able to match any star-shaped surface to any other, and
- has a low number of free parameters.

We find that a radial displacement of each surface-point fulfills this condition. E. g., if the radial positions of the surface of object 1 are parameterized in spherical coordinates by $r = s_1(\varphi, \vartheta)$ and the surface of object 2 by $s_2(\varphi, \vartheta)$, then the deformation model can be parameterized as $\gamma(\varphi, \vartheta)$ such that

$$s_2(\varphi, \vartheta) = s_1(\varphi, \vartheta) + \gamma(\varphi, \vartheta) . \quad (9.2)$$

The next question that arises is how this deformation should act on the remaining (non-surface-) points of the object. If we want to avoid an introduction of further parameters, we find two alternatives. The first alternative is to approximate this deformation by a radial translation, such that for each point

9 Specialized Invariants for Microscopical Spherical Particles (MiSP)

$(\varphi_1, \vartheta_1, r_1)$ in the first object the, corresponding point $(\varphi_2, \vartheta_2, r_2)$ in the second object is defined by

$$\begin{aligned}\varphi_2 &= \varphi_1 \\ \vartheta_2 &= \vartheta_1 \\ r_2 &= r_1 + \gamma(\varphi_1, \vartheta_1) .\end{aligned}\tag{9.3}$$

The second alternative is to approximate this deformation by a radial scaling, such that

$$\begin{aligned}\varphi_2 &= \varphi_1 \\ \vartheta_2 &= \vartheta_1 \\ r_2 &= r_1 \cdot \gamma_s(\varphi_1, \vartheta_1) ,\end{aligned}\tag{9.4}$$

where $\gamma_s(\varphi, \vartheta)$ is determined in analogy to $\gamma(\varphi, \vartheta)$ such that it deforms the surface of object 1 to the surface of object 2 by a scaling,

$$s_2(\varphi, \vartheta) = s_1(\varphi, \vartheta) \cdot \gamma_s(\varphi, \vartheta) .\tag{9.5}$$

It depends on the particular application, which of these two approaches provides the better modeling for the corresponding structures of two individuals depends on the application. For pollen recognition, it seems that the correspondences are better found by a radial translation, because the size of the corresponding structures seem to be largely independent of the overall size of the pollen (see figure 9.3).

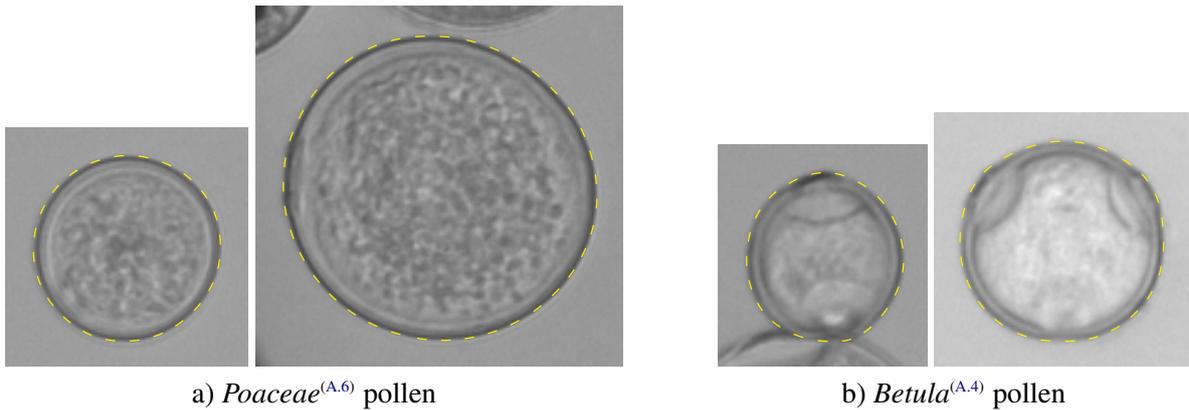


Figure 9.3: Small and big pollen grains of *Poaceae*^(A.6) (the most important allergenic taxa) and *Betula*^(A.4). The corresponding structures (the thickness of the exine (the outer border), the size of porates, and the granularity of the inner texture) have nearly the same size, independent of the overall size of the pollen

After such a radial deformation, the corresponding structures besides the segmentation border are only roughly localized. A reasonable assumption is that the uncertainty of the position of such a structure increases with its distance from the reliably corresponding points on the border. This will be described by an additional arbitrary deformation model, which allows any deformation. For the final Haar integration, we will have to specify probabilities for the occurrence of a certain realization of this deformation field.

9.1.4 Integration of the Deformation Models into the Transformation Group

As motivated in the previous section, the deformation model consists of two parts. The global deformations are modeled by a shift in radial direction e_r , which depends only on the angular coordinates (see figure 9.4a).

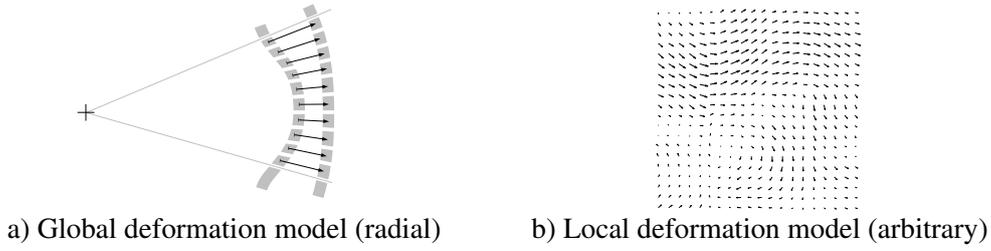


Figure 9.4: Possible realizations of the deformation models

For arbitrary rotations, described in spherical coordinates $\mathbf{x} = (x_r, x_\varphi, x_\vartheta)$, this model is

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\gamma}(\mathbf{x}) \quad \text{with} \quad \boldsymbol{\gamma}(\mathbf{x}) = \gamma(x_\varphi, x_\vartheta) \cdot \mathbf{e}_r(x_\varphi, x_\vartheta) . \quad (9.6)$$

For rotations around the z-axis described in cylindrical coordinates $\mathbf{x} = (x_r, x_\varphi, x_z)$ we get

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\gamma}(\mathbf{x}) \quad \text{with} \quad \boldsymbol{\gamma}(\mathbf{x}) = \gamma(x_\varphi) \cdot \mathbf{e}_r(x_\varphi) . \quad (9.7)$$

Please note, that this deformation is well defined only for $r > -\gamma(\varphi)$, which is no problem in the present application, because the features are computed “far” away from the center. To keep the Haar integrals over this deformations finite, we later will introduce a synthetic channel, that defines the integration limits according to the segmentation borders.

The smaller local deformations are described by an arbitrary displacement field $\mathbf{D}(\mathbf{x})$ (see figure 9.4b)

$$\mathbf{x}' = \mathbf{x} + \mathbf{D}(\mathbf{x}) . \quad (9.8)$$

For the later partial Haar-integration (Haasdonk et al., 2004) over all possible realizations of this displacement field, it is sufficient to know only the probability of the occurrence of a certain relative displacement \mathbf{r} within this field as

$$p(\mathbf{D}(\mathbf{x} + \mathbf{d}) - \mathbf{D}(\mathbf{x}) = \mathbf{r}) = p_d(\mathbf{r}; \|\mathbf{d}\|) \quad \forall \mathbf{x}, \mathbf{d} \in \mathbb{R}^3 \quad (9.9)$$

where we select $p_d(\mathbf{r}; \|\mathbf{d}\|)$ to be a rotationally symmetric Gaussian distribution with a standard deviation $\sigma = \|\mathbf{d}\| \cdot \sigma_d$,

$$p_d(\mathbf{r}; \|\mathbf{d}\|) = \frac{1}{\left(\|\mathbf{d}\| \sigma_d \sqrt{2\pi}\right)^3} \cdot \exp\left(\frac{-\|\mathbf{r}\|^2}{2\left(\|\mathbf{d}\| \sigma_d\right)^2}\right) . \quad (9.10)$$

While we achieve full invariance to radial deformations by full Haar-integration, we only can reach robustness to local deformations by the probability-weighted Haar-integration. But this non-invariance in the second case is exactly the desired behavior. In combination with appropriate kernel functions this results in a continuous mapping of objects (with weak or strong local deformations) into the feature space.

9.2 The Kernel Functions for MiSP Invariants

9.2.1 Robustness to Arbitrary Gray Value Transformations.

A simple but very powerful method to become robust to arbitrary gray value transformations is to operate not on the raw gray values but on the gradient direction and on the gradient magnitude instead. This has been widely known since the great success of the SIFT features (Lowe, 2004) at the latest.

This splitting of the information into gradient magnitude and gradient direction was already successfully used within the HI framework for the construction of robust invariants in (Schulz, 2005), (Schulz et al., 2006) and (Reisert and Burkhardt, 2006).

For a deeper understanding, we will carry out a detailed analysis of its invariance properties of this approach here.

As described in chapter 2 there are multiple sources that cause certain transformations of the raw gray values. The simplest model assumes a constant factor α and an offset β , that transforms each gray value v to

$$v_r = \alpha \cdot v + \beta . \quad (9.11)$$

A typical source for an offset is scattered light, that reaches the CCD sensor. A factor α may be introduced by absorption effects, aging of the used light sources, misadjustment of the camera parameters, the AD converter, etc.

Beside these simple affine gray-value transformations, we find several sources for nonlinear gray-value changes, e. g. effects of bleaching, or just the intra-class variations of the various structures within a pollen grain. This can be modeled by an arbitrary nonlinear gray value transformation f_ℓ (ℓ stands for lookup table)

$$v_r = f_\ell(v) . \quad (9.12)$$

To become robust to such gray value transformations, we extract the gradient magnitude G_m and the gradient direction \mathbf{G}_d at each position of the data set:

$$G_m(X) := \|\nabla X\| \quad (9.13)$$

$$\mathbf{G}_d(X) := \frac{\nabla X}{\|\nabla X\|} . \quad (9.14)$$

Both of these quantities are invariant to gray value offsets β , because they are based on derivatives. The gradient direction is invariant to a scaling of the gray values α

$$\mathbf{G}_d(\alpha \cdot X) = \frac{\nabla(\alpha \cdot X)}{\|\nabla(\alpha \cdot X)\|} = \frac{\alpha \cdot \nabla X}{\alpha \cdot \|\nabla X\|} = \frac{\nabla X}{\|\nabla X\|} = \mathbf{G}_d(X) \quad (9.15)$$

while the gradient magnitude operator commutes with the scaling of gray values,

$$G_m(\alpha \cdot X) = \alpha \cdot G_m(X) . \quad (9.16)$$

This property will be used later in the post-processing step to make the features invariant to gray value scaling.

The most important property of this transformation is that the gradient direction is invariant to an arbitrary nonlinear gray value transformation f_ℓ as long as this mapping is strictly monotonic increasing,

$$\dot{f}_\ell(v) > 0 \quad \forall v \in \mathbb{R} \quad (9.17)$$

The proof is quite intuitive: the gradient direction is always perpendicular to the iso-lines (in 2D) or iso-surfaces (in 3D) in the given data set. As the gray value mapping is independent of the position in the data set, all points with gray value v that build the iso-surface in the original data set will be mapped to the new gray value $f_\ell(v)$, and build the identical iso-surface in the transformed data set.

The mathematical proof is also straightforward. By defining a new coordinate system with the e_1 -axis in the direction of the gradient, e_2 and e_3 perpendicular to it,

$$\begin{aligned} \mathbf{e}_1 &= \frac{\nabla X(\mathbf{x})}{\|\nabla X\|(\mathbf{x})} \\ \mathbf{e}_2 &\perp \mathbf{e}_1 \\ \mathbf{e}_3 &\perp \mathbf{e}_2 \wedge \mathbf{e}_3 \perp \mathbf{e}_1 \end{aligned} \quad (9.18)$$

the partial derivatives at position \mathbf{x} in this new coordinate system disappear in direction \mathbf{e}_2 and \mathbf{e}_3 ,

$$\begin{aligned} \frac{\partial X}{\partial e_1}(\mathbf{x}) &= \|\nabla X\|(\mathbf{x}) \\ \frac{\partial X}{\partial e_2}(\mathbf{x}) &= 0 \\ \frac{\partial X}{\partial e_3}(\mathbf{x}) &= 0 . \end{aligned} \tag{9.19}$$

Using the chain rule, we see that in the image with transformed gray values the derivatives in direction \mathbf{e}_2 and \mathbf{e}_3 also disappear,

$$\begin{aligned} \frac{\partial f_\ell(X)}{\partial e_1}(\mathbf{x}) &= \dot{f}_\ell(X(\mathbf{x})) \cdot \frac{\partial X}{\partial e_1}(\mathbf{x}) = \dot{f}_\ell(X(\mathbf{x})) \cdot \|\nabla X\|(\mathbf{x}) \\ \frac{\partial f_\ell(X)}{\partial e_2}(\mathbf{x}) &= \dot{f}_\ell(X(\mathbf{x})) \cdot \frac{\partial X}{\partial e_2}(\mathbf{x}) = 0 \\ \frac{\partial f_\ell(X)}{\partial e_3}(\mathbf{x}) &= \dot{f}_\ell(X(\mathbf{x})) \cdot \frac{\partial X}{\partial e_3}(\mathbf{x}) = 0 , \end{aligned} \tag{9.20}$$

and we can see immediately, why the mapping must be strictly monotonic increasing.

9.2.2 Synthetic Channels with Segmentation Results.

To feed the segmentation information into the HI framework we simply render the surface (confocal data set) or the contour of the sharpest layer (transmitted light data set) as delta-peaks into a new channel S and extend the kernel-function with two additional points that sense the gray value in this channel (see fig 9.5)

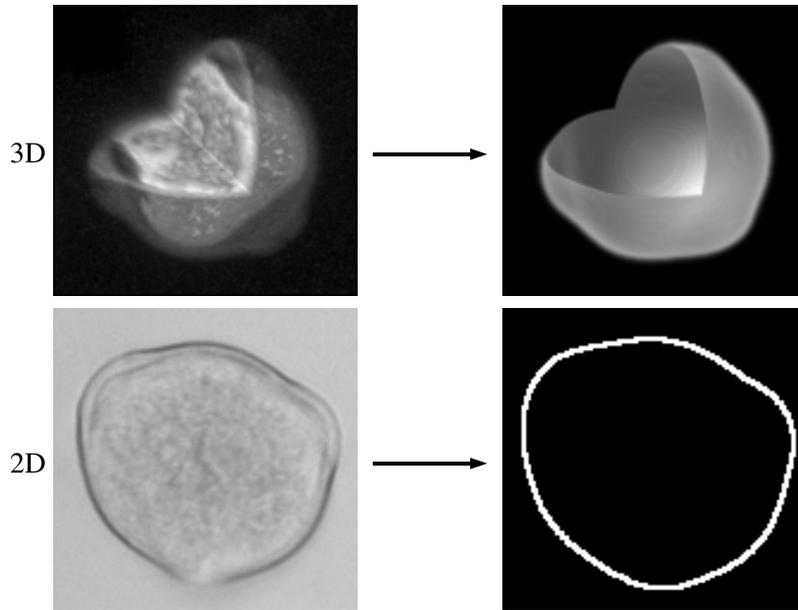


Figure 9.5: Example for the synthetic channels containing the segmentation information. Note that in the present application these channels are not computed voxel-by-voxel, but are described by an appropriate parameterization.

The only condition for this technique is that the computation of the synthetic channel and the action of the transformation group can be exchanged without the result being changed (see section 6.5)

9.2.3 Resulting Kernel Functions.

To achieve the requested properties we construct 4-point kernels. 2 points \mathbf{a}_1 and \mathbf{a}_2 of the kernel sense the segmentation channel and the other 2 points $\mathbf{b}_1 = \mathbf{a}_1 + \mathbf{q}_1$ and $\mathbf{b}_2 = \mathbf{a}_2 + \mathbf{q}_2$ sense the gradient ∇X of the gray values,

$$k_1[\mathbf{p}](S, X) = S(\mathbf{a}_1) \cdot \|\nabla X\|(\mathbf{b}_1) \cdot \delta\left(c_1 - \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{b}_1)\right) \cdot S(\mathbf{a}_2) \cdot \|\nabla X\|(\mathbf{b}_2) \cdot \delta\left(c_2 - \frac{\mathbf{a}_2}{\|\mathbf{a}_2\|} \cdot \frac{\nabla X}{\|\nabla X\|}(\mathbf{b}_2)\right), \quad (9.21)$$

while the delta-functions restrict the kernel to “see” only gradients with the given “directions” c_1 and c_2 , respectively. Not all combinations of \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{q}_1 , \mathbf{q}_2 , c_1 and c_2 make sense, because the Haar integration returns identical features for all kernels which are equivalent under the given transformation group. Furthermore for certain combinations, only trivial features will be returned. To ensure, that only non-trivial and non-identical features are created, we will introduce a low-dimensional parameterization \mathbf{p} later. Examples of this kernel function are depicted in figure 9.6.

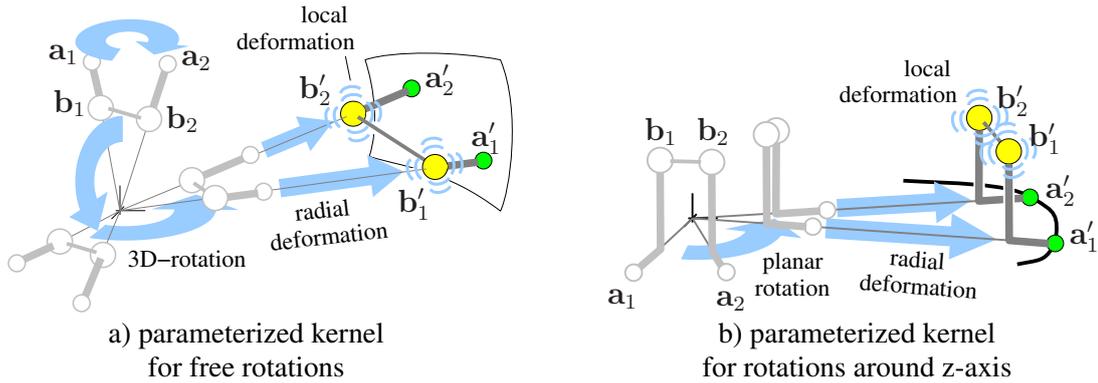


Figure 9.6: 4-point kernel function to extract the “structural” MiSP invariants. According to the transformation group, the kernels are rotated, radially deformed and locally deformed. Only at such transformations, where both kernel points \mathbf{a}_1 and \mathbf{a}_2 hit a segmentation surface (or another signal) in the synthetic channel, the gradients of the dataset are “sensed” at \mathbf{b}_1 and \mathbf{b}_2 and contribute to the current Haar integral

The sensing of the gradients with a certain direction is illustrated in figure 9.7.

As mentioned above, the resulting “structural” features are fully invariant to the global radial deformations. To extract the shape information, we select the parameterized kernel

$$k_2[\mathbf{a}_1, \mathbf{a}_2](S) = \|\boldsymbol{\gamma}(\mathbf{a}_1)\| S(\mathbf{a}_1) \cdot \|\boldsymbol{\gamma}(\mathbf{a}_2)\| S(\mathbf{a}_2), \quad (9.22)$$

that operates on the synthetic channel (see figure 9.8 for an example). When we use the simple scheme of creating the synthetic channels described above, the resulting features extract the same information as the magnitudes of the Fourier coefficients of the contour in the 2D case and the magnitudes of the spherical harmonic coefficients of the surface in the 3D case.

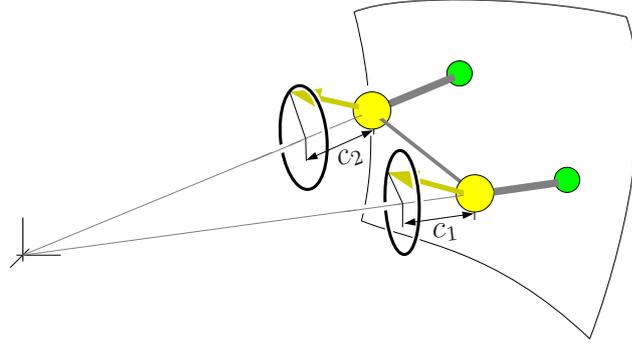


Figure 9.7: Illustration of the “sensing” of the gradients by the kernel points. Only at those positions, where the projection of the normalised gradients match the given c_1 and c_2 , the gradient magnitudes contribute to the current Haar integral

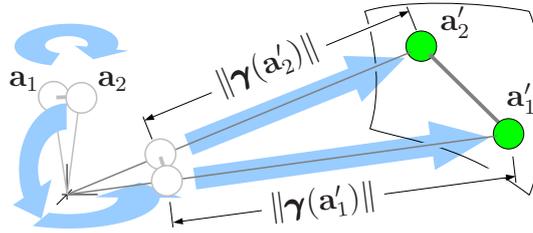


Figure 9.8: 2-point kernel function to extract the “shape” MiSP invariants from the synthetic channel. The transformation group contains only rotation and radial deformation.

9.3 Resulting Haar Integral

With the group of radial deformations $G_{\boldsymbol{\gamma}}$, the group of arbitrary deformations $G_{\mathbf{D}}$ and the group of rotations $G_{\mathbf{R}}$ the final Haar integral becomes:

$$T = \int_{G_{\mathbf{R}}} \int_{G_{\boldsymbol{\gamma}}} \int_{G_{\mathbf{D}}} f(g_{\mathbf{R}} g_{\boldsymbol{\gamma}} g_{\mathbf{D}} S, g_{\mathbf{R}} g_{\boldsymbol{\gamma}} g_{\mathbf{D}} X) p(\mathbf{D}) dg_{\mathbf{D}} dg_{\boldsymbol{\gamma}} dg_{\mathbf{R}} . \quad (9.23)$$

where $p(\mathbf{D})$ is the probability of the occurrence of the local displacement field \mathbf{D} . The transformation of the data set is described by $(gX)(\mathbf{x}) =: X(\mathbf{x}')$, where

$$\mathbf{x}'(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D}) = \underbrace{\mathbf{R}\mathbf{x}}_{\text{rotation}} + \underbrace{\boldsymbol{\gamma}(\mathbf{R}\mathbf{x})}_{\text{global deformation}} + \underbrace{\mathbf{D}(\mathbf{R}\mathbf{x} + \boldsymbol{\gamma}(\mathbf{R}\mathbf{x}))}_{\text{local deformation}} . \quad (9.24)$$

To ensure a strong coupling of \mathbf{a}'_1 and \mathbf{b}'_1 (and $\mathbf{a}'_2, \mathbf{b}'_2$ accordingly) we only use kernels, where these two points will be treated equally by the global transformation, i. e., the kernel must fulfill the condition $\boldsymbol{\gamma}(\mathbf{R}\mathbf{a}_i) = \boldsymbol{\gamma}(\mathbf{R}\mathbf{b}_i), \forall \mathbf{R}$ (illustrated by thick connections in fig 9.9).

The transformed kernel points are

$$\begin{aligned} \mathbf{a}'_i(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D}) &= \mathbf{R}\mathbf{a}_i + \boldsymbol{\gamma}(\mathbf{R}\mathbf{a}_i) + \mathbf{D}(\mathbf{R}\mathbf{a}_i + \boldsymbol{\gamma}(\mathbf{R}\mathbf{a}_i)) \\ \mathbf{b}'_i(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D}) &= \mathbf{R}\mathbf{b}_i + \boldsymbol{\gamma}(\mathbf{R}\mathbf{a}_i) + \mathbf{D}(\mathbf{R}\mathbf{b}_i + \boldsymbol{\gamma}(\mathbf{R}\mathbf{a}_i)) . \end{aligned} \quad (9.25)$$

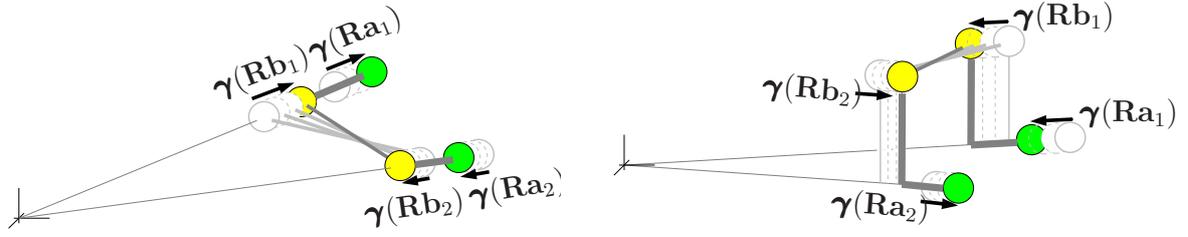


Figure 9.9: The kernel points \mathbf{b}_1 and \mathbf{b}_2 must be selected such that they are treated equally by the global deformation as \mathbf{a}_1 and \mathbf{a}_2 : For the spherical deformation model they must be co-linear; for the cylindrical deformation model they must have an identical x_φ (when expressed in cylinder coordinates). This property is illustrated here by a stiff coupling of the two points

Now inserting the kernel into the Haar integral gives:

$$\begin{aligned}
 T = & \int_{G_{\mathbf{R}}} \int_{G_{\boldsymbol{\gamma}}} \int_{G_{\mathbf{D}}} \\
 & S\left(\mathbf{a}'_1(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right) \cdot \|\nabla_{\mathbf{x}'} X\| \left(\mathbf{b}'_1(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right) \cdot \delta\left(c_1 - \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|} \left(\mathbf{b}'_1(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right)\right) \\
 & \cdot S\left(\mathbf{a}'_2(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right) \cdot \|\nabla_{\mathbf{x}'} X\| \left(\mathbf{b}'_2(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right) \cdot \delta\left(c_2 - \frac{\mathbf{a}_2}{\|\mathbf{a}_2\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|} \left(\mathbf{b}'_2(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right)\right) \\
 & \cdot p(\mathbf{D}) \\
 & \cdot d\mathbf{D}\left(\mathbf{Ra}_1 + \boldsymbol{\gamma}(\mathbf{Ra}_1)\right) \cdot d\mathbf{D}\left(\mathbf{Rb}_1 + \boldsymbol{\gamma}(\mathbf{Ra}_1)\right) \\
 & \cdot d\mathbf{D}\left(\mathbf{Ra}_2 + \boldsymbol{\gamma}(\mathbf{Ra}_2)\right) \cdot d\mathbf{D}\left(\mathbf{Rb}_2 + \boldsymbol{\gamma}(\mathbf{Ra}_2)\right) \\
 & \cdot d\boldsymbol{\gamma}(\mathbf{Ra}_1) \cdot d\boldsymbol{\gamma}(\mathbf{Ra}_2) \\
 & \cdot d\mathbf{R} \ ,
 \end{aligned} \tag{9.26}$$

where $\nabla_{\mathbf{x}'}$ denotes the del operator in the transformed coordinate system. The uncommon notation like $d\boldsymbol{\gamma}(\mathbf{Ra}_1)$ is necessary, because each displacement field is described here by an infinite number of parameters (one displacement for each location in the 3D space). During the integration the outer integral continuously “selects” the integration parameter for the inner integral.

This is the most general form of the integral allowing any type of synthetic channel, that is created according to (6.5).

9.4 Fast Simultaneous Computation of the Invariants

For the synthetic channels in the MiSP invariants that are created from a single surface or contour, we can significantly simplify the integral if we can assume a star-shaped object (which is granted for all considered pollen types). For a star-shaped object we will find only one nonzero response of S for each \mathbf{Ra}_i during the integration over all global deformations $\boldsymbol{\gamma}(\mathbf{Ra}_i)$. By parameterizing this surface as $\mathbf{s}(\mathbf{Ra}_i)$ such that

$$\begin{aligned}
 S\left(\mathbf{a}'_i(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D})\right) &= \delta\left(\mathbf{a}'_i(\mathbf{R}, \boldsymbol{\gamma}, \mathbf{D}) - \mathbf{s}(\mathbf{Ra}_i)\right) \\
 &= \delta\left(\mathbf{Ra}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i) + \mathbf{D}\left(\mathbf{Ra}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i)\right) - \mathbf{s}(\mathbf{Ra}_i)\right) \ .
 \end{aligned} \tag{9.27}$$

So for a nonzero response of the synthetic channel, $\boldsymbol{\gamma}(\mathbf{Ra}_i)$ must be

$$\boldsymbol{\gamma}(\mathbf{Ra}_i) = \mathbf{s}(\mathbf{Ra}_i) - \mathbf{Ra}_i - \mathbf{D}(\mathbf{Ra}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i)) . \quad (9.28)$$

For this specific $\boldsymbol{\gamma}(\mathbf{Ra}_i)$ the transformed kernel points \mathbf{b}'_i are located at:

$$\mathbf{b}'_i = \mathbf{Rb}_i + \mathbf{s}(\mathbf{Ra}_i) - \mathbf{Ra}_i - \mathbf{D}(\mathbf{Ra}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i)) + \mathbf{D}(\mathbf{Rb}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i)) . \quad (9.29)$$

Now the positions only depend on relative displacements of the local deformation field. With the additional precondition that the probability for the occurrence of a certain relative displacement \mathbf{r} only depends on the distance of the two considered points (9.9) we can fully eliminate the dependency of the global transformation $\boldsymbol{\gamma}(\mathbf{Ra}_i)$ and replace the four integrals over the local displacement field $p(\mathbf{D}) \cdot d\mathbf{D}(\mathbf{Ra}_1 + \boldsymbol{\gamma}(\mathbf{Ra}_1)) \cdot d\mathbf{D}(\mathbf{Rb}_1 + \boldsymbol{\gamma}(\mathbf{Ra}_1)) \cdot d\mathbf{D}(\mathbf{Ra}_2 + \boldsymbol{\gamma}(\mathbf{Ra}_2)) \cdot d\mathbf{D}(\mathbf{Rb}_2 + \boldsymbol{\gamma}(\mathbf{Ra}_2))$ by the integration over all relative deformations weighted with their probability $p_d(\mathbf{r}_1; \|\mathbf{d}_1\|) p_d(\mathbf{r}_2; \|\mathbf{d}_2\|) d\mathbf{r}_1 d\mathbf{r}_2$. The resulting \mathbf{b}'_i is

$$\begin{aligned} \mathbf{b}'_i(\mathbf{R}, \mathbf{s}, \mathbf{r}_i) &= \mathbf{s}(\mathbf{Ra}_i) + \mathbf{Rb}_i - \mathbf{Ra}_i + \mathbf{r}_i \\ &= \mathbf{s}(\mathbf{Ra}_i) + \mathbf{R} \cdot (\mathbf{b}_i - \mathbf{a}_i) + \mathbf{r}_i , \end{aligned} \quad (9.30)$$

with the probability

$$\begin{aligned} p_d(\mathbf{r}_i; \|\mathbf{Rb}_i + \boldsymbol{\gamma}(\mathbf{Ra}_i) - \mathbf{Ra}_i - \boldsymbol{\gamma}(\mathbf{Ra}_i)\|) &= p_d(\mathbf{r}_i; \|\mathbf{Rb}_i - \mathbf{Ra}_i\|) \\ &= p_d(\mathbf{r}_i; \|\mathbf{b}_i - \mathbf{a}_i\|) . \end{aligned} \quad (9.31)$$

By using the definition of $\mathbf{q}_i := \mathbf{b}_i - \mathbf{a}_i$ we finally get

$$\mathbf{b}'_i(\mathbf{R}, \mathbf{r}_i) = \mathbf{s}(\mathbf{Ra}_i) + \mathbf{Rq}_i + \mathbf{r}_i , \quad (9.32)$$

with the probability

$$p_d(\mathbf{r}_i; \|\mathbf{q}_i\|) . \quad (9.33)$$

Now the full Haar integral can be written as

$$\begin{aligned} T &= \int_{G_{\mathbf{R}}} \int_{\mathbb{R}^3} \|\nabla_{\mathbf{x}'} X\|(\mathbf{b}'_1(\mathbf{R}, \mathbf{r}_1)) \cdot \delta\left(c_1 - \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|}(\mathbf{b}'_1(\mathbf{R}, \mathbf{r}_1))\right) p_d(\mathbf{r}_1; \|\mathbf{q}_1\|) d\mathbf{r}_1 \\ &\quad \cdot \int_{\mathbb{R}^3} \|\nabla_{\mathbf{x}'} X\|(\mathbf{b}'_2(\mathbf{R}, \mathbf{r}_2)) \cdot \delta\left(c_2 - \frac{\mathbf{a}_2}{\|\mathbf{a}_2\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|}(\mathbf{b}'_2(\mathbf{R}, \mathbf{r}_2))\right) p_d(\mathbf{r}_2; \|\mathbf{q}_2\|) d\mathbf{r}_2 d\mathbf{R} . \end{aligned} \quad (9.34)$$

Due to the restriction that \mathbf{a}_1 and \mathbf{b}_1 (and therefore also \mathbf{q}_1) are colinear in the case of 3D rotations, the first term in the integral returns a scalar value that is only dependent on the resulting direction \mathbf{Ra}_1 and not on the rotation parameters themselves. For a selected parameter set, it can therefore be replaced by a precomputed scalar function on a sphere with radius $\|\mathbf{q}_1\|$

$$\begin{aligned} X_1(\mathbf{Rq}_1) &:= \int_{\mathbb{R}^3} \|\nabla_{\mathbf{x}'} X\|(\mathbf{b}'_1(\mathbf{R}, \mathbf{r}_1)) \cdot \delta\left(c_1 - \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|}(\mathbf{b}'_1(\mathbf{R}, \mathbf{r}_1))\right) p_d(\mathbf{r}_1; \|\mathbf{q}_1\|) d\mathbf{r}_1 \\ &\quad \forall \mathbf{R} \in G_{\mathbf{R}} , \end{aligned} \quad (9.35)$$

and correspondingly for the second term we can find a

$$X_2(\mathbf{R}\mathbf{q}_2) := \int_{\mathbb{R}^3} \|\nabla_{\mathbf{x}'} X\|(\mathbf{b}'_2(\mathbf{R}, \mathbf{r}_2)) \cdot \delta\left(c_2 - \frac{\mathbf{a}_2}{\|\mathbf{a}_2\|} \cdot \frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|}(\mathbf{b}'_2(\mathbf{R}, \mathbf{r}_2))\right) p_d(\mathbf{r}_2; \|\mathbf{q}_2\|) d\mathbf{r}_2$$

$$\forall \mathbf{R} \in G_{\mathbf{R}} \quad (9.36)$$

With this the Haar integral becomes

$$T = \int_{G_{\mathbf{R}}} X_1(\mathbf{R}\mathbf{q}_1) \cdot X_2(\mathbf{R}\mathbf{q}_2) d\mathbf{R} \quad (9.37)$$

Now we can use the framework introduced in section 7.7 for the fast but still fully rotation invariant approximation of the solution with a spherical harmonics expansion.

For rotations only around the z-axis the colinearity restriction is not needed and the resulting X_1 are scalar functions defined on a circle. There the fast approximation of the solution can be achieved by a Fourier series expansion (see section 7.6).

9.5 Interpretation

For a more intuitive understanding of the MiSP invariants this section provides an interpretation of the individual terms within the final formulas (9.35), (9.36) and (9.37):

$\|\nabla_{\mathbf{x}'} X\|$ is the gradient magnitude of the volumetric data set. The fact that it is computed in the transformed coordinate system influences the magnitude only marginally for the deformations that are considered in this application, and is neglected here. This allows to compute the gradient magnitude in the original coordinate system. An example for the gradient magnitude of a volumetric data set of a *Betula*^(A.4) pollen grain is shown in figure 9.10b.

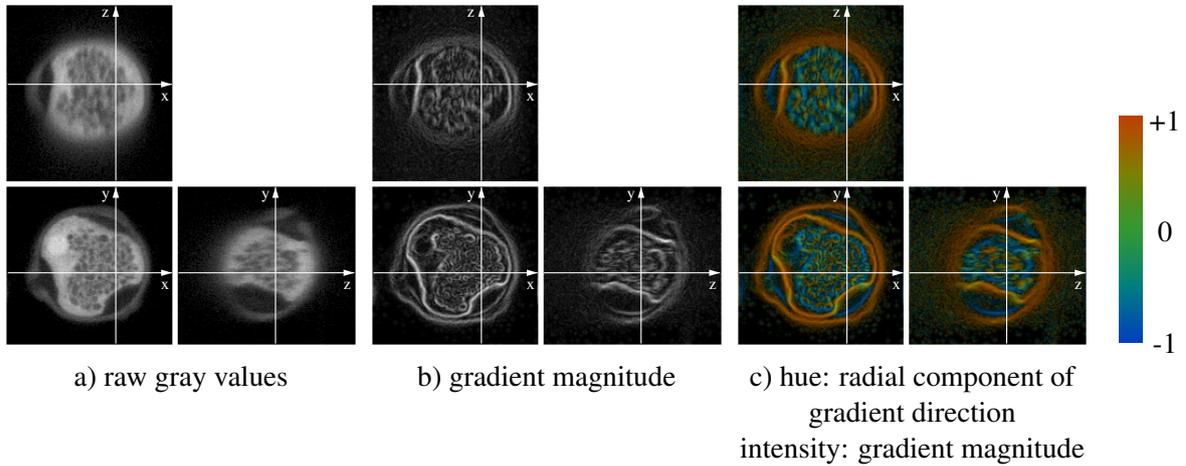


Figure 9.10: Orthogonal slices of a confocally recorded volume of a *Betula*^(A.4) pollen grain. For high robustness to nonlinear gray scale transformations, the MiSP invariants do not operate on the plain gray values, but on the gradient direction and gradient magnitude instead.

The term $\frac{\nabla_{\mathbf{x}'} X}{\|\nabla_{\mathbf{x}'} X\|}$ is the (vectorial) gradient in the transformed coordinate system normalized to unit length. Here the fact that the gradient is computed in the transformed coordinate system is important: In combination with the dot product with $\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$ it returns always the radial component of the normalized

gradient. Here again we neglect the influences of the deformation on the gradient direction such that we can display it in the original dataset. In figure 9.10c it is overlaid as hue on the gradient magnitude image. The reason for using the gradient direction and gradient magnitude instead of operating on the plain gray values is their robustness to arbitrary gray value transformations (discussed in detail in section 9.2.1).

The two resulting scalar datasets (the gradient magnitude and radial component of the gradient direction) are scanned by the transformed kernel points \mathbf{b}'_1 and \mathbf{b}'_2 during the integration over all possible rotations of the kernel function. The scanned region of each kernel point is $\mathbf{s}(\mathbf{R}\mathbf{a}_i) + \mathbf{R}\mathbf{q}_i + \mathbf{r}_i$ (see (9.32)), where $\mathbf{s}(\mathbf{R}\mathbf{a}_i)$ is a point on the segmentation surface and $\mathbf{R}\mathbf{q}_i$ a vector with a fixed length that points from that surface point to the center. For all rotations the sum of these two vectors describe a “shell” with the radial distance $\|\mathbf{q}_i\|$ to the segmentation border. The last term \mathbf{r}_i (originating from the local deformation model) together the Gaussian probability $p_d(\mathbf{r}_i; \|\mathbf{q}_i\|)$ and the integration $d\mathbf{r}_i$ can be interpreted as a convolution of this shell with a Gaussian, such that the scanned area is smoothly extended into its neighborhood. This extension is greater if the shell has a bigger distance from the outer surface. This area is depicted as white overlay in the orthogonal slices in the top row of figure 9.11b and 9.11c). This “smoothing” has a clear interpretation: By interpreting the segmentation surface as a set of fix points that are reproducibly found for each pollen grain, and by taking the local deformations into account, it is obvious, that the relative position of a certain structure to these fix points varies more if it is farther away. Correspondingly the scan area must be enlarged to increase the probability that it scans the corresponding structures in two different pollen grains.

Before this smoothing (which is a result of the integration over the local deformation model) takes effect, The δ -function in (9.35) selects only those points of the structure, whose radial component of the gradient direction matches the given c_1 . All the other points do not contribute to the integral. As a result, the dense 3D data is projected into a sparse representation in the 4D space (see figure 9.12).

This is a very important part of the MiSP invariants. This technique increases the specificity of the resulting features and at the same time increases the robustness to randomly oriented gradients originating from noise.

Without this technique, the integration over the local deformation model would result in a simple low-pass filtering, which just removes all fine structures. This would of course also result in a high similarity of structure and its deformed version, but it would not be possible to distinguish this structure from a background-only image (see figure 9.13)

In contrast to this the technique that is used by the MiSP invariants first distributes the voxels according to their gradient direction into several images, where the reproducibly oriented points of the structure will end up in the same image, while the randomly oriented points of the background will be equally distributed over all images, resulting in a background that is significantly reduced. After this operation, the integration over the local deformation model will not “smooth away” the fine structures, resulting in the desired similarity properties (see figure 9.14)

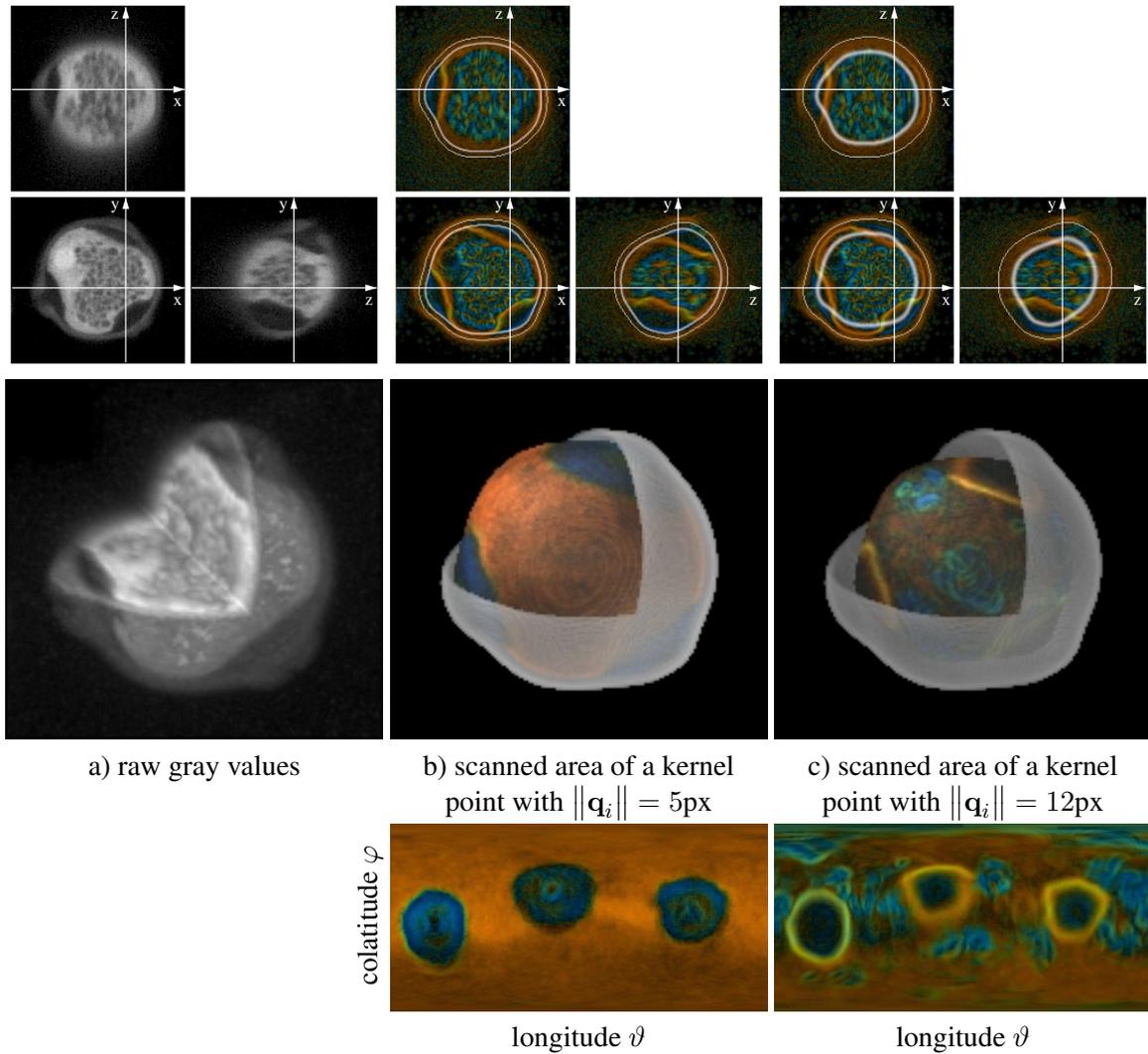


Figure 9.11: Interpretation of the MiSP invariants: Each kernel point \mathbf{b}_i scans the radial component of the gradient direction (displayed as hue) and the gradient magnitude (displayed as intensity) on a “shell” with distance $\|\mathbf{q}_i\|$ (and a “smoothing” proportional to $\|\mathbf{q}_i\|$) to the segmentation surface. The segmentation surface and the resulting scan areas for different $\|\mathbf{q}_i\|$ are displayed as white overlays in the orthogonal slices in the top row and as surface renderings in the middle row. In the bottom row the scanned areas are shown as function of the longitude and the colatitude. In this representation the three porates of the *Betula*^(A.4) pollen grain become clearly visible

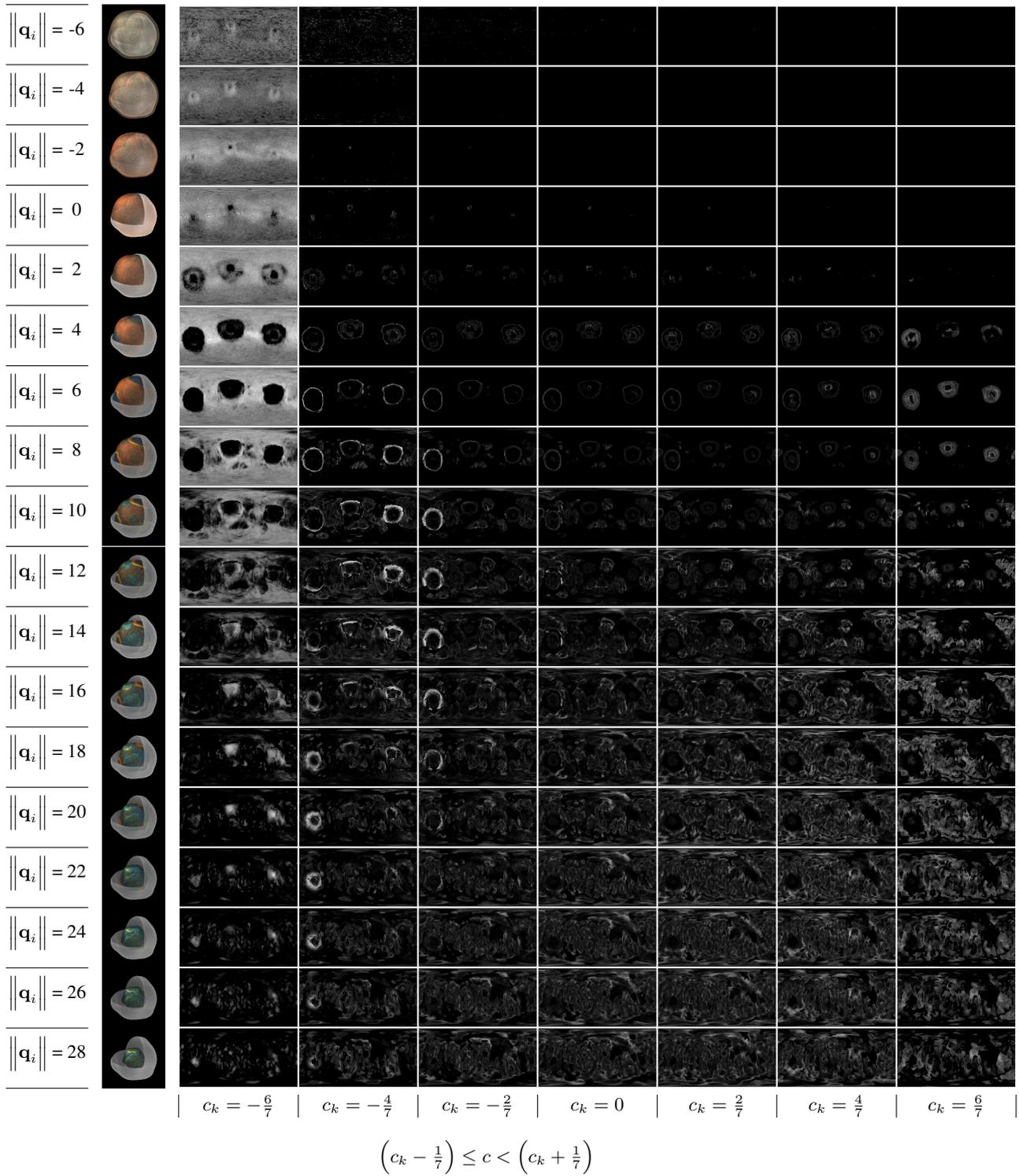


Figure 9.12: Interpretation of MiSP invariants part 2: Projection of the original dense 3D space (represented by the volume renderings) to the sparse 4D space that is spanned by the parameters. The table shows for each parameter combination of $\|\mathbf{q}_i\|$ (distance to segmentation surface) and c (radial component of the normalized gradient) the sensed signal as function of the longitude ϑ and the colatitude φ

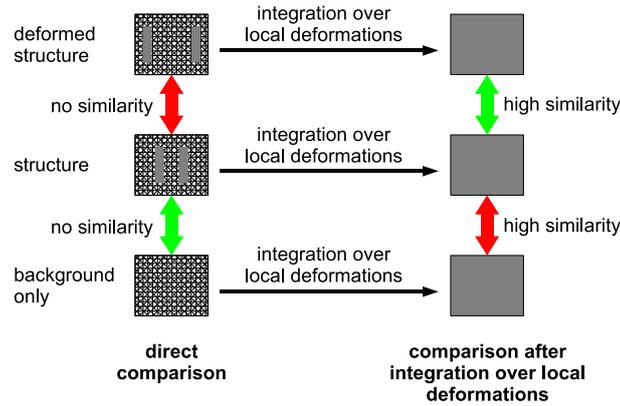


Figure 9.13: Depiction of the effects of an integration over a local deformation model without using the gradient direction. A structure on a random background and its deformed version show no similarity in a direct (voxel-by-voxel) comparison and a high similarity after the integration over a deformation model (which results here in a simple smoothing). At the same time this transformation increases the similarity to a background-only image.

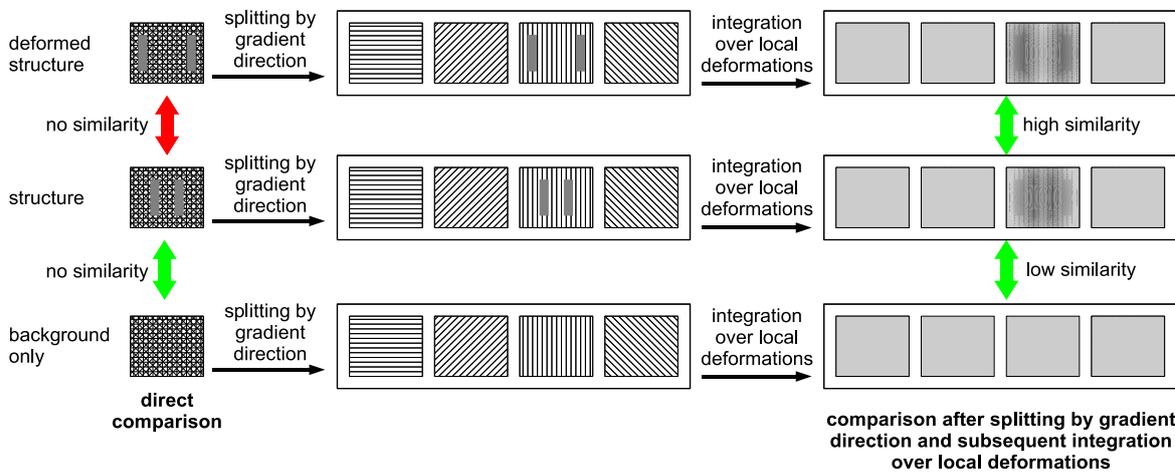


Figure 9.14: Depiction of the effects of an integration over a local deformation model that makes use of the gradient direction. Due to the projection of the original structures into a higher dimensional space, the integration over the local deformations does not smooth away the differences between the structure and background.

9.6 Pseudo Code

A pseudo code for the computation of the MiSP invariants for 3D rotations is shown in algorithm 9.1

Algorithm 9.1 MiSP invariants for 3D rotations invariants= MiSP(X , mask)

Input: volumetric data set X , volumetric mask where object voxels are set to 1

Output: invariants

```

// compute 4D profile
1: initialize 4D profile to zero    // dimensions are the distance to border ("shell"), radial compo-
    nent of gradient direction ("gradDir"),  $\varphi$  and  $\vartheta$ 
2: initialize 4D weight array with same dimensions as profile
3: compute center of mask as object center
4:  $D$  := array (3D) containing radial distance to mask border
5:  $G$  := radial component of gradient direction from  $X$ 
6:  $M$  := gradient magnitude from  $X$ 
7: for all  $(\varphi, \vartheta)$  do
8:   for all positions  $\mathbf{x}$  on a beam from object center in  $(\varphi, \vartheta)$  direction do
9:     profile( $D(\mathbf{x}), G(\mathbf{x}), \varphi, \vartheta$ ) +=  $M(\mathbf{x})$     // Array access uses tri-linear interpolation
10:     weight( $D(\mathbf{x}), G(\mathbf{x}), \varphi, \vartheta$ ) += 1
11:   end for
12: end for
13: profile := profile / weight
    // Integration over local deformations
14: for all shells in profile do
15:    $\sigma$  :=  $\sigma_d$  · shell distance to border
16:   smoothProfile(shell) =  $\sum_d \text{Gauss}_\sigma(d) \cdot \text{profile}(\text{shell} + d)$ 
17: end for
    // perform the sperical harmonics transformations
18:  $W$  := array (4D)    // dimensions are shell, gradDir, SH degree  $\ell$ , SH order  $m$ 
19: for all shells in smoothProfile do
20:   for all gradDir in smoothProfile do
21:      $W(\text{shell}, \text{gradDir}) := \text{SH-Transform}(\text{smoothProfile}(\text{shell}, \text{gradDir}))$ 
22:   end for
23: end for
    // compute the invariants
24: for all kernel parameter sets  $(q_1, q_2, c_1, c_2, \varphi_{\text{circ}})$  do
25:    $T = \sum_{l=0}^N P_0^l(\cos \varphi_{\text{circ}}) \cdot \sum_{m=0}^l \Re\left(W(q_1, c_1, \ell, m) \cdot W^*(q_2, c_2, \ell, m)\right)$ 
26:   append  $T$  to list of invariants
27: end for

```

In the first part (lines 1 – 17) the dense 3D data is mapped into the sparse 4D space (depicted in figure 9.12). The computation of the radial distance to the border in line 4 can be approximated by the fast distance transform from Felzenszwalb and Huttenlocher (2004).

In the second part (lines 18 – 23) the spherical harmonics transformation for each shell and gradient direction is done. Here we can use the fast spherical harmonics transformation from Healy et al. (2003).

In the last part (lines 24 – 27) the invariants are computed from the spherical harmonics based on equation (7.67).

10 Experimental Results

10.1 Relevant Statistical Measures

One important aspect in a reasonable evaluation and comparison of the results are the used statistical measures. For the experiments with pure pollen samples, containing approximately the same number of samples per class, this is relatively easy. The recognition rate,

$$\text{reconition rate} = \frac{\text{correct classifications}}{\text{total number of samples}} , \quad (10.1)$$

ranging from 0% to 100% is a meaningful value for the comparison of different experiments or approaches.

In applications with extremely different numbers of samples per class and of different importance of these classes, the overall recognition rate might not carry any valuable information. E. g., in the case of pollen recognition on real air samples, we have an enormous amount of dust particles. Even though the segmentation already rejects all non-spherical particles, we still have a proportion of about 9:1 for dust particles per pollen. So a classifier that just assigns every particle to the dust class already returns a remarkable recognition rate of 90%.

For pollen recognition we need measures, that are independent of the total number of particles, and that are not biased by the high recognition rates on dust particles. For that we select precision and recall with the definition that is used in information retrieval. By abbreviating number of true positives (hits) with n_{tp} , the number of false positives (false alarms) with n_{fp} , and the number of false negatives (misses) with n_{fn} , we can write

$$\text{precision} = \frac{n_{tp}}{n_{tp} + n_{fn}} \quad \text{recall} = \frac{n_{tp}}{n_{tp} + n_{fp}} . \quad (10.2)$$

Or a little more intuitive for the example of *Alnus*^(A.2) pollen this results in

$$\text{precision} = \frac{\|\{\text{correctly classified Alnus}\}\|}{\|\{\text{classified as Alnus}\}\|} \quad \text{recall} = \frac{\|\{\text{correctly classified Alnus}\}\|}{\|\{\text{true Alnus}\}\|} . \quad (10.3)$$

This is illustrated in figure 10.1.

10.1.1 Estimation of the Pollen Concentration in the Air

The most important measure that we want to compute from the resulting pollen counts is the pollen concentration in the air. So the goal is to optimize the recognition such that the computed pollen concentration has the lowest possible error.

For this we have to analyze the different error sources within the whole process:

Errors in the Sampling Process

The first and very important error that arises when measuring a particle concentration in air is the error that results from the random sampling. If we assume a spatially and temporarily constant concentration

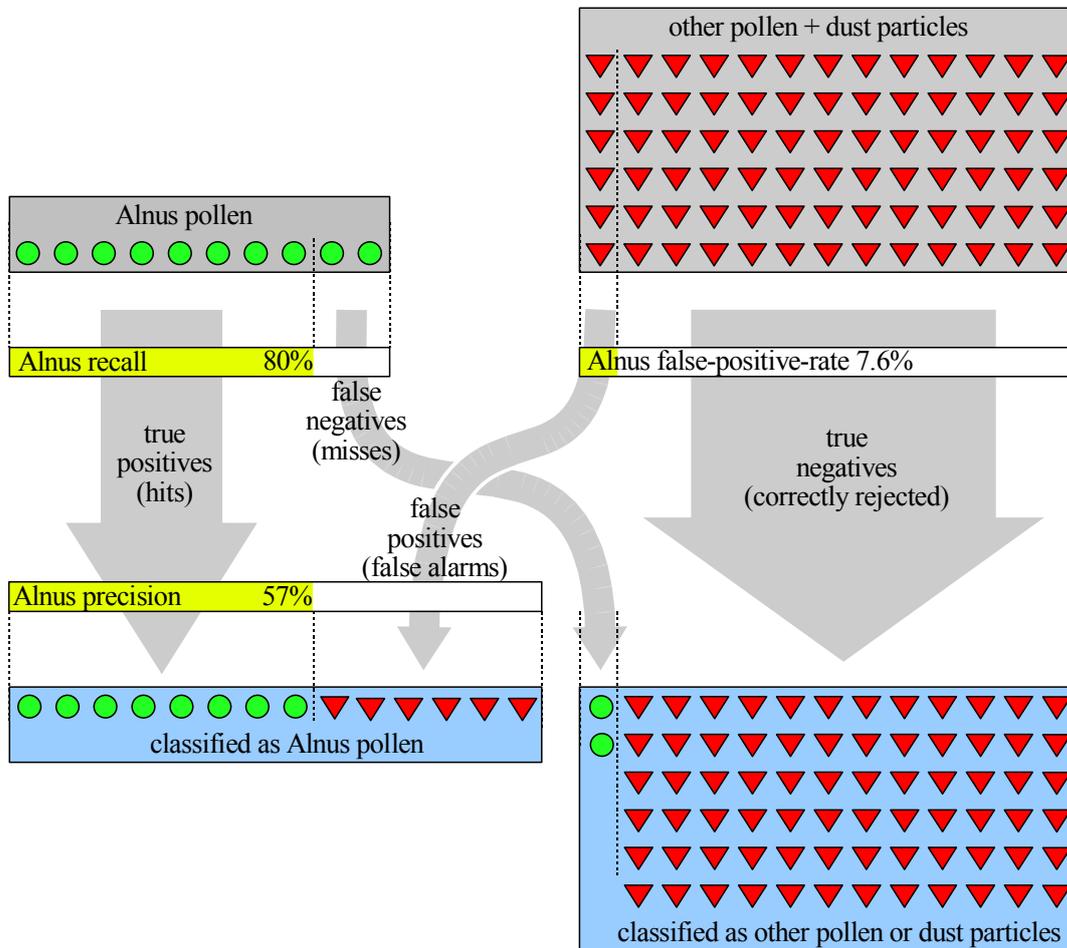


Figure 10.1: For pollen recognition in air samples the overall recognition rate is not a meaningful measure, because a correct classification of a particle from the largest class (dust particles) has no relevance for the application. Instead we compute precision and recall for each pollen taxa.

10 Experimental Results

for the considered pollen taxon of c_{taxon} (given in m^{-3}) we can model the number of pollen k , that are found in a randomly sampled volume v (given in m^3) by using the Poisson distribution. The mean number of pollen in this volume is

$$n_{\text{taxon}} = v \cdot c_{\text{taxon}} \quad (10.4)$$

By defining the corresponding random variable as N_{taxon} , and the Poisson distribution as $\text{Poi}(\lambda)$ with

$$\Pr(\text{Poi}(\lambda) = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad (10.5)$$

we can write

$$\begin{aligned} N_{\text{taxon}} &\sim \text{Poi}(n_{\text{taxon}}) \\ \mathbb{E}\{N_{\text{taxon}}\} &= n_{\text{taxon}} \\ \text{Var}\{N_{\text{taxon}}\} &= n_{\text{taxon}}. \end{aligned} \quad (10.6)$$

The same statistics apply for the other pollen and particles. Their concentration is denoted by c_{other} , and n_{other} is defined accordingly,

$$\begin{aligned} N_{\text{other}} &\sim \text{Poi}(n_{\text{other}}) \\ \mathbb{E}\{N_{\text{other}}\} &= n_{\text{other}} \\ \text{Var}\{N_{\text{other}}\} &= n_{\text{other}}. \end{aligned} \quad (10.7)$$

For an easier development of the statistics it is sufficient (without changing the result) to take only other spherical particles into account, and ignore the remaining.

Errors in the Detection.

The first step of the segmentation is the detection of spherical particles in the volumetric data set. The threshold for the sphere detector was set so low, that no pollen was lost in a certain amount of typical air samples. Therefore we neglect errors in this step and assume, that all pollen grains on the sample are cropped and forwarded to the recognition. As we limited the considered ‘‘other particles’’ also to spherical particles, the same applies here as above, and we can neglect the errors.

Errors in the Recognition.

So, the finally occurring errors are the errors in the recognition. If we assume a constant quality of the pollen, the probability to correctly recognize a certain pollen grain is the recall p_r for this pollen taxon, that was computed from the test data set.

In this case the number of true positives n_{tp} can be modeled by the binomial distribution, from the true number of pollen on the sample k as

$$p(n_{\text{tp}}; k; p_r) = \binom{k}{n_{\text{tp}}} p_r^{n_{\text{tp}}} (1 - p_r)^{k - n_{\text{tp}}}, \quad (10.8)$$

where the mean of this distribution is $k p_r$ and the variance $k p_r (1 - p_r)$. For the combined probabilities including sampling and recognition, we do not need to compute the successive application of these two distributions, but can (for the identical result) combine the recognition rate directly with the concentration, such that the result is again a Poisson distribution around $p_r n_{\text{taxon}}$

$$N_{\text{tp}} \sim \text{Poi}(p_r n_{\text{taxon}}) \quad (10.9)$$

with the respective mean and variance

$$\begin{aligned} E\{N_{\text{tp}}\} &= p_r n_{\text{taxon}} \\ \text{Var}\{N_{\text{tp}}\} &= p_r n_{\text{taxon}} \end{aligned} \quad (10.10)$$

The second source of errors in the recognition is the occurrence of false positives. Due to the different sources of the particles which are falsely classified to the considered pollen taxa, the models might be very different. For a first order approximation we neglect these different sources for false positives and assume, that the probabilities p_{fp} to become a false positive are equal for all the other spherical particles. This results in a kind of “background noise” that is present on every sample and depends on the total number of other (spherical) particles n_{other} in the considered volume of air. In analogy to the derived formulae for the pollen the number of false positives n_{fp} is Poisson distributed around $p_{\text{fp}} n_{\text{other}}$,

$$N_{\text{fp}} \sim \text{Poi}\left(p_{\text{fp}} n_{\text{other}}\right) \quad (10.11)$$

Total Error

Due to the fact that the number of particles classified as pollen n_{est} is the sum of the true positive and the false positive pollen

$$n_{\text{est}} = n_{\text{tp}} + n_{\text{fp}} \quad (10.12)$$

As N_{tp} and N_{fp} are statistically independent, we can use the summation property of the Poisson distribution: If $X_1 \sim \text{Poi}(\lambda_1)$ and $X_2 \sim \text{Poi}(\lambda_2)$ are independent, then $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$. Therefore the random variable N_{est} is

$$N_{\text{est}} \sim \text{Poi}\left(p_r n_{\text{taxon}} + p_{\text{fp}} n_{\text{other}}\right) \quad (10.13)$$

and by using the air volume v and the concentrations,

$$N_{\text{est}} \sim \text{Poi}\left(p_r v c_{\text{taxon}} + p_{\text{fp}} v c_{\text{other}}\right) \quad (10.14)$$

accordingly the mean and the variance are

$$\begin{aligned} E\{N_{\text{est}}\} &= v \cdot (p_r c_{\text{taxon}} + p_{\text{fp}} c_{\text{other}}) \\ \text{Var}\{N_{\text{est}}\} &= v \cdot (p_r c_{\text{taxon}} + p_{\text{fp}} c_{\text{other}}) \end{aligned} \quad (10.15)$$

Finally, to estimate the concentration c_{est} , we have to correct n_{est} for the systematic errors with the known recall and false-positive-rate as

$$c_{\text{est}} = \frac{1}{v p_r} n_{\text{est}} - p_{\text{fp}} c_{\text{other}} \quad (10.16)$$

Accordingly the variance of the estimated concentration will be

$$\begin{aligned} \text{Var}\{C_{\text{est}}\} &= \frac{1}{v^2 p_r^2} v \cdot (p_r c_{\text{taxon}} + p_{\text{fp}} c_{\text{other}}) \\ &= \frac{1}{v} \cdot \left(\frac{1}{p_r} c_{\text{taxon}} + \frac{p_{\text{fp}}}{p_r^2} c_{\text{other}} \right) \end{aligned} \quad (10.17)$$

and the standard deviation

$$\sigma_{C_{\text{est}}} = \frac{1}{\sqrt{v}} \cdot \sqrt{\frac{1}{p_r} c_{\text{taxon}} + \frac{p_{\text{fp}}}{p_r^2} c_{\text{other}}} \quad (10.18)$$

From this we can compute the relative error as

$$\begin{aligned}\sigma_{\text{rel}} &= \frac{\sigma_{C_{\text{est}}}}{C_{\text{taxon}}} \\ &= \frac{1}{\sqrt{v} \cdot C_{\text{taxon}}} \cdot \sqrt{\frac{1}{p_r} C_{\text{taxon}} + \frac{p_{\text{fp}}}{p_r^2} C_{\text{other}}}\end{aligned}\quad (10.19)$$

and by some rearranging we can find a form that separates this formula into the sampling error and a factor depending on the recognition system,

$$\sigma_{\text{rel}} = \underbrace{\frac{1}{\sqrt{v C_{\text{taxon}}}}}_{\text{sampling error}} \cdot \underbrace{\sqrt{\frac{1}{p_r} + \frac{p_{\text{fp}}}{p_r^2} \cdot \frac{C_{\text{other}}}{C_{\text{taxon}}}}}_{\text{factor due to recognition errors}}. \quad (10.20)$$

10.1.2 Discussion

By taking a closer look at the resulting error we can see the following:

- The sampling volume v appears only in the sampling error, so independent of all other errors, we can always become better by increasing this sampling volume or correspondingly the evaluation area on the sample. A doubling of the sampling volume will decrease the sampling error by a factor of $\sqrt{2}$.
- The “factor due to recognition errors” is 1 for a perfect recognition system with recall $p_r = 1$ and false-positive-rate $p_{\text{fp}} = 0$, and increases with decreasing recall or increasing fp-rate.
- The second term in the “factor due to recognition errors” is always present – a constant background noise – and can be reduced, if we manage to decrease the false-positive rate p_{fp} or by increasing the recall p_r , e. g. in a setup with 90% recall and 1% false-positive-rate, an increase of the recall from 90% to 100% allows an increase of the false-positive-rate from 1% to 1.23% for the identical contribution of this second component
- By computing the ratio of concentrations of the other particles to the considered pollen taxon $\frac{C_{\text{other}}}{C_{\text{taxon}}}$ from the pollenmonitor data set, we find a mean factor of about 66 for *Urtica*^(D.15) (the most frequent pollen taxa in our dataset), a factor of 120 for *Poaceae*^(A.6) (the most important pollen taxa), up to to 3000 for *Artemisia*^(A.1) (the pollen taxa with the highest allergenic impact at low concentrations). From this we can see, that a slight increase of the false-positive-rate will have a dramatic impact on the error.

10.1.3 A Better Measure for the False-Positives

As mentioned above, the false-positive-rate p_{fp} is no meaningful measure to compare different pollen recognition systems, because it can only be given relative to the number of segmented particles, and this number heavily changes depending on the selected algorithms and thresholds.

By a closer look to the dependencies of the final error on this false-positive-rate, we find, that it always appears in a product with the concentration of other particles C_{other} . By defining a “false-positive-concentration”

$$C_{\text{fp}} := p_{\text{fp}} \cdot C_{\text{other}}, \quad (10.21)$$

we can introduce a measure, which does not depend on the segmentation parameters anymore, and which has a clear physical meaning. It is the mean number of false positives, that we will find in one m^3 of air. We can compare this measure (as a kind of background-noise) directly to the pollen-concentrations that we want to measure to correctly predict the allergenic impact of a certain situation.

10.2 Results on the Confocal Data Set

10.2.1 Euclidean Invariants on the Confocal Data Set

Our first experiments on automated pollen recognition were done on the “confocal data set” which was briefly introduced in section 3.1. As mentioned there, the recording quality of the *Larix*^(D.8) and *Pinus*^(D.10) pollen was too low, such that they were not included in the experiments. Additionally three raw-data files were corrupted (containing 3 *Alnus viridis*^(A.3), 1 *Carpinus*^(B.1) and 1 *Corylus*^(A.5))¹ and therefore have been left out from the further processing. The remaining raw data volumes were cropped into single volumes using a cubic region around each grain that was found with a given threshold. The resulting 385 volumetric data sets of pollen grains from 26 taxa were smoothly scaled down by a factor of 2 to reduce the computational costs and were padded with zeros to a size of $128 \times 128 \times 128$ voxels to ensure that the features are not influenced by the size of the dataset.

To become invariant to linear gray value transformations, the individual data sets were normalized to unit variance. After that 14 invariants were computed for each data set by evaluating the two-point-kernel functions,

$$\begin{aligned} f_1(X) &= X(0, 0, 0) \cdot X(0, 0, 2) \\ f_2(X) &= \sqrt{X(0, 0, 0)} \cdot \sqrt{X(0, 0, 2)} \end{aligned} \quad (10.22)$$

at 7 different scalings of the object (1:1, 1:2, 1:4, 1:8, 1:16, 1:32 and 1:64). The computations were done by the fast algorithms introduced in section 7.4, using the freely available FFT implementation FFTW (Frigo and Johnson, 2005).

For the classification a set of 26 two-class Support Vector Machines (“SVM”) were used, each of them trained to separate one class from the rest. The kernel function for the SVM (not to be confused with the kernel functions in the Haar integral) was chosen to be the radial basis function (“RBF”),

$$k(\mathbf{a}, \mathbf{b}) = e^{-\gamma \|\mathbf{a}-\mathbf{b}\|^2} \quad (10.23)$$

There are two free parameters in the SVM training, the regularization parameter c , which defines the “cost” for training samples on the wrong side of the hyper-plane, and the kernel-parameter γ , which describes the “distance measure” in the feature space. A big γ results in a “fine-structured” border between the classes, while a smaller γ results in a “coarse”, more generalized border.

Due to the limited size of the data set, a leave-one-out validation was applied to measure the recognition rate. To reduce the risk of over-fitting only a very low number of different SVM-parameter sets were tested. The best found results are reported in table 10.1. The overall recognition rate is 91.9%, which was at the time when we published these results (Ronneberger et al., 2002b,a) better than the published state of the art, so far.

With the development of our highly optimized SVM implementation libsvm1 that allows to run the complete leave-one-out validation on this data set within 1 second (on an Opteron 885 at 2.6GHz) we could significantly improve these results: We applied a prior scaling of each feature to unit standard deviation within the whole dataset

$$T'_i(X_j) = \frac{T_i(X_j)}{\sqrt{\frac{1}{N} \sum_{k=1}^N (T_i(X_k))^2}} \quad \forall i \in \{1, \dots, M\} \text{ and } j \in \{1, \dots, N\} \quad (10.24)$$

where $M = 14$ is the dimension of the feature vector and $N = 385$ is the number of samples in the full dataset. For these scaled features one can easily find SVM-training-parameter combinations for cost c and kernel-parameter γ which return a leave-one-out recognition rate of 95.6% (see table 10.2). Please note that this kind of experiments always have a certain risk of over-fitting.

¹With a later written converter it was able to recover the raw data of the first two corrupted files, such that in the newer experiments 389 pollen grains are used.

Table 10.1: Confusion table for the “confocal data set” with general gray-scale invariants and non-optimized SVM

		Classified As...																														
		number of training samples	number of test samples	Artemisia	Alnus	Alnus viridis	Betula	Corylus	Poaceae	Secale	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Salix	Acer	Chenopodium	Compositae	Cruciferae	Aesculus	Juglans	Platanus	Taxus	Tilia	Ulmus	Urtica	recall (%)	rec. per group (%)	
True Class	Artemisia	15	15	13	1	1	86.7	
	Alnus	15	15	.	15	100	
	Alnus viridis	12	12	.	.	12	100	
	Betula	15	15	.	.	.	14	1	93.3	92.1
	Corylus	14	14	.	1	.	.	13	92.9	
	Poaceae	15	15	15	100	
	Secale	15	15	11	3	1	.	.	73.3	
	Carpinus	14	14	14	100	
	Quercus	15	15	11	.	1	1	2	73.3	
	Fraxinus	15	15	12	1	2	80	87.8
	Plantago	15	15	2	13	86.7	
	Rumex	15	15	15	100	
	Fagus	15	15	15	100	
	Populus	15	15	14	.	.	.	1	93.3	97.8
	Salix	15	15	15	100	
	Acer	15	15	14	1	.	.	93.3	
	Chenopodium	15	15	1	1	.	.	.	1	12	80	
	Compositae	15	15	15	100	
	Cruciferae	15	15	1	1	.	13	86.7	
	Aesculus	15	15	15	100	
	Juglans	15	15	1	.	1	13	86.7	92.1
	Platanus	15	15	15	100	
	Taxus	15	15	15	.	.	.	100	
	Tilia	15	15	1	14	.	.	93.3	
	Ulmus	15	15	1	2	.	12	.	.	80	
	Urtica	15	15	1	.	.	14	.	93.3	
	precision (%)				100	93.8	100	100	100	88.2	100	93.3	91.7	85.7	76.5	100	83.3	82.4	100	87.5	80	83.3	100	100	100	100	78.9	100	87.5	100	100	
	prec. per group (%)				96.9			89			88			91.6																		

Data type: 3D confocal recordings from manually prepared samples
Training + Test data set: full data set (leave-one-out validation)
Features: 14 gray scale invariants (fully Euclidean motion)
Classifier: SVM (RBF kernel, One-vs-Rest)
Overall Recognition Rate: 91.9% (354/385)

Table 10.2: Confusion table for the “confocal data set” with general gray-scale invariants with feature scaling and extensive grid search for the SVM training parameters

		Classified As...																				recall (%)	rec. per group (%)											
		number of training samples	number of test samples	Artemisia	Alnus	Alnus viridis	Betula	Corylus	Poaceae	Secale	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Salix	Acer	Chenopodium	Compositae	Cruciferae	Aesculus	Juglans	Platanus	Taxus	Tilia	Ulmus	Urtica	recall (%)	rec. per group (%)			
True Class	Artemisia	15	15	14	93.3			
	Alnus	15	15	.	15	100			
	Alnus viridis	12	12	.	.	11	91.7			
	Betula	15	15	.	.	.	14	.	.	.	1	93.3	95		
	Corylus	14	14	14	100			
	Poaceae	15	15	15	100			
	Secale	15	15	13	1	1	.	.	86.7			
	Carpinus	14	14	14	100		
	Quercus	15	15	12	.	1	1	1	80		
	Fraxinus	15	15	14	1	93.3	91.9	
	Plantago	15	15	.	.	.	1	1	13	86.7		
	Rumex	15	15	15	100		
	Fagus	15	15	15	100		
	Populus	15	15	1	13	.	.	.	1	86.7	95.6	
	Salix	15	15	15	100		
	Acer	15	15	15	100		
	Chenopodium	15	15	1	14	93.3		
	Compositae	15	15	15	100		
	Cruciferae	15	15	15	100		
	Aesculus	15	15	15	100		
	Juglans	15	15	15	100	97.6	
	Platanus	15	15	.	.	1	14	93.3			
	Taxus	15	15	15	.	.	.	100			
	Tilia	15	15	15	.	.	100			
	Ulmus	15	15	1	1	.	.	13	.	86.7			
	Urtica	15	15	15	100			
	precision (%)				100	100	91.7	93.3	100	100	100	87.5	92.3	87.5	92.9	100	93.8	100	100	93.8	87.5	93.8	100	100	100	100	93.3	100	93.8	86.7	100			
	prec. per group (%)				98			91.9			97.7			95.3																				

Data type: 3D confocal recordings from manually prepared samples
Training + Test data set: full data set (leave-one-out validation)
Features: 14 gray scale invariants (fully Euclidean motion) (scaled to unit standard deviation)
Classifier: SVM (RBF kernel, One-vs-Rest) (extensive grid search for γ and c)
Overall Recognition Rate: 95.6% (368/385)

10.2.2 Graph-Cut Segmentation on the Confocal Data Set

By taking a first look at the confocal pollen data, it seemed to be easy to segment the bright pollen from the dark background just by the application of an appropriate threshold and some additional morphological steps. These techniques gave satisfactory results for the boundaries in x-y-direction, but failed at the borders in z-direction (especially the lower border), which are significantly blurred by the recording process.

In order to extract a robust surface, the graph-cut based algorithm described in section 8.4 was used.

The probabilities for each gray value v to belong to background $p_{bg}(v)$ or to foreground $p_{fg}(v)$ were modelled with Gaussian distributions. The parameters were computed relative to the mean gray value \bar{v} as

$$\mu_{bg} = 1 \qquad \mu_{fg} = \bar{v} \qquad (10.25)$$

$$\sigma_{bg} = \frac{1}{\sqrt{-2 \ln \frac{1}{2}}} \cdot 0.5 \cdot \bar{v} \qquad \sigma_{fg} = \frac{1}{\sqrt{-2 \ln \frac{1}{2}}} \cdot \bar{v} \ . \qquad (10.26)$$

See figure 10.2 for an example.

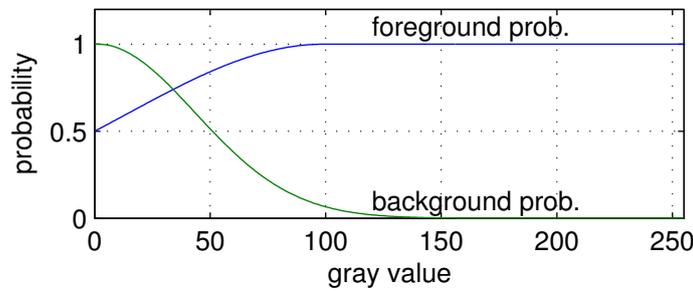


Figure 10.2: Example for the modelled foreground- and background probabilities for a mean gray value of 100

The pre-factor $\frac{1}{\sqrt{-2 \ln \frac{1}{2}}}$ was selected such that the parameter could be provided in a more intuitively scale relative to the gauss width at half maximum.

The probabilities that two neighboring voxels belong to the same class are also modeled with a Gaussian, where σ_v was heuristically determined to be

$$\sigma_v = \frac{4}{\sqrt{-2 \ln \frac{1}{2}}} \ . \qquad (10.27)$$

The relative weight between the voxel-to-voxel edges and the voxel-to-terminal edges was set to 4:1.

With this edge weights, the graph was built using the 125-neighborhood for each voxel. To obtain a single smooth mask for all pollen taxa and to reduce the typical graph-cut artifacts (borders in the direction of the voxel-to-voxel edges are preferred), some pre- and post-processing steps have been applied. The pre-processing steps are

1. smoothing and gamma correction (with $\gamma = 2$) of the raw data to reduce noise and emphasize the important, but dark structures at the border
2. down-sampling of the array by a factor of 4 in each dimension for faster processing and a larger neighborhood for the graph-cut.

3. plane-wise gray-fill-holes algorithm to fill inner holes in the pollen grains (especially needed for *Taxus*^(D.12) pollen)

The post-processing steps are:

1. up-scaling of the generated mask by a factor of 4
2. repeated smoothing and binarization of the resulting mask, which removes small bumps while keeping the coarse shape untouched.
3. connected component labeling to crop the data sets with multiple pollen into individual data sets.

With the selected parameters and the pre- and post-processing, all 389 pollen grains were segmented satisfactorily. Two typical results are shown in figure 10.3 and figure 10.4. Only for some pollen grains with an extremely blurred lower border, the results were suboptimal (see figure 10.5), but in these cases it is even hard for a human to specify the correct position of the border.

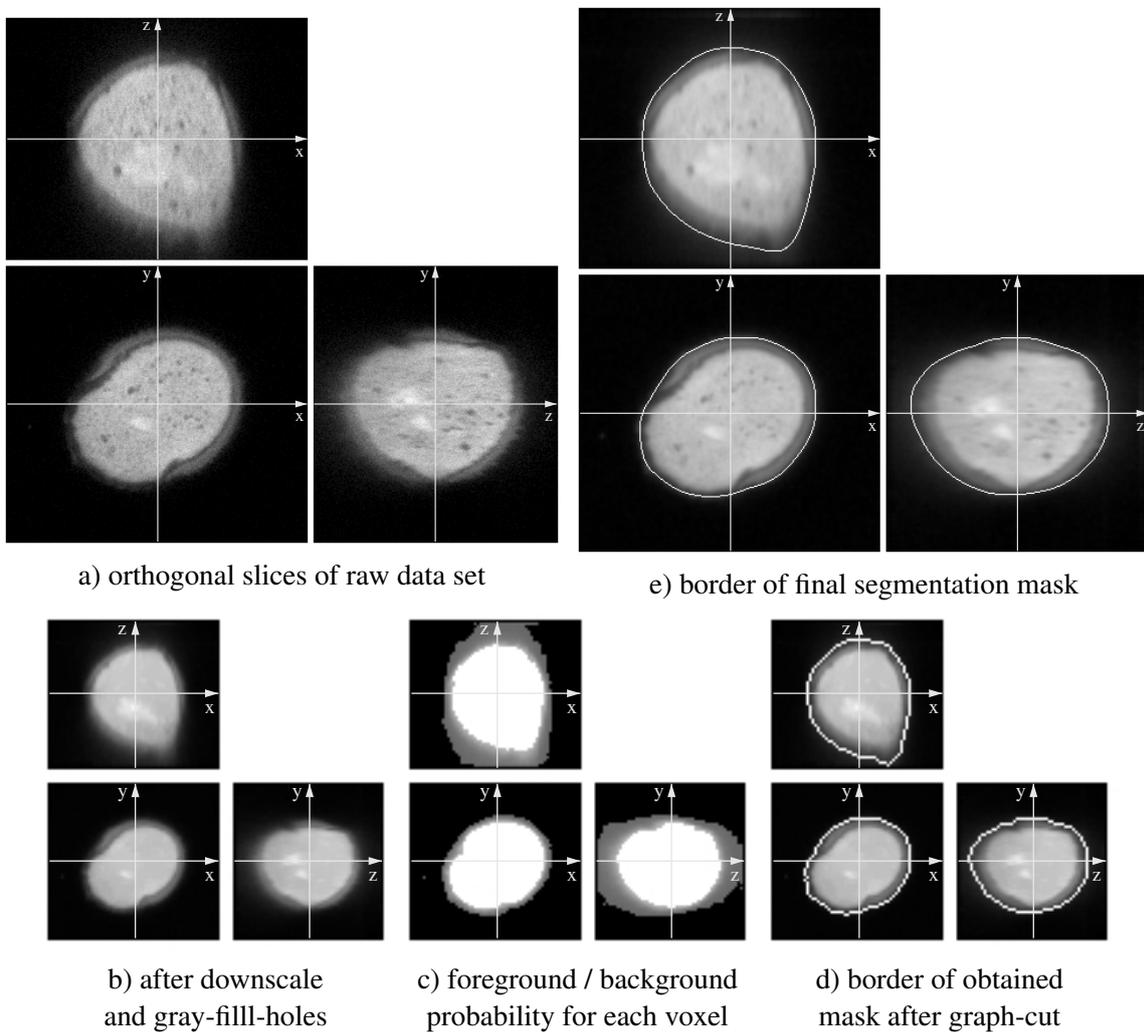


Figure 10.3: Graph-cut based segmentation of an *Acer*^(D.1) pollen

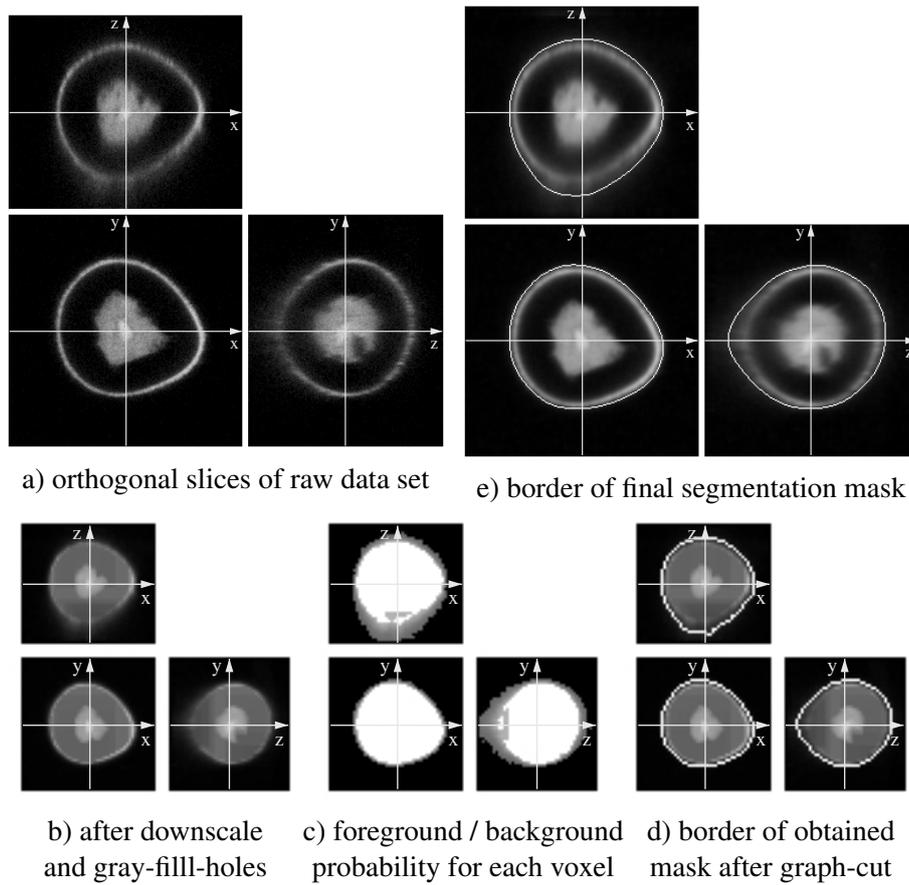


Figure 10.4: Graph-cut based segmentation of a *Taxus*^(D.12) pollen. For this pollen taxa the pre-processing with the gray-fill-holes algorithm was essential to avoid two distinct regions.

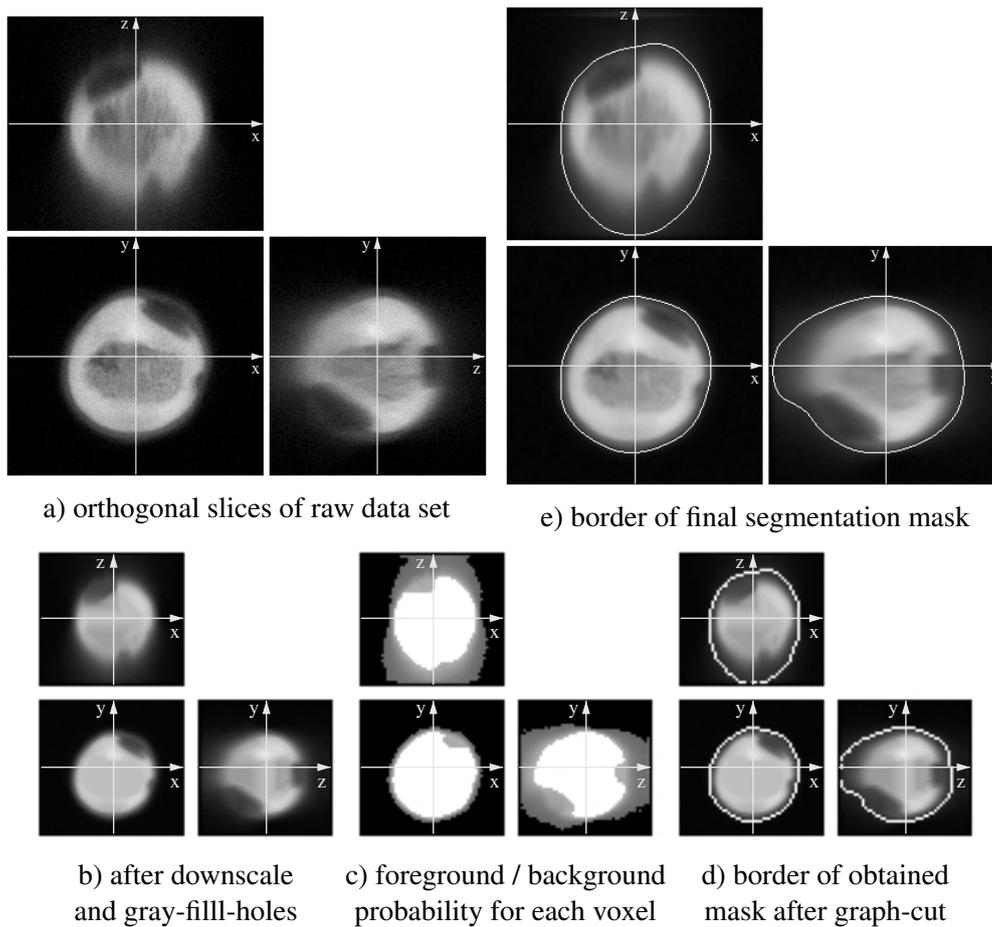


Figure 10.5: In rare cases the results of the graph-cut segmentation were suboptimal at the lower border of the object (here a *Corylus*^(A.5) pollen). But even for a human it is hard to specify the correct position of the border in this data set.

10.2.3 MiSP Invariants on the Confocal Data Set

After the extraction of the outer border with the graph-cut based segmentation, the MiSP invariants (introduced in chapter 9) can be applied to this data set.

For the experiments described here we only used kernels with $\|\mathbf{q}_1\| = \|\mathbf{q}_2\|$ and $c_1 = c_2$. For the application on the confocal data set (allowing full 3D rotations) this results in 3 parameters for the kernel: The distance q to the segmentation surface, the relative direction of the gradient c and the desired angular resolution n (see figure 10.6)

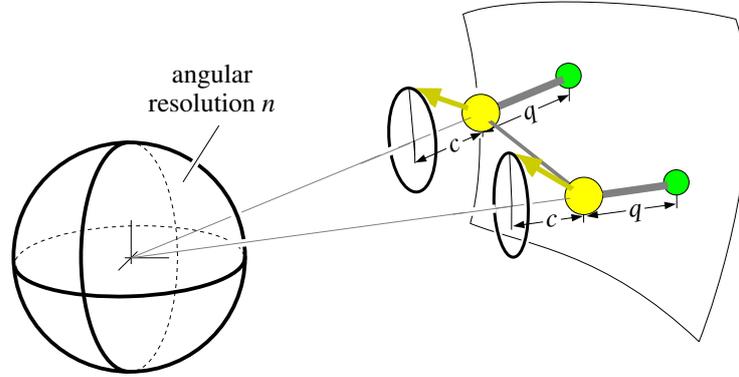


Figure 10.6: Parameterization of the kernel function with 3 parameters q, c, n for invariance to full 3D rotations. The thick yellow arrows denote the local gradient direction in the data set. Only if the radial component of the gradient direction matches c (which is given for all points on the small circle) then the corresponding gradient magnitude contributes to the Haar integral. The illustrated surface part symbolizes the outer surface of the segmentation mask.

For the computation, each voxel of the dense 3D data is first projected into the sparse representation in the 4D kernel parameter space, defined by each “arm” of the kernel function $(q, c, \varphi, \vartheta)$. The advantage of this sparseness is that fine detail information from the original images “survive” the smoothing effects of the partial Haar-integration over the local deformation model and the extraction of rotation invariant features (see section 9.5).

For the selected kernels with $\|\mathbf{q}_1\| = \|\mathbf{q}_2\|$ and $c_1 = c_2$ a further reduction of the complexity can be achieved. In this case the spherical harmonics coefficients $W_{2,\ell,m}$ and $W'_{3,\ell,m}$ in equation (7.67), are identical. The final features are only a nonlinear combination of the magnitudes of the computed SH-coefficients. As this final recombination does not introduce additional information we can omit it, and instead can use the magnitudes of the SH-coefficients directly.

The best quantizations for the parameters were found by visual inspection of intermediate results and some leave-one-out experiments. Due to the very high recognition rate (only a few pollen are misclassified) it did not make sense to base a parameter search on the final recognition rate. The statistical relevance of one more or one less correctly classified pollen is not very high. The best found parameters are listed in table 10.3.

By the use of these quantizations, we obtain $N_q \times N_c \times n = 64 \times 7 \times 2 = 896$ “structural” MiSP invariants. For the same reason as above, we did not compute the “shape” features. In a leave-one-out validation with these features using a 1-nearest-neighbor classifier on the full data set we obtained a recognition rate of 99.2% (386 of 389 pollen). The details and volume renderings of the 3 incorrectly classified pollen are shown in table 10.4.

Due to the extraction of the structural features relative to the segmentation border a certain scale invariance is obtained. This may be an explanation for the first two incorrectly classified pollen there. We suppose that the addition of the “shape” features might have solved this problem, or at least would

Table 10.3: Selected parameters for the computation of the MiSP invariant on the confocal data set

parameter	value(s)	description
q	64 steps from $2\mu\text{m}$ outside to $10\mu\text{m}$ inside	radial localization of the “shells” relative to the segmentation border
c	7 steps from -1 to +1	radial component of the normalized gradient
n	2	spherical harmonic bands for the approximation of the kernel function
σ_d	0.1	relative deformations for the local deformation model

lead to different (equally sized) wrong nearest neighbors.

In a leave-one-out validation with a support vector machine, using the histogram-intersection kernel (which is equivalent to the L1 norm) we obtain a recognition rate of 99.5% (387/389): Only 2 pollen are incorrectly classified: One *Alnus*^(A.2) pollen was classified as *Corylus*^(A.5) and one *Betula*^(A.4) pollen was classified as *Alnus viridis*^(A.3) (see table 10.5).

The *Alnus*^(A.2) pollen is the same as in the 1NN experiment. The *Betula*^(A.4) pollen that was classified to *Alnus viridis*^(A.3) is a slight improvement, because *Betula*^(A.4) and *Alnus viridis*^(A.3) belong to the same family. But nevertheless from the reduction from 3 to 2 incorrectly classified pollen, we can not draw any reasonable conclusions for the expected performance on a separate test data set.

Table 10.4: Confusion table for the “confocal data set” with MiSP invariants

		Classified As...																																		
		number of training samples	number of test samples	Artemisia	Alnus	Alnus viridis	Betula	Corylus	Poaceae	Secale	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Salix	Acer	Chenopodium	Compositae	Cruciferae	Aesculus	Juglans	Platanus	Taxus	Tilia	Ulmus	Urtica	recall (%)	rec. per group (%)					
True Class	Artemisia	15	15	15	100					
	Alnus	15	15	.	14	.	1	93.3					
	Alnus viridis	15	15	.	.	15	100					
	Betula	15	15	.	.	.	14	1	93.3	98.1				
	Corylus	14	14	14	100					
	Poaceae	15	15	15	100					
	Secale	15	15	15	100					
	Carpinus	15	15	15	100				
	Quercus	15	15	15	100				
	Fraxinus	15	15	14	1	93.3	98.7				
	Plantago	15	15	15	100					
	Rumex	15	15	15	100				
	Fagus	15	15	15	100				
	Populus	15	15	15	100	100			
	Salix	15	15	15	100				
	Acer	15	15	15	100				
	Chenopodium	15	15	15	100				
	Compositae	15	15	15	100				
	Cruciferae	15	15	15	100				
	Aesculus	15	15	15	100				
	Juglans	15	15	15	100	100			
	Platanus	15	15	15	100				
	Taxus	15	15	15	100				
	Tilia	15	15	15	100				
	Ulmus	15	15	15	.	.	.	100				
	Urtica	15	15	15	.	100				
	precision (%)				100	100	100	100	93.3	100	100	100	100	88.2	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
	prec. per group (%)				99			97.4			100			100			100																			

Data type: 3D confocal recordings from manually prepared samples
Training + Test data set: full data set (leave-one-out validation)
Features: 896 MiSP invariants (only “structure”, no “shape”; full 3D-rotation)
Classifier: 1 Nearest Neighbor (L1-norm)
Overall Recognition Rate: 99.2% (386/389)

The three incorrectly classified pollen grains with their nearest neighbors

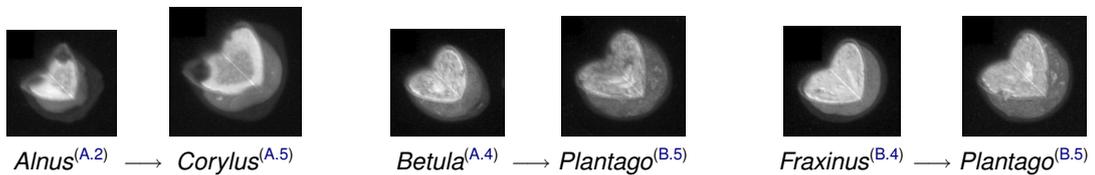
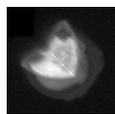


Table 10.5: Confusion table for the “confocal data set” with MiSP invariants and SVM classification

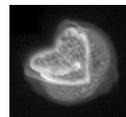
		Classified As...																													
True Class	number of training samples	number of test samples	Artemisia	Alnus	Alnus viridis	Betula	Corylus	Poaceae	Secale	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Salix	Acer	Chenopodium	Compositae	Cruciferae	Aesculus	Juglans	Platanus	Taxus	Tilia	Ulmus	Urtica	recall (%)	rec. per group (%)	
	Artemisia	15	15	15	100
Alnus	15	15	.	14	.	.	1	93.3	93.3
Alnus viridis	15	15	.	.	15	100	
Betula	15	15	.	.	1	14	93.3	98.1
Corylus	14	14	14	100	
Poaceae	15	15	15	100	
Secale	15	15	15	100	
Carpinus	15	15	15	100	
Quercus	15	15	15	100	
Fraxinus	15	15	15	100	100
Plantago	15	15	15	100	
Rumex	15	15	15	100	
Fagus	15	15	15	100	
Populus	15	15	15	100	100
Salix	15	15	15	100	
Acer	15	15	15	100	
Chenopodium	15	15	15	100	
Compositae	15	15	15	100	
Cruciferae	15	15	15	100	
Aesculus	15	15	15	100	
Juglans	15	15	15	100	100
Platanus	15	15	15	100	
Taxus	15	15	15	100	
Tilia	15	15	15	.	.	.	100	
Ulmus	15	15	15	.	.	100	
Urtica	15	15	15	.	100	
precision (%)			100	100	93.8	100	93.3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
prec. per group (%)					98.1						100				100																

Data type: 3D confocal recordings from manually prepared samples
Training + Test data set: full data set (leave-one-out validation)
Features: 896 MiSP invariants (only “structure”, no “shape”; full 3D-rotation)
Classifier: SVM, histogram intersection kernel
Overall Recognition Rate: 99.5% (387/389)

The two incorrectly classified pollen grains



Alnus^(A.2)
 → Corylus^(A.5)



Betula^(A.4)
 → Alnus viridis^(A.3)

10.2.4 Comparison to the State of the Art

In this section we want to compare our results to the state of the art of pollen recognition from pure pollen samples (see section 1.2.1). The extended table is shown in table 10.6.

Our first results using general gray-scale invariants (Ronneberger et al., 2002a) are also presented in table 10.1 were already much better (higher recognition rates on more different classes) than the state of the art.

With the identical 14 features we could increase the recognition rate to 95.6% just by scaling the features to unit variance and a grid search for the best SVM parameters.

In 2004 Qing Wang further improved the results to 97.9% using these 14 features and a modified gray level normalization (unpublished internal report, (Wang, 2004)). By increasing the number of features to 38 (by variation of the kernel size and the smoothing) she could increase the recognition further to 98.7%.

In 2006 Marco Reisert got significantly lower recognition rates on the same data set with a much larger number of 7680 features, see (Reisert and Burkhardt, 2006). The recognition rates were 96.9% using an SVM and 94.5% using a INN classifier.

The recognition rates obtained by Qing Wang could only be outperformed by the utilization of prior information from a segmentation and the integration of a deformation model. The corresponding results are presented in this thesis (tables 10.4 and 10.5) and in (Ronneberger et al., 2007).

A slightly different approach was studied together with Rafael Baumgartner on a “honey pollen data set” (see section 10.5.2 and Baumgartner (2005)). The use of multi-channel microscopical data can improve the recognition rates significantly.

Table 10.6: Comparison of our results to the results of other groups (**Pure Pollen Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate
Massey University New Zealand								
(Hodgson et al., 2005)	pure	2D	morph.	texture, shape	neural net.	184	4	96%
(Allen et al., 2006)	"	"	"	"	"	1,400	7	98%
"	"	"	"	"	"	3,800	19	89%
University of Sydney Australia								
(Jones, 2000)	pure	2D	?	texture, shape	binary trees	16,220	46	76.3%
							80	70.7%
University of Vigo Spain								
(Rodríguez-Damian et al., 2006)	pure	2D	morph, snake	texture, shape	neural net, SVM	100	3	89%
Delft University of Technology Netherlands								
(Chen et al., 2006)	pure	3D	thresh.	texture, shape, specialized	"linear normal classifier"	254	3	97.2%
University of Freiburg Germany								
(Ronneberger et al., 2002a) and table 10.1	pure	3D confocal	not needed	14 gray scale invariants	SVM	385	26	92%
table 10.2	"	"	"	"	SVM (grid search)	"	"	95.6%
(Wang, 2004)	"	"	"	14 gray scale invariants	"	"	"	97.9%
(Wang, 2004)	"	"	"	38 gray scale invariants (modified)	"	"	"	98.7%
(Reisert and Burkhardt, 2006)	"	"	"	7680 directional SH features	1NN	"	"	94.5%
(Reisert and Burkhardt, 2006)	"	"	"	"	SVM	"	"	96.9%
(Ronneberger et al., 2007) and table 10.4	"	"	graph-cut	896 MiSP invariants	1NN	389	"	99.2%
table 10.5	"	"	"	"	SVM	"	"	99.5%
(Baumgartner, 2005) and table 10.17	pure	3D, 4 channels	thresh.	multi-channel features	SVM	1117	20	96.5%

pure: pure pollen samples
SVM: Support Vector Machine

10.3 MiSP Detector on Pollenmonitor Data Set

To detect and segment the pollen grains in the pollenmonitor samples the MiSP detector (described in section 8.2) was applied.

Some typical results of the detection and the subsequent refinement of the border by snakes are shown in figures 10.7, 10.8, 10.9 and 10.10.

Figure 10.7 shows a low density air sample from March, that resulted in a perfect detection, segmentation and subsequent recognition.

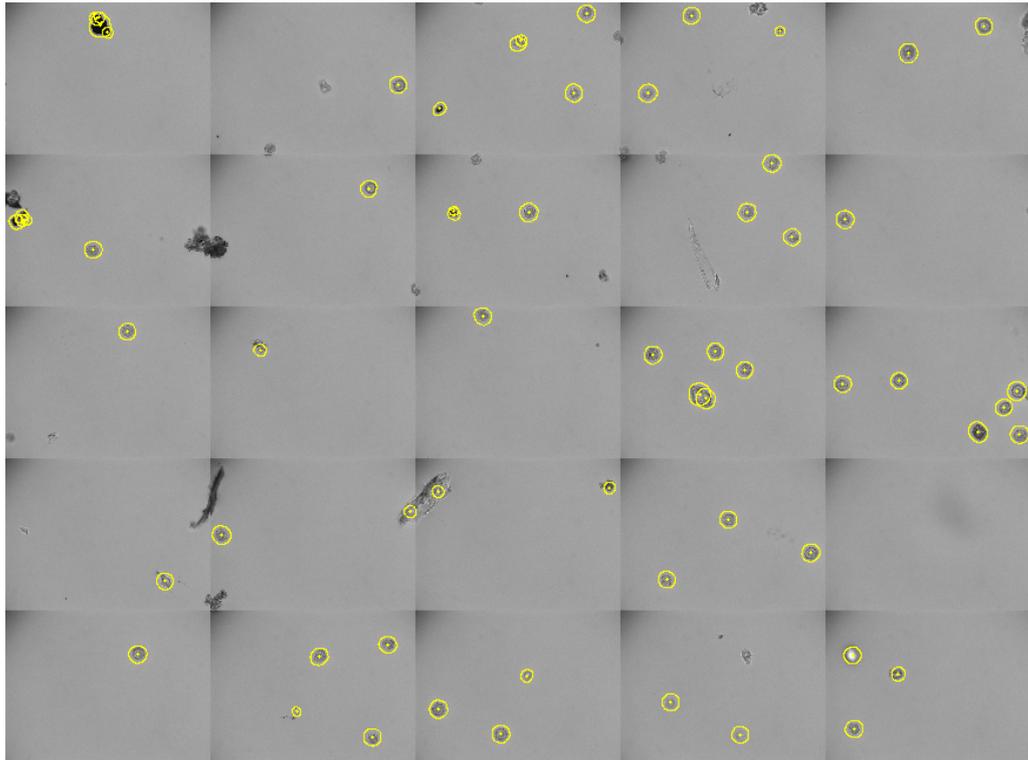
An extreme situation is found in the sample which was taken just a few hours before (figure 10.8). It contains water droplets and dense particle clusters that originate from melted snow flakes. Here the particles are located in very different layers, and some pollen grains have not been detected (e. g. a *Corylus*^(A.5) pollen at the lower right border between the dust particles within the upper left water droplet). The missed detection of such pollen grains (that were in all probability transported within the snow flake) is negligible, at least in the present application which provides raw data for the prediction of the allergenic impact. Please note that pollen which are too close to the border of an image stack are rejected by the segmentation.

Another extreme situation is shown in figure 10.9, an overcrowded air sample from April. Here the high amounts of pollen grains corresponds to their concentration in the air and the agglomeration has presumably taken place directly on the sample and not in the air before. Nearly all of the objects which are not marked are touching the border. Only very few pollen were missed by the detection. Compared to the uncertainties (Poisson noise) from the random sampling of an air volume (see section 10.1), this error is also negligible.

In the air sample in figure 10.10 (recorded in May) a further extreme situation is shown: Enormous amounts of *Picea*^(D.9) pollen cover the sample. This non-allergenic pollen taxon exhibits significant departures from a spherical shape and is therefore not correctly detected and segmented. The fractional parts of such pollen were consequently labeled (and recognized) as non-pollen objects.

High resolution results of the segmentation can be found in the appendices A, B, C and D. There for each pollen grain the sharpest layer and the found segmentation border is shown.

Detection and Segmentation



0 100 200 300 400 500 750 1000µm
Scale 60:1

Recognition

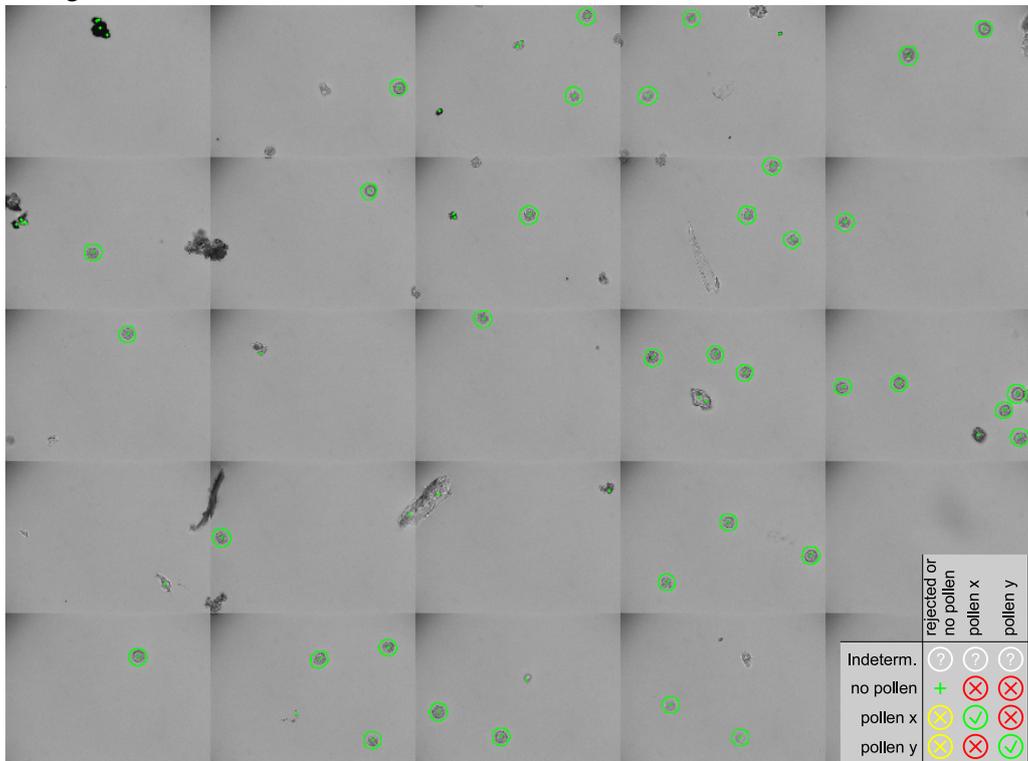
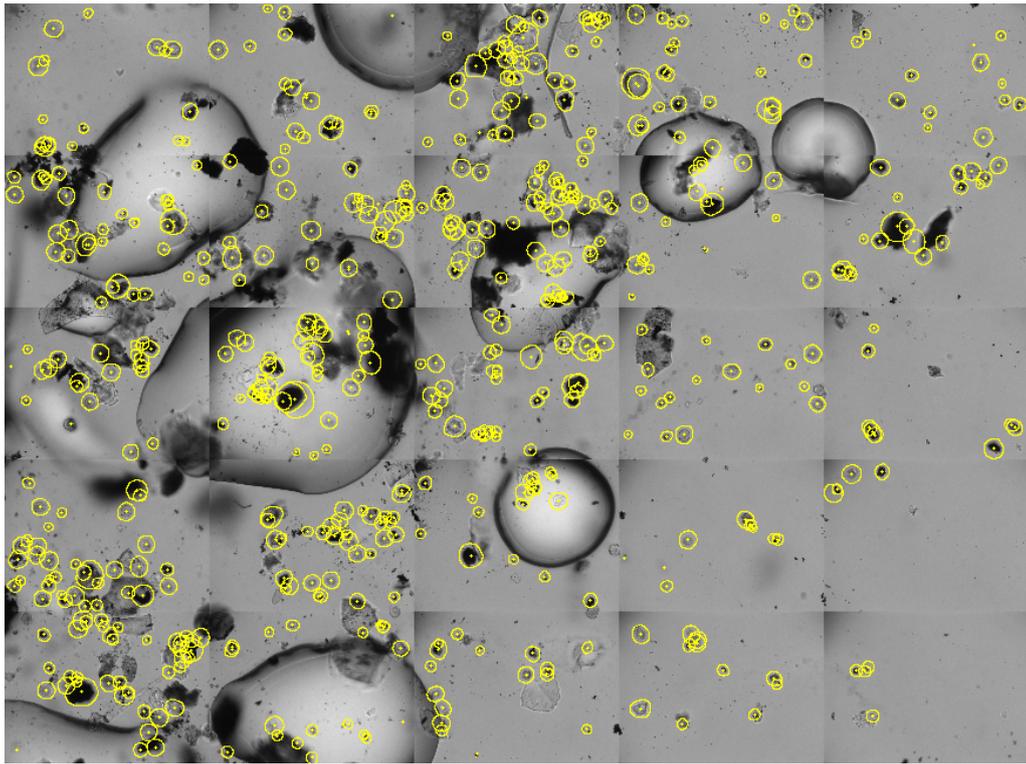


Figure 10.7: Perfect detection, segmentation and recognition on a low density air sample from March 11, 6:00 a.m.

Detection and Segmentation



0 100 200 300 400 500 750 1000µm
Scale 60:1

Recognition

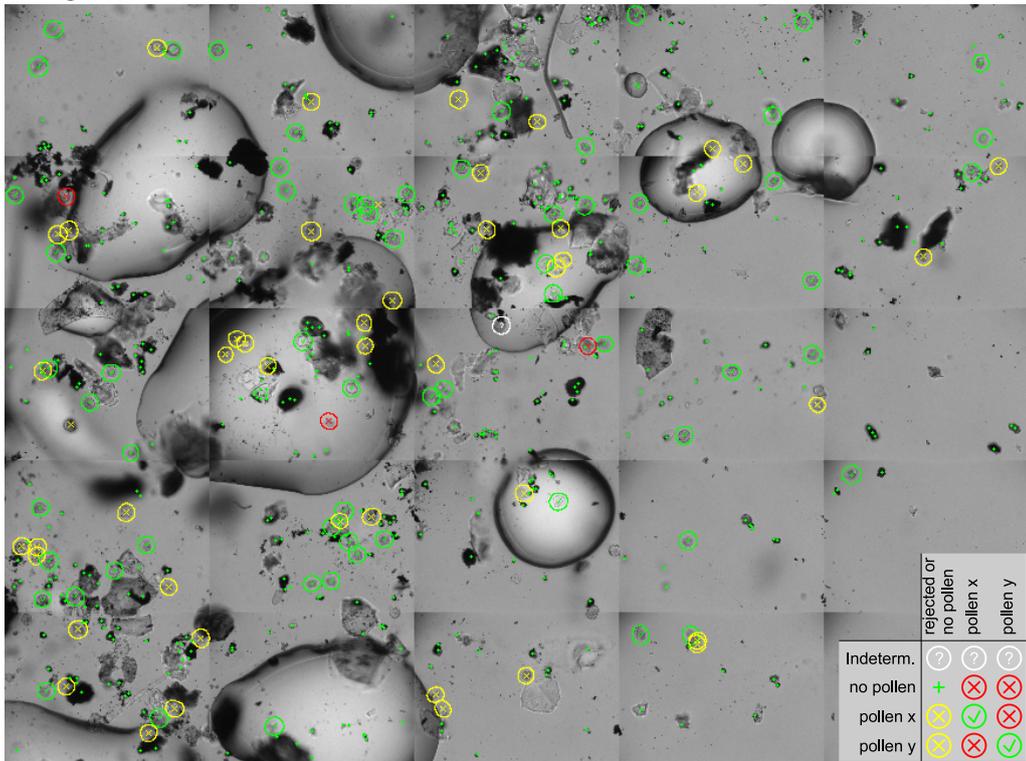
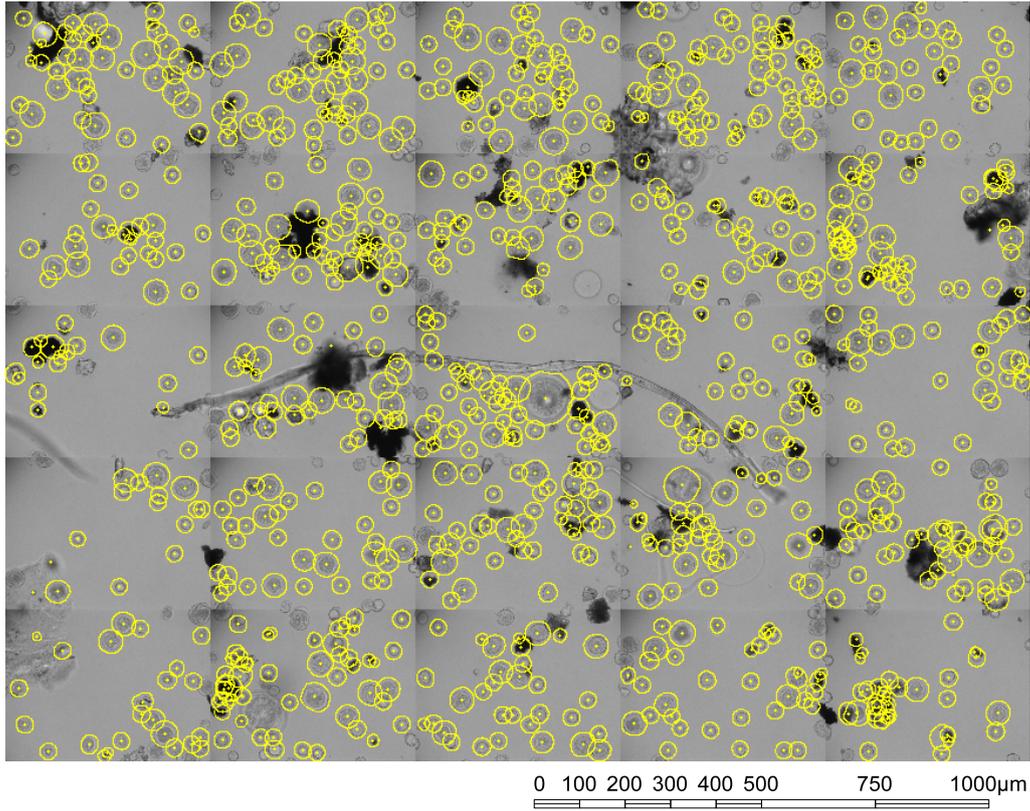


Figure 10.8: Detection, segmentation and recognition results of a highly cluttered air sample from March 11, 2:00 a.m. This air sample contains water droplets and dense particle clusters that originate from melted snow flakes.

Detection and Segmentation



Recognition

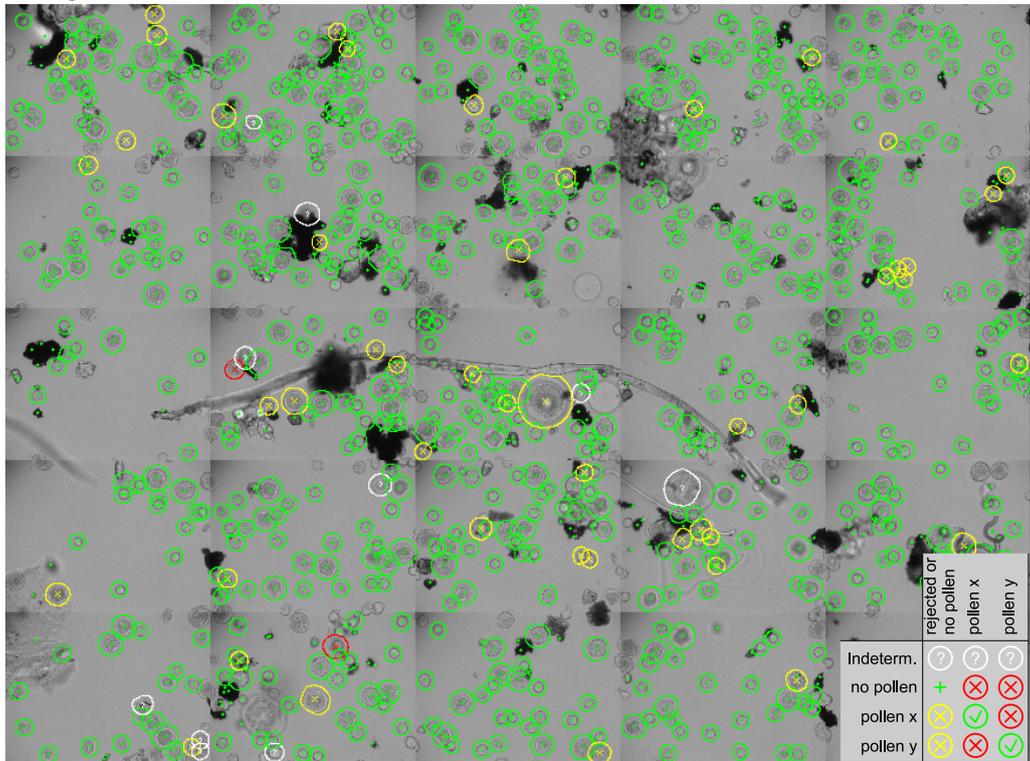
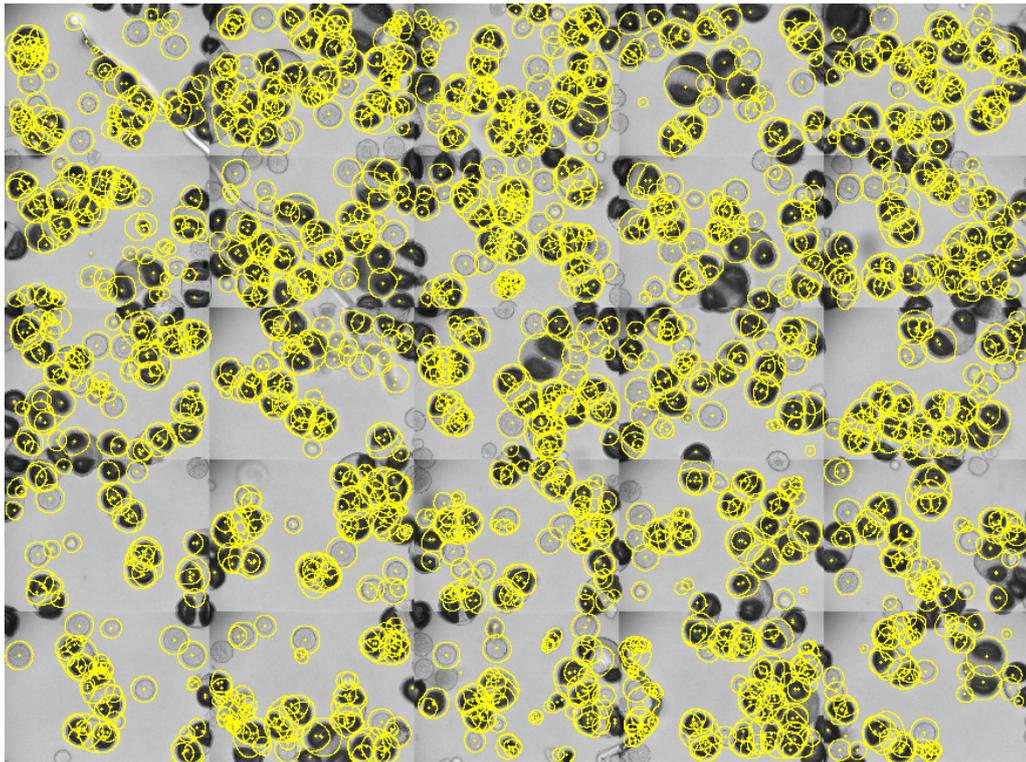


Figure 10.9: Detection, segmentation and recognition results of a high density air sample from April 20, 8:00 p.m.

Detection and Segmentation



0 100 200 300 400 500 750 1000µm

Scale 60:1

Recognition

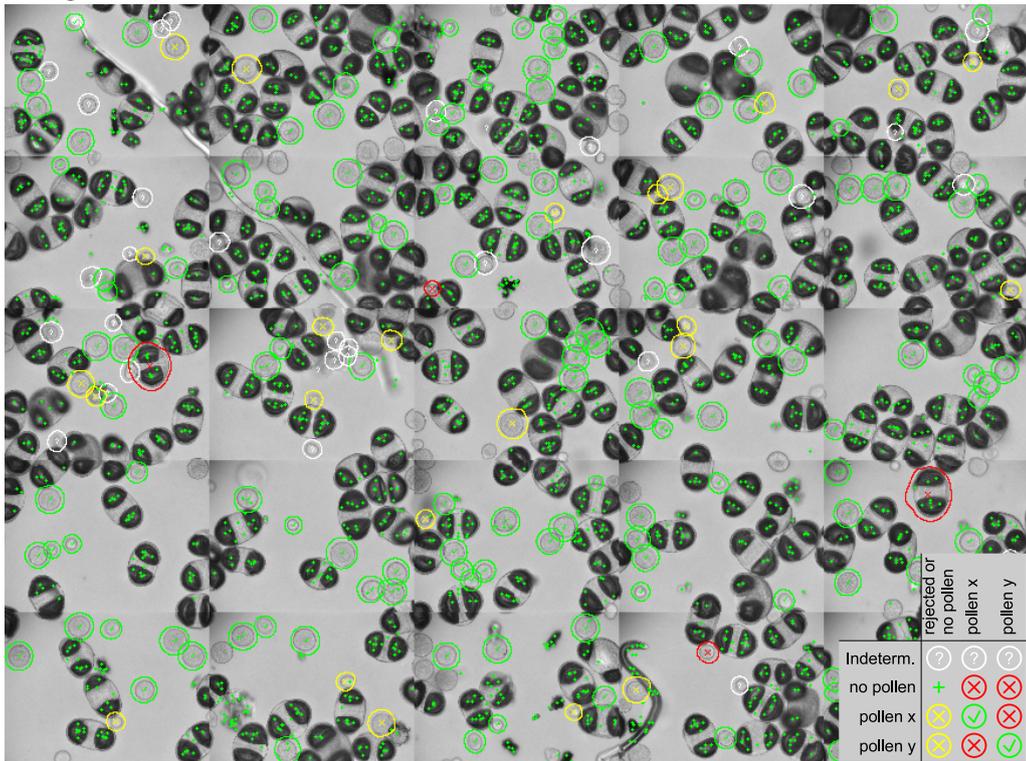


Figure 10.10: Detection, segmentation and recognition results of an air sample from May 4, 5:00 p.m from Zürich. High amounts of *Picea*^(D,9) pollen grains cover the sample. This type of non-spherical pollen cannot be detected and segmented with the proposed algorithms. The segmented fractional parts of these pollen grains are labeled as “non-pollen”

10.4 MiSP Invariants on Pollenmonitor Data Set

10.4.1 Features

The feature extraction is performed by the MiSP invariants described in chapter 9. For the application on the pollenmonitor data set (rotational invariance only around the z-axis), \mathbf{q} from equation (9.34) is split into a radial distance q_r from the segmentation border and the z-distance q_z from the central plane.

The best sampling of the parameter space of the kernel functions (corresponding to the inner class deformations of the objects), was found by cross validation on the training data set. The selected parameters for the “structural” MiSP invariants are listed in table 10.7. The resulting feature vector has $N_{q_r} \times N_{q_z} \times N_c \times n = 31 \times 11 \times 16 \times 16 = 87296$ components. All features are proportional to a scaling of the raw gray values with a factor α . So we can reach invariance to this type of gray value transformations by the application of a post-processing (see section 6.5.1). We have chosen the normalization of the feature vector to unit sum here.

Table 10.7: Selected parameters for the computation of the “structural” MiSP invariants on the pollenmonitor data set

parameter	value(s)	description
N_{q_r}	31 steps from $0\mu\text{m}$ to $9.6\mu\text{m}$ inside	radial localization of the “shells” relative to the segmentation border
N_{q_z}	11 steps from $7.5\mu\text{m}$ above to $7.5\mu\text{m}$ below	vertical localization of the “shells” relative to the sharpest layer
N_c	16 steps from -1 to +1	radial component of the normalized gradient
n	16	Fourier coefficients for the approximation of the kernel function
σ_d	0.1	relative deformations for the local deformation model

For the “shape” MiSP invariants that are computed using the kernel from eq. 9.22 on the synthetic channel, the only free parameter is the number of Fourier coefficients n that are used for the approximation of the kernel function. Here $n = 8$ was found to give the best results.

For the combination of the “structural” and the “shape” features into one feature vector an additional parameter has to be found. It depends on the relative amount of information that is contained in the structure and in the shape of the considered objects and the original range of the features. For the presented application a relative weighting of 0.01 for the shape features was found to return the best results.

10.4.2 Training and Test Data Sets

The pollenmonitor data set is described in detail in section 3.2. For the experiments the 326 air samples from Freiburg and Zürich were split into approximately equal sized sets by using the air samples with an even index as training set and that with odd index as test set. To be able to test the transferability of the approach, the air samples from Zürich were not used for training.

For the training set only the “clean” pollen (not agglomerated, not contaminated, well segmented, etc.) were manually selected from the air samples with an even index from Freiburg, resulting in 1,968 pollen grains. We only use clean pollen, because the addition of contaminated pollen to the training set has bad effects on the final precision: The feature extraction cannot distinguish between the relevant

structures and the contamination, such that there is always a high risk that an unknown contaminated pollen will be assigned to that pollen with a similar contamination pattern instead of that pollen with the same taxon.

To find a representative set of “non-pollen” objects from the available 85,000 particles a 2-step approach was used. The first training set was combined from the “clean” pollen and the non-pollen from a few air samples. With this training set a support vector machine was trained, using the RBF-kernel (radial basis function) and the one-vs-rest multi-class approach. The best parameters were selected using cross-validation on this training data set.

After that the resulting SVM was used to classify all remaining particles from the air samples with an even index. The incorrectly classified “non-pollen” objects were added to the final SVM-training set, resulting in 1048 non-pollen objects.

Due to the different characteristics of the images recorded in Zürich (because of the different location and the misadjustment of the optics) we present not only the full test data set (air samples with an odd index) but also the partial results for each city. The total numbers of training and test pollen in the different data sets are listed in table 10.8

Table 10.8: Training and Test Sets

	total	training set Freiburg	selected for SVM training	test set Freiburg	test set Zürich	test set Freiburg + Zürich
number of air samples:	326	147		146	16	162
corresponding air volume (in m ³):	282.94	135.28		128.25	9.31	137.56
number of recorded stacks: (1392 × 1040 × 70 voxel each)	27 280	13 043		12 365	898	13 263
transmitted light raw data (in Tera Byte):	2.764	1.322		1.253	0.091	1.344
number of dust particles:	171 010	85 313	1 048	63 342	16 101	79 443
total number of pollen:	22 750	10 976	1 968	9 994	1 815	11 759
number of indetermin. pollen:	3 838	1 849	0	2 623	261	2 884
number of pollen taxa:	36	31	20	33	17	33

10.4.3 Optimization for High Precision or High Recall

Depending on the application a high recall or a high precision of the results is preferable. In a fully automated environment, where the results shall be used directly without further human intervention (e. g. for the indication of the current pollen concentration on a web site or for pollen alerts that are sent via SMS to allergic people) we identified in section 10.1.2 that a low number of false-positives (and therefore a high precision) is the most important factor of the system.

If the pollen recognition is used, e. g., for scientific studies where more precise concentration measurements are needed, it might be advantageous to optimize the system for a high recall with a subsequent manual removal of the false-positives by a pollen expert. This procedure is still much faster than a manual counting of the whole sample.

The used one-vs-rest multi-class approach for the support vector machines provides an easy way of optimization for high recall or optimization for a high precision: During the classification process each two-class SVM (which is trained to separate one specific pollen taxon from the rest) returns a decision value that is positive if the test object may belong to this taxon or negative if it may belong to the rest.

There are now two main possibilities to combine these decision values to a final classification result: The first is to always assign the object to that taxon with the highest decision value. This forces the classifier to assign every object to one of the available classes and therefore usually results in a high recall but with a high number of false positives. The alternative is to only assign the test object to a

class if exactly one of the decision values is positive and all the others are negative. In cases with more than one positive decision value or only negative decision values, the object is rejected and classified as unknown. With this technique only the “sure” candidates are assigned to a specific class and the “unsure” are rejected, which results usually in a low number of false-positives (and therefore in a high precision) at the expense of the recall.

10.4.4 Results

For all experiments, the used training data set was the same, containing only pollen from Freiburg. The trained SVM was applied to classify different test data sets with and without rejection. The results were grouped for the different pollen taxa according to their allergenic relevance.

The grouped results are listed in table 10.9. Those results, that were taken into account for the optimization of the system are set in boldface:

The most important demand in pollen forecasts is a high precision for the highly allergenic pollen taxa. All other pollen taxa are not included in the forecasts at present.

For other applications of the MiSP invariants the most interesting fact is their performance on a large number of different objects. Here either a high precision or a high recall might be of interest.

Table 10.9: Overall recognition results on the pollenmonitor data set. The results that were taken into account for the optimization of the system are set in boldface.

		Freiburg	Zürich	Freiburg+Zürich
		precision / recall	precision / recall	precision / recall
with rejection (high precision)	all pollen	97.2% / 84.6%	94.3% / 83.1%	96.7% / 84.3%
	highly allergenic	98.7% / 87.5%	92.4% / 64.0%	98.5% / 86.5%
	mod. allergenic	96.4% / 87.8%	95.0% / 75.9%	96.1% / 85.5%
	lowly allergenic	92.8% / 85.3%	97.4% / 94.7%	96.1% / 91.9%
no rejection (high recall)	all pollen	92.0% / 91.4%	89.2% / 92.6%	91.5% / 91.6%_x
	highly allergenic	94.6% / 93.5%	76.2% / 81.6%	93.8% / 93.0%
	mod. allergenic	89.9% / 93.7%	89.7% / 89.6%	89.9% / 92.9%
	lowly allergenic	79.1% / 90.4%	95.9% / 98.4%	90.5% / 96.0%

The precision and recall per pollen taxon for the full test set “Freiburg+Zürich” are provided in figure 10.11. We can clearly see the effect, that omitting the rejection increases the recall, while the precision is reduced.

10 Experimental Results

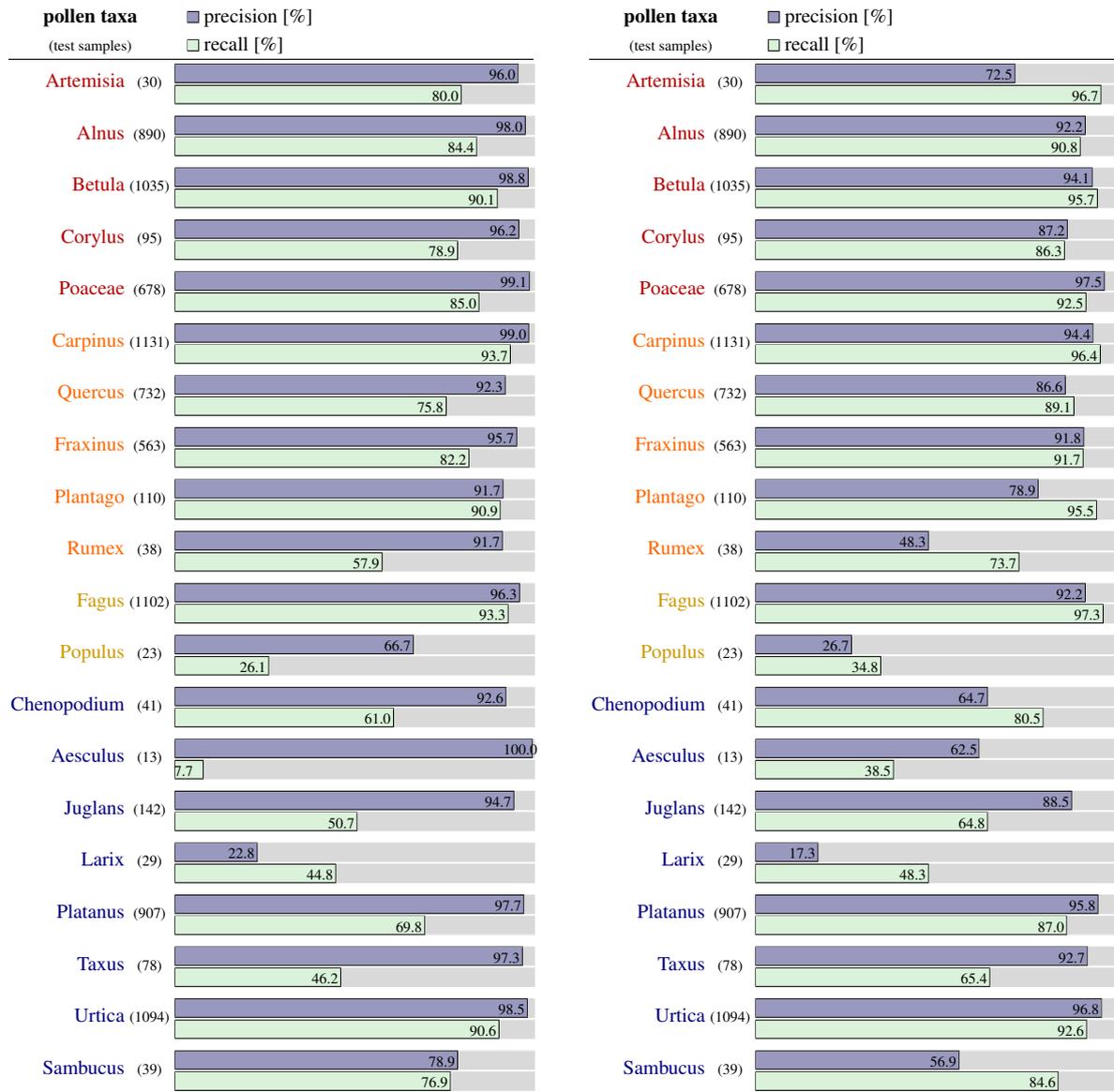


Figure 10.11: Comparison of precision and recall for the classification with rejection (left) and without rejection (right) on the full test set (Freiburg+Zürich). The color of the pollen names (red,orange,yellow,blue) represents their allergenic level (high,moderate,low,none)

10.4.5 Detailed Results

Every set of statistics can only provide a reduced view of the real performance of a system. So we made a considerable effort to find representations of the results which provide a view to the real performance which is as unbiased as possible.

For a better description of the overall recognition rate the full confusion tables for each experiment are provided in the tables 10.10 and 10.11 for the test set “Freiburg”, in the tables 10.12 and 10.13 for the test set “Zürich”, and in the tables 10.14 and 10.15 for the full test set “Freiburg+Zürich”.

In these tables the text color (red, orange, yellow, black) indicates the allergenic level of the taxa (high, moderate, low, none). The background color (red, orange, yellow, gray) is chosen according to the impact of a misclassification in a non-supervised operation of the pollenmonitor. Red indicates a high impact while gray denotes a low impact. Pollen taxa, that were not in the training data set are written in brackets. For the computation of the precision the “indeterm.” pollen are not taken into account.

But even these tables provide only a rough impression of the true performance of the system. Due to the very high variabilities in the air samples, concerning the particle density on the sample, the type of other particles, the degradation and contamination of the pollen grains, etc., the correct or incorrect classification of a certain pollen grain should be weighted with its “difficulty”. But instead of introducing a subjective “difficulty” here, we decided to provide detailed appendices (A, B, C, and D) with a representative subset of the correctly and incorrectly classified pollen grains for each taxon. To obtain an objective subset of the true-positives, the rejected, the false-negatives, and the unknown-positives², the pollen grains in each group were sorted by the decision value of the support vector machine, which can be interpreted as a kind of certainty for the classification. From this sorted list every n 'th pollen grain was taken (starting with the “best” down to the “worst”), where n was computed such that the number of retrieved images fits on a double page. The decision values are provided there below each pollen grain, and for each incorrectly classified pollen grain the true and the classified taxon is listed. (In the electronic version of this document each pollen name is hyper-linked to its page in the appendix). For each pollen grain only the sharpest layer from the 3D stack is displayed together with the found segmentation border. The reduction to only the sharpest layer hides important information in several cases, especially for the incorrectly classified pollen grains, where the “reasons” for the incorrect classification are often located in the other layers. But this 3D information was considered to be not so important as the display of many different individuals of this pollen taxon on the double page.

²the unknown-positives are those pollen grains which were labeled as “indeterm.” by the pollen experts, which means that they were not able to unequivocally classify them.

Table 10.10: Confusion Table for Freiburg; Highest Precision

True Class	number of training samples	number of test samples	Classified As...																				recall (%)	rec. per group (%)
			rejected	no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Chenopodium	Aesculus	Juglans	Larix	Platanus	Taxus		
indeterm.	-	2623	(710)	(1482)	(3)	(7)	(22)	(5)	(36)	(13)	(52)	(16)	(49)	(16)	(33)	(1)	(4)	(1)	(43)	(10)	(117)	(3)		
no pollen	1048	63342	1486	61765	-	7	2	1	1	4	16	2	-	-	12	-	1	1	14	7	-	15	8	97.5
Artemisia	24	30	6	-	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	80
Alnus	145	882	94	41	745	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	84.5
Betula	231	936	55	8	3	870	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	92.9 87.5
Corylus	44	94	13	6	-	-	75	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	79.8
Poaceae	424	672	83	7	-	1	-	572	4	2	-	2	1	-	-	-	-	-	-	-	-	-	-	85.1
(Secale)	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Ambrosia)	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	(0)
Carpinus	193	1110	47	9	-	-	4	-	1047	-	-	-	-	1	1	-	-	1	-	-	-	-	-	94.3
Quercus	59	283	60	7	-	1	-	-	1	209	3	1	-	1	-	-	-	-	-	-	-	-	-	73.9
Fraxinus	160	538	76	9	-	-	-	-	-	-	447	-	-	-	2	-	-	-	-	4	-	-	-	83.1 87.8
Plantago	63	109	8	-	-	1	-	-	-	1	-	99	-	-	-	-	-	-	-	-	-	-	-	90.8
Rumex	14	36	15	-	-	-	-	-	-	-	-	1	20	-	-	-	-	-	-	-	-	-	-	55.6
(Ostr. Carp.)	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
Fagus	135	311	25	4	-	-	-	2	-	1	-	-	-	279	-	-	-	-	-	-	-	-	-	89.7
Populus	13	23	13	-	-	-	-	-	-	2	1	-	-	6	-	-	-	-	-	-	1	-	-	26.1
(Castanea)	-	11	5	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Salix)	-	13	6	2	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	3	-	-	-	(0)
Chenopodium	17	41	13	-	-	-	-	-	-	-	-	3	-	-	-	25	-	-	-	-	-	-	-	61
Aesculus	2	13	9	2	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	7.69
Juglans	25	104	35	3	-	-	-	-	1	-	-	-	-	5	-	-	60	-	-	-	-	-	-	57.7
Larix	3	29	2	14	-	-	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-	44.8
Platanus	161	847	220	21	-	1	-	-	-	-	11	-	-	-	-	-	-	-	594	-	-	-	-	70.1
Taxus	27	72	28	8	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	34	-	-	47.2
Urtica	165	1094	46	55	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	991	-	90.6
Sambucus	63	39	5	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30	-	76.9
(Acer)	-	3	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Humulus)	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Compositae)	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Picea)	-	5	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	(0)
(Pinus)	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Tilia)	-	5	4	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Ulmus)	-	12	11	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Cyperaceae)	-	2	2	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	(0)
precision (%)					96	98.3	98.9	97.4	99.5	99.1	87.8	95.7	92.5	90.9	93.6	66.7	92.6	100	96.8	43.3	97.7	97.1	98.5	78.9
prec. per group (%)						98.7				96.4				92.8					96.7					

Training data set: Freiburg (even air sample indices)

Test data set: Freiburg (odd air sample indices)

Classification Scheme: Reject, if 0 or more than 1 positive decision values

Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence

Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)

Classifier: SVM (RBF-kernel, One-vs-Rest)

Overall Precision: 97.2% (6141/6320)

Overall Recall: 84.6% (6141/7263)

Table 10.11: Confusion Table for Freiburg; Highest Recall

		Classified As...																				rec. per group (%)		
True Class	number of training samples	no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Chenopodium	Aesculus	Juglans	Larix	Platanus	Taxus	Urtica		Sambucus	recall (%)
	indeterm.	- 2623	(1813)	(11)	(29)	(50)	(12)	(66)	(24)	(100)	(45)	(76)	(44)	(53)	(6)	(20)	(1)	(3)	-	(88)	(24)	(150)	(8)	
no pollen	1048 63342	63053	1	38	23	2	3	32	38	4	1	10	37	9	8	-	4	21	14	-	29	15	99.5	
Artemisia	24 30	1	29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96.7
Alnus	145 882	69	1	801	10	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	90.8
Betula	231 936	20	-	6	909	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	97.1 93.5
Corylus	44 94	12	-	-	-	82	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87.2
Poaceae	424 672	23	-	-	4	1	622	7	6	-	3	3	-	1	-	-	-	-	-	1	1	-	-	92.6
(Secale)	- 1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	(0)
(Ambrosia)	- 2	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	(0)
Carpinus	193 1110	21	-	-	5	-	-	1075	3	-	-	-	3	1	-	-	2	-	-	-	-	-	-	96.8
Quercus	59 283	17	-	3	1	1	-	1	244	7	2	-	2	2	1	-	-	-	1	1	1	-	-	86.2
Fraxinus	160 538	20	1	4	-	-	1	-	2	496	3	-	-	3	-	1	-	-	-	7	-	-	-	92.2 93.7
Plantago	63 109	-	-	1	1	-	-	1	2	-	104	-	-	-	-	-	-	-	-	-	-	-	-	95.4
Rumex	14 36	-	-	-	-	1	2	-	1	-	4	26	-	-	-	-	-	-	-	2	-	-	-	72.2
(Ostr. Carp.)	- 1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
Fagus	135 311	7	-	-	-	-	3	3	1	-	-	-	294	-	-	-	3	-	-	-	-	-	-	94.5
Populus	13 23	1	-	1	-	1	-	-	3	7	-	1	-	8	-	-	-	-	-	-	1	-	-	34.8
(Castanea)	- 11	8	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	2	(0)
(Salix)	- 13	3	1	-	2	-	-	-	-	1	-	-	-	-	-	-	-	-	4	-	2	-	-	(0)
Chenopodium	17 41	-	-	-	1	-	-	-	-	-	7	-	-	-	33	-	-	-	-	-	-	-	-	80.5
Aesculus	2 13	2	1	2	-	-	-	-	-	1	1	-	-	-	-	5	-	-	1	-	-	-	-	38.5
Juglans	25 104	4	-	-	-	-	1	7	1	-	-	-	19	-	-	72	-	-	-	-	-	-	-	69.2
Larix	3 29	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	-	-	-	-	-	-	48.3
Platanus	161 847	62	4	4	2	-	-	-	-	21	-	9	-	-	1	-	-	734	1	3	6	-	-	86.7
Taxus	27 72	16	1	-	1	1	-	2	1	1	1	-	-	-	-	-	-	-	1	47	-	-	-	65.3
Urtica	165 1094	76	1	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	1	-	1013	1	-	92.6
Sambucus	63 39	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33	-	84.6
(Acer)	- 3	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Humulus)	- 1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Compositae)	- 1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Picea)	- 5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	(0)
(Pinus)	- 1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Tilia)	- 5	2	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
(Ulmus)	- 12	-	-	1	-	-	-	-	9	-	-	-	-	1	1	-	-	-	-	-	-	-	-	(0)
(Cyperaceae)	- 2	1	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(0)
precision (%)			72.5	93	94.8	91.1	98.3	95.5	76	92	82.5	50	82.6	30.8	73.3	62.5	88.9	36.8	95.8	92.2	96.8	57.9		
prec. per group (%)				94.6						89.9			79.1						93.2					

Training data set: Freiburg (even sample indices)
Test data set: Freiburg (odd sample indices)
Classification Scheme: Highest decision value always wins
Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence
Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)
Classifier: SVM (RBF-kernel, One-vs-Rest)
Overall Precision: 92% (6641/7222)
Overall Recall: 91.4% (6641/7263)

Table 10.12: Confusion Table for Zürich; Highest Precision

True Class	number of training samples	number of test samples	Classified As...														recall (%)	rec. per group (%)							
			rejected	no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus			Chenopodium	Aesculus	Juglans	Larix	Platanus	Taxus	Urtica
indeterm.	-	261	(116)	(36)	· (1)	(1)	· (4)	·	(63)	(2)	·	·	(15)	(1)	(1)	· (1)	· (17)	·	(2)	(1)					
no pollen	1048	16101	396	15665	·	·	1	·	·	8	·	·	·	8	·	·	·	22	1	·	·	·	·	97.3	
Artemisia	24	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Alnus	145	8	2	·	·	6	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	75
Betula	231	99	31	2	·	1	63	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	63.6 64
Corylus	44	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	0
Poaceae	424	6	1	·	·	·	·	4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	66.7
Carpinus	193	21	5	1	·	·	·	·	13	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	61.9
Quercus	59	449	88	12	·	·	·	·	·	346	·	·	·	2	·	·	1	·	·	·	·	·	·	·	77.1
Fraxinus	160	25	6	·	·	1	·	1	·	·	16	1	·	·	·	·	·	·	·	·	·	·	·	·	64 75.9
Plantago	63	1	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	100
Rumex	14	2	·	·	·	·	·	·	·	·	·	·	2	·	·	·	·	·	·	·	·	·	·	·	100
Fagus	135	791	37	3	·	·	·	·	·	1	·	·	·	749	·	·	1	·	·	·	·	·	·	·	94.7
Populus	13	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Chenopodium	17	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Aesculus	2	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Juglans	25	38	14	3	·	·	·	·	·	1	·	·	·	8	·	·	12	·	·	·	·	·	·	·	31.6
Larix	3	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Platanus	161	60	19	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	39	·	·	·	·	·	65 51
Taxus	27	6	4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	2	·	·	·	33.3
Urtica	165	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
Sambucus	63	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-
(Acer)	-	8	1	1	·	·	·	·	·	4	1	·	·	1	·	·	·	·	·	·	·	·	·	·	(0)
(Humulus)	-	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	(0)
(Picea)	-	35	6	24	·	·	·	·	·	·	·	·	·	·	·	·	·	5	·	·	·	·	·	·	(0)
(Cyperaceae)	-	3	1	·	·	·	·	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	(0)
precision (%)						75	98.4	0	66.7	92.9	95.3	94.1	50	100	97.4		85.7	0	97.5	100					
prec. per group (%)						92.4				95				97.4			63.9								

Training data set: Freiburg (even sample indices)

Test data set: Zürich (odd sample indices)

Classification Scheme: Reject, if 0 or more than 1 positive decision values

Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence

Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)

Classifier: SVM (RBF-kernel, One-vs-Rest)

Overall Precision: 94.3% (1253/1329)

Overall Recall: 83.1% (1253/1507)

Table 10.13: Confusion Table for Zürich; Highest Recall

True Class	number of training samples	number of test samples	Classified As...														recall (%)	rec. per group (%)								
			no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Chenopodium			Aesculus	Juglans	Larix	Platanus	Taxus	Urtica	Sambucus	
indeterm.	-	261	(76)	(7)	(4)	(1)	(7)	(2)	(86)	(4)	(2)	(4)	(28)	(2)	(2)	(4)	(29)	(2)	(1)							
no pollen	1048	16101	16008	1	11			11	12			4	11		5		36	1			(2)	(1)		99.4		
Artemisia	24	-																							-	
Alnus	145	8	1	7																					87.5	
Betula	231	99	7	5	81	2	2					1			1										81.8 81.6	
Corylus	44	1			1																				0	
Poaceae	424	6					5						1												83.3	
Carpinus	193	21	2					15	4																71.4	
Quercus	59	449	23	1		2	1		408	1	4		3	4		2									90.9	
Fraxinus	160	25	1	1		1				20	1						1								80 89.6	
Plantago	63	1									1														100	
Rumex	14	2										2													100	
Fagus	135	791	11						1				778		1										98.4 98.4	
Populus	13	-																							-	
Chenopodium	17	-																							-	
Aesculus	2	-																							-	
Juglans	25	38	3						2				13		20										52.6	
Larix	3	-																							-	
Platanus	161	60	5														55								91.7 76	
Taxus	27	6	1														1	4							66.7	
Urtica	165	-																							-	
Sambucus	63	-																							-	
(Acer)	-	8	1						4	2			1												(0)	
(Humulus)	-	1					1																		(0)	
(Picea)	-	35	26						2								7								(0)	
(Cyperaceae)	-	3	1						1	1															(0)	
precision (%)				46.7	87.1	0	50	51.7	94.4	87	14.3	33.3	96.4	0	0	87	0	94.8	100	0						
prec. per group (%)																										

Training data set: Freiburg (even sample indices)

Test data set: Zürich (odd sample indices)

Classification Scheme: Highest decision value always wins

Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence

Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)

Classifier: SVM (RBF-kernel, One-vs-Rest)

Overall Precision: 89.2% (1396/1565)

Overall Recall: 92.6% (1396/1507)

Table 10.14: Confusion Table for Freiburg+Zürich; Highest Precision

True Class	number of		Classified As...																				recall (%)	rec. per group (%)				
	training samples	test samples	rejected	no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Chenopodium	Aesculus	Juglans	Larix	Platanus	Taxus			Urtica	Sambucus		
indeterm.	-	2884	(826)	(1518)	(3)	(8)	(23)	(5)	(40)	(13)	(115)	(18)	(49)	(16)	(48)	(2)	(5)	.	(2)	.	(60)	(10)	(119)	(4)				
no pollen	1048	79443	1882	77430	.	7	3	1	1	4	24	2	.	.	20	.	1	.	1	36	8	.	15	8	97.5			
Artemisia	24	30	6	.	24	80		
Alnus	145	890	96	41	.	751	2	84.4		
Betula	231	1035	86	10	.	4	933	1	1	90.1	86.5	
Corylus	44	95	14	6	.	.	.	75	78.9		
Poaceae	424	678	84	7	.	.	1	.	576	4	2	.	2	1	1	85		
(Secale)	-	1	1	(0)		
(Ambrosia)	-	2	1	1	(0)		
Carpinus	193	1131	52	10	.	.	4	.	.	1060	2	.	.	.	1	1	.	.	1	93.7		
Quercus	59	732	148	19	.	1	.	.	.	1	555	3	1	.	3	.	.	.	1	75.8		
Fraxinus	160	563	82	9	.	1	.	.	1	.	.	463	1	.	2	4	82.2	85.5	
Plantago	63	110	8	.	.	1	1	.	100	90.9		
Rumex	14	38	15	1	22	57.9		
(Ostr. Carp.)	-	1	1	(0)		
Fagus	135	1102	62	7	2	.	2	.	.	.	1028	.	.	.	1	93.3		
Populus	13	23	13	2	1	.	.	6	1	.	.	.	26.1	91.9	
(Castanea)	-	11	5	6	(0)		
(Salix)	-	13	6	2	.	.	1	1	3	(0)		
Chenopodium	17	41	13	3	.	.	.	25	61		
Aesculus	2	13	9	2	1	1	7.69		
Juglans	25	142	49	6	1	1	.	.	.	13	.	.	.	72	50.7		
Larix	3	29	2	14	13	44.8		
Platanus	161	907	239	23	.	1	11	633	69.8	76.9	
Taxus	27	78	32	8	1	.	1	36	.	.	.	46.2		
Urtica	165	1094	46	55	1	1	991	.	.	90.6		
Sambucus	63	39	5	4	30	.	76.9		
(Acer)	-	11	1	1	7	1	.	.	1	(0)		
(Humulus)	-	2	1	1	(0)		
(Compositae)	-	1	.	1	(0)		
(Picea)	-	40	6	26	8	(0)		
(Pinus)	-	1	.	1	(0)		
(Tilia)	-	5	4	1	(0)		
(Ulmus)	-	12	11	1	(0)		
(Cyperaceae)	-	5	3	1	1	1	1	(0)		
precision (%)					96	98	98.8	96.2	99.1	99	92.3	95.7	91.7	91.7	96.3	66.7	92.6	100	94.7	22.8	97.7	97.3	98.5	78.9				
prec. per group (%)							98.5					96.1			96.1						95.3							

Training data set: Freiburg (even sample indices)

Test data set: Freiburg + Zürich (odd sample indices)

Classification Scheme: Reject, if 0 or more than 1 positive decision values

Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence

Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)

Classifier: SVM (RBF-kernel, One-vs-Rest)

Overall Precision: 96.7% (7394/7649)

Overall Recall: 84.3% (7394/8770)

Table 10.15: Confusion Table for Freiburg+Zürich; Highest Recall

		Classified As...																								
True Class	number of training samples	no pollen	Artemisia	Alnus	Betula	Corylus	Poaceae	Carpinus	Quercus	Fraxinus	Plantago	Rumex	Fagus	Populus	Chenopodium	Aesculus	Juglans	Larix	Platanus	Taxus	Urtica	Sambucus	recall (%)	rec. per group (%)		
	number of test samples	(1889)	(11)	(36)	(54)	(13)	(73)	(26)	(186)	(49)	(78)	(48)	(81)	(8)	(22)	(1)	(7)	(117)	(24)	(152)	(9)					
indeterm.	- 2884	(1889)	(11)	(36)	(54)	(13)	(73)	(26)	(186)	(49)	(78)	(48)	(81)	(8)	(22)	(1)	(7)	(117)	(24)	(152)	(9)					
no pollen	1048 79443	79061	1	39	34	2	3	43	50	4	1	14	48	9	13	4	57	15	29	16			99.5			
Artemisia	24 30	1	29																					96.7		
Alnus	145 890	70	1	808	10								1												90.8	
Betula	231 1035	27	11	990	3	2					1				1										95.7 93	
Corylus	44 95	12		1	82																				86.3	
Poaceae	424 678	23		4	1	627	7	6		3	3		1	1				1	1						92.5	
(Secale)	- 1												1												(0)	
(Ambrosia)	- 2														2										(0)	
Carpinus	193 1131	23			5			1090	7				3	1			2								96.4	
Quercus	59 732	40		4	1	3	1	1	652	8	6		5	6	1	2		1	1						89.1	
Fraxinus	160 563	21	1	5		2			2	516	4			3		1			8						91.7 92.9	
Plantago	63 110			1	1			1	2		105														95.5	
Rumex	14 38					1	2		1		4	28							2						73.7	
(Ostr. Carp.)	- 1								1																(0)	
Fagus	135 1102	18					3	3	2				1072			4									97.3	
Populus	13 23	1		1		1			3	7	1		8							1					34.8	
(Castanea)	- 11	8														1									(0)	
(Salix)	- 13	3	1		2					1								4			2				(0)	
Chenopodium	17 41				1						7				33										80.5	
Aesculus	2 13	2	1	2						1	1					5			1						38.5	
Juglans	25 142	7					1	7	3								92								64.8	
Larix	3 29	15																14							48.3	
Platanus	161 907	67	4	4	2					21	9					1			789	1	3	6			87 86.6	
Taxus	27 78	17	1		1	1			2	1	1	1							2	51					65.4	
Urtica	165 1094	76	1									2							1		1013	1			92.6	
Sambucus	63 39	6																				33			84.6	
(Acer)	- 11	1							7	2			1												(0)	
(Humulus)	- 2					1	1																		(0)	
(Compositae)	- 1	1																							(0)	
(Picea)	- 40	28						2										10							(0)	
(Pinus)	- 1	1																							(0)	
(Tilia)	- 5	2							3																(0)	
(Ulmus)	- 12			1					9					1	1										(0)	
(Cyperaceae)	- 5	2						1	3	1															(0)	
precision (%)			72.5	92.2	94.1	87.2	97.5	94.4	86.6	91.8	78.9	48.3	92.2	26.7	64.7	62.5	88.5	17.3	95.8	92.7	96.8	56.9				
prec. per group (%)				93.8					89.9				90.5						91.1							

Training data set: Freiburg (even sample indices)
Test data set: Freiburg + Zürich (odd sample indices)
Classification Scheme: Highest decision value always wins
Data type: Real air samples from pollenmonitor, 3D transmitted light, no fluorescence
Features: 87,296 MiSP invariants ("structure" and "shape"; only φ_z -rotation)
Classifier: SVM (RBF-kernel, One-vs-Rest)
Overall Precision: 91.5% (8037/8787)
Overall Recall: 91.6% (8037/8770)

Quantification of the False Positives

Another important measure that was introduced in section 10.1.2 is the “false-positives concentration” (false-positive events per m^3 of air), which allows to directly see the influence of the false-positives for the prediction of the allergenic impact in the pollen forecasts. Due to high seasonal variation of the particles that cause false-positive events (in nearly all cases these are other morphologically similar pollen taxa) a mean value of this false-positives concentration over the year does not provide any useful information. Furthermore the available threshold values for the allergenic impact of certain pollen taxa at a certain concentration in the air are based on daily mean values.

So for a meaningful representation the true and the predicted daily mean pollen concentrations were computed and were compared to the corresponding false-positive concentration. This is shown in figures 10.12 and 10.13 for *Corylus*^(A.5), *Alnus*^(A.2) and *Betula*^(A.4) pollen, and in figures 10.14 and 10.15 for *Poaceae*^(A.6) and *Artemisia*^(A.1) pollen.

10.4.6 Comparison to the State of the Art

The comparison of our results on real-world samples to the state of the art (see section 1.2.1) is listed in table 10.16.

The results that are reached with the presented approaches by far outperform all other published results as well as in the number of considered pollen taxa as in the recognition rate and the precision.

Table 10.16: Comparison of our results to the results of other groups (**Real-world Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate	precision
Bangor University UK (France et al., 2000)	pollen +debris	2D	thresh.	neural net.	neural net.	204	3	81.7%	96.7%
INRIA, Sophia-Antipolis France (Boucher et al., 2002)	air	3D color	morph. +color	shape, color, specialized	Mahalanobis	350	30	77%	?
Caltech and New York University, US (Ranzato et al., 2007)	air	2D	DOG + circle fit	texture	Fisher linear discr. + GMM	3104	8	64,9%	30%
University of Freiburg Germany (Ronneberger et al., 2007) and table 10.14	air	3D	misp detector + Snakes	87,296 MiSP invariants (structure + shape)	SVM (reject)	22,750	33	84.3%	96.7%
table 10.15	"	"	"	"	SVM (no rej.)	"	"	91.6%	91.5%
table 10.14	"	"	"	"	SVM (reject)	"	5 h.a.	86.5%	98.5%

DOG: Difference of Gaussians
GMM: Gaussian mixture model
h.a.: highly allergenic pollen taxa

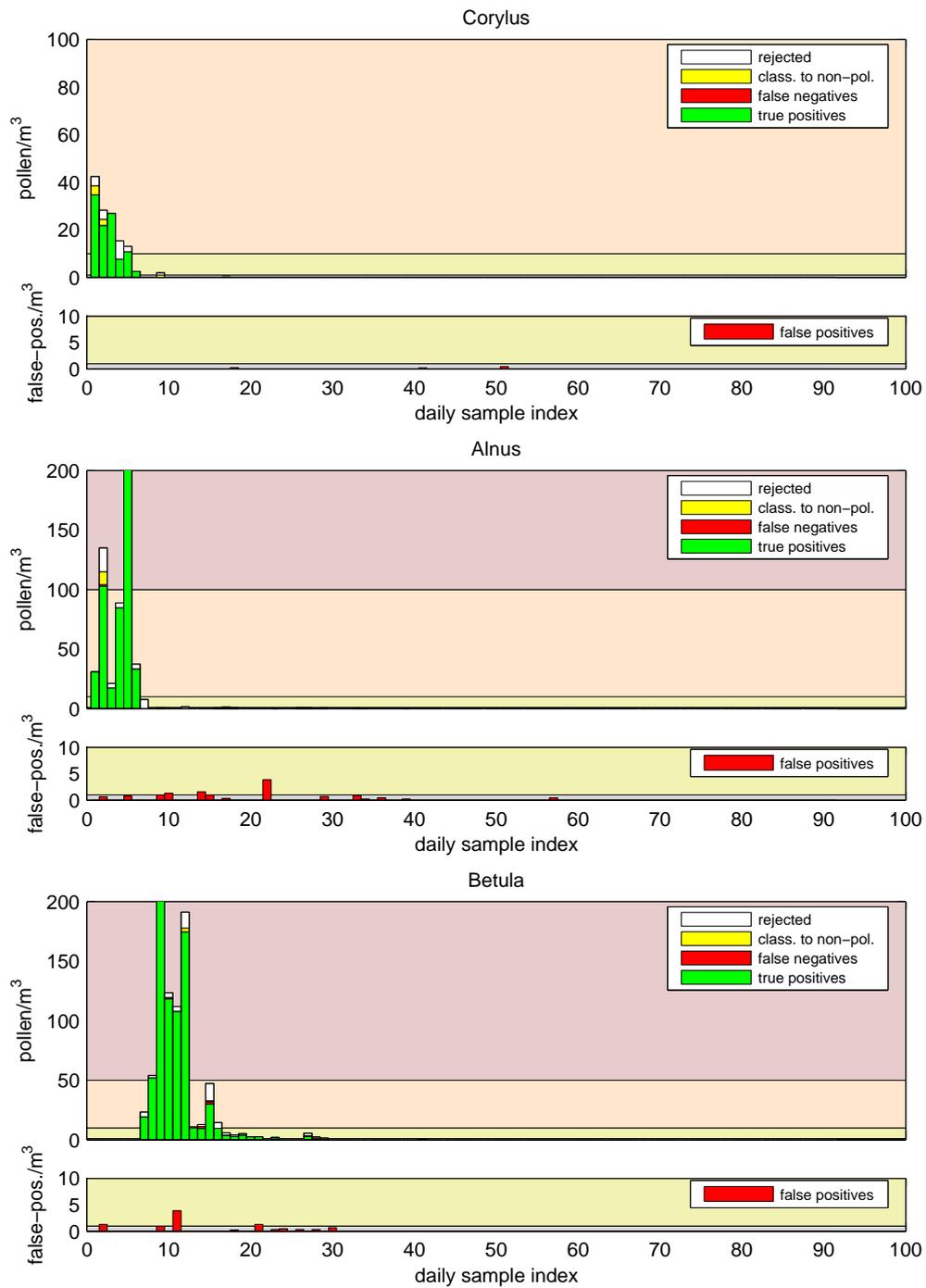


Figure 10.12: Daily mean concentrations (true and predicted) for *Corylus*^(A.5), *Alnus*^(A.2) and *Betula*^(A.4) pollen compared to the according false-positive-concentrations. The background color in the plots represents the allergic stress at the particular concentration: red=high, orange=moderate, yellow=low, gray=none. These results are obtained by the classification scheme “with rejection”

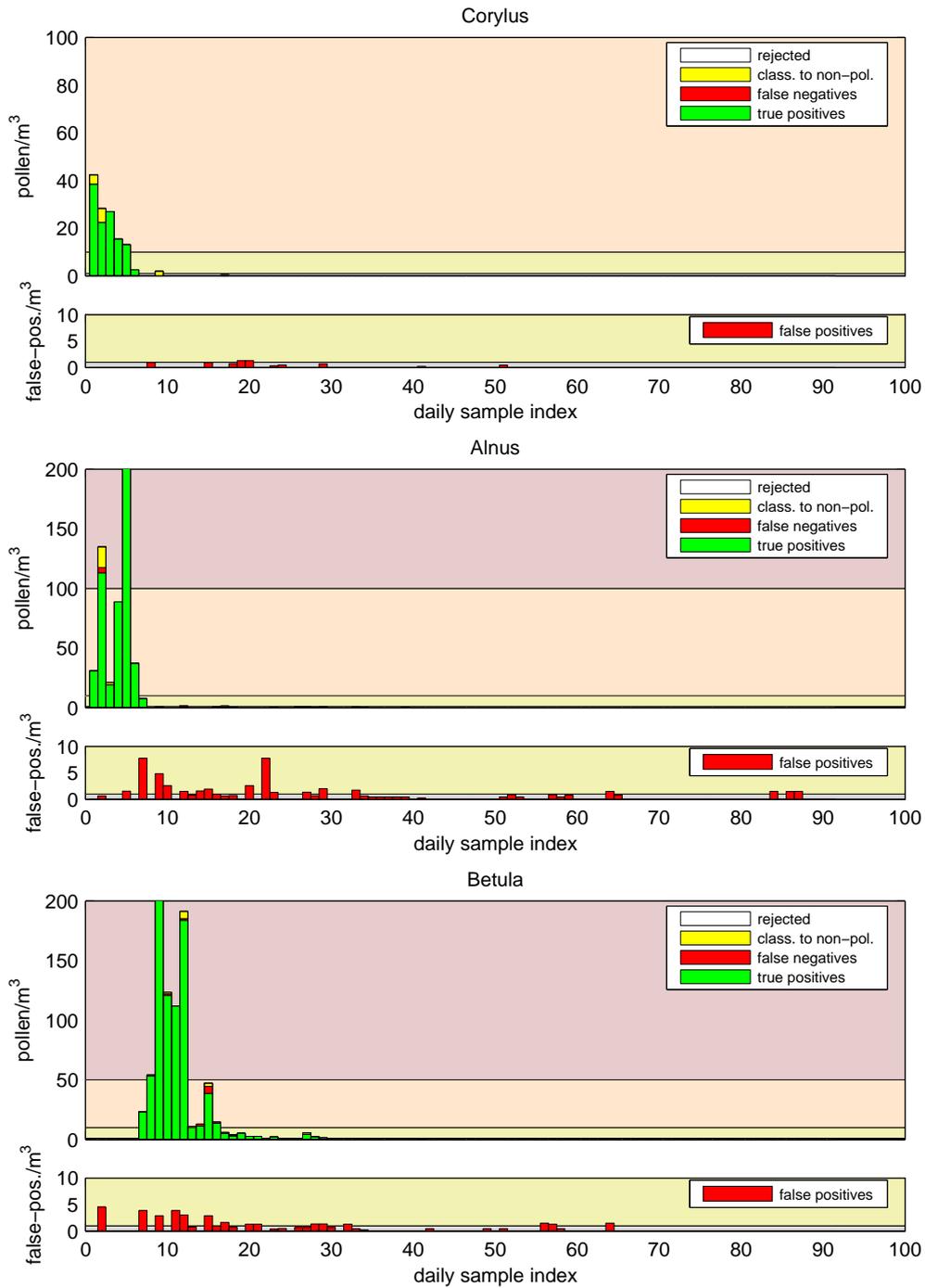


Figure 10.13: Daily mean concentrations (true and predicted) for *Corylus*^(A.5), *Alnus*^(A.2) and *Betula*^(A.4) pollen compared to the according false-positive-concentrations. These results are obtained by the classification scheme “no rejection”

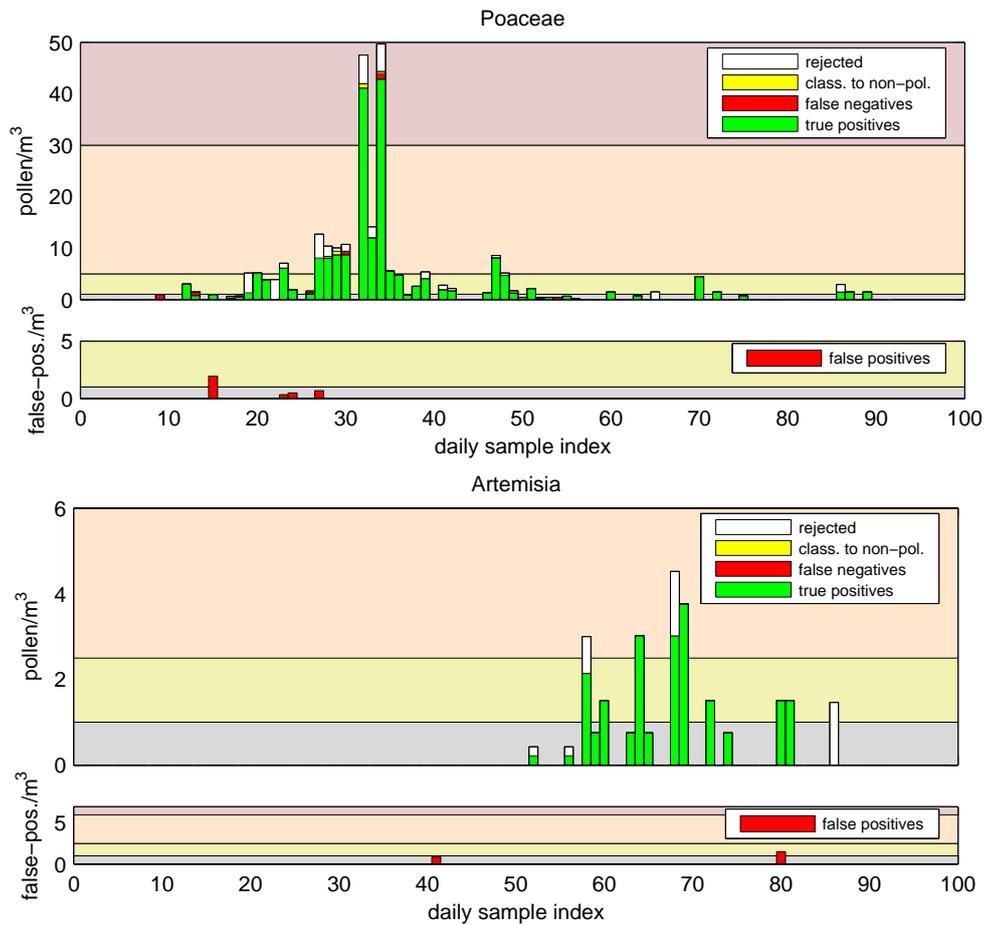


Figure 10.14: Daily mean concentrations (true and predicted) for *Poaceae*^(A.6) and *Artemisia*^(A.1) pollen compared to the according false-positive-concentrations. These results are obtained by the classification scheme “with rejection”

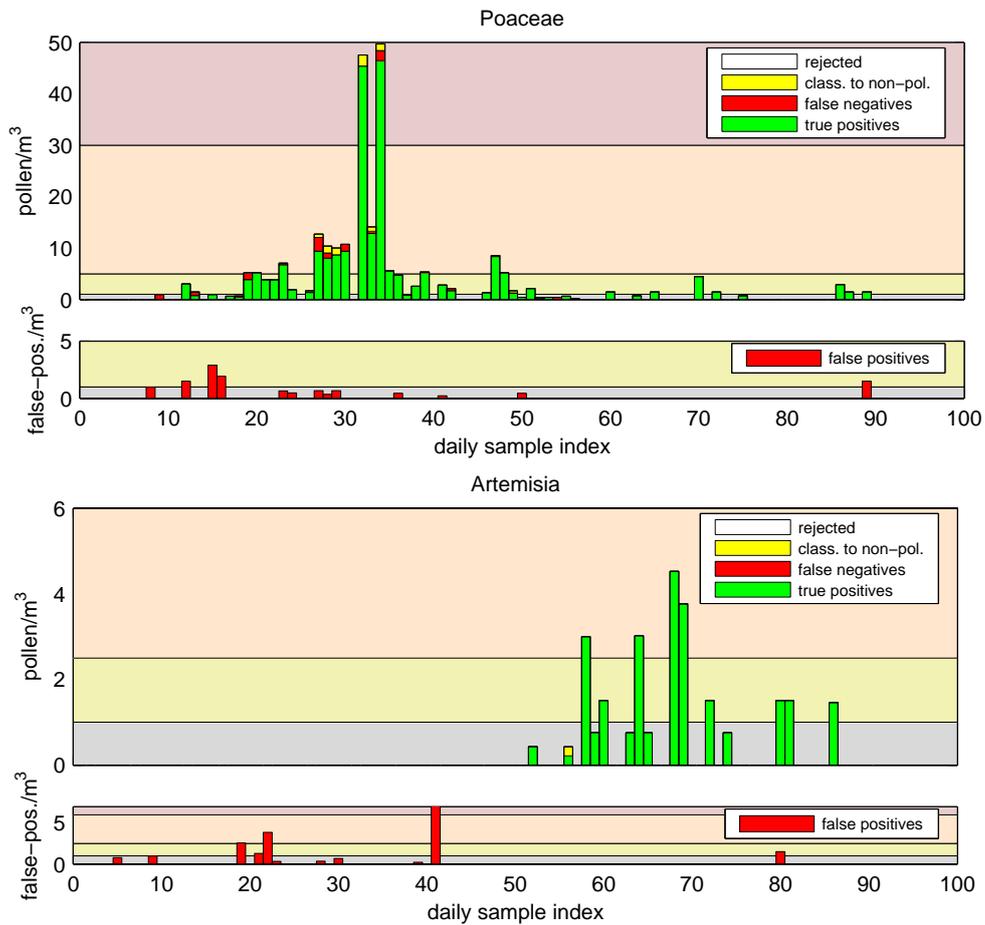


Figure 10.15: Daily mean concentrations (true and predicted) for *Poaceae*^(A.6) and *Artemisia*^(A.1) pollen compared to the according false-positive-concentrations. These results are obtained by the classification scheme “no rejection”

10.5 Summary of Further Results

Many of the theoretical results presented in this thesis have been extended and experimentally verified in close collaboration with several students during the supervision of their master thesis.

Because many of these results significantly influenced the presented work and furthermore allow the reader to get an idea how the presented theoretical results perform in a real application, we present a short summary of the results of these studies here.

10.5.1 Deconvolution

The very promising results on the high-resolved confocally recorded pollen grains and the first disillusioning results with the same features on data sets recorded with a conventional microscope gave reason for several studies on an improvement of the raw data quality of such conventionally recorded data sets.

The most obvious technique is to try a pure software-solution, the deconvolution of the recorded data with the point spread function of the used microscope. The problem of existing approaches is that they are optimized to produce plausible, good-looking results of the input data and use several assumptions (continuity, discontinuity, edges, etc) for the “reconstruction” of the original signal.

In collaboration with Thorsten Schmidt a deconvolution technique was developed that is purely based on measured data and does not invent additional information at places where the original information is lost. An example of a deconvolved *Corylus*^(A.5) pollen grain is shown in figure 10.16.

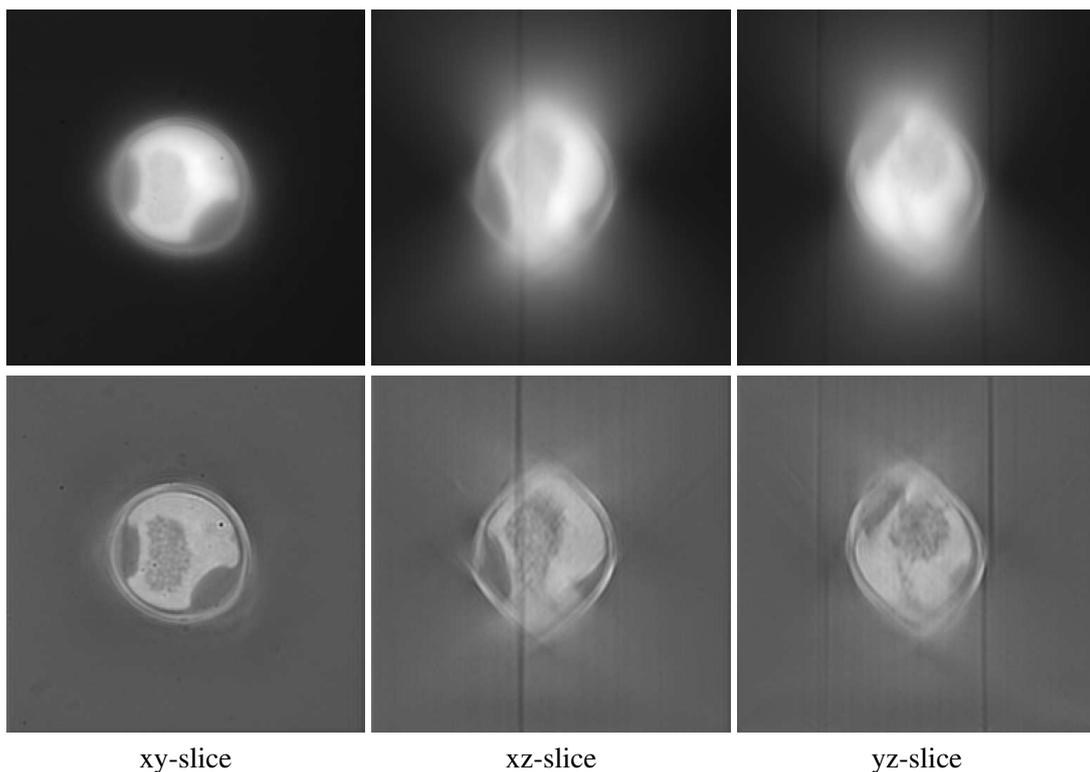


Figure 10.16: Deconvolution of a *Corylus*^(A.5) pollen grain with an assumption free approach that is based only on measured data (from (Schmidt, 2003)). The raw data was recorded using conventional fluorescence microscope equipped with a 100x oil immersion objective with a numerical aperture of 1.3

We can see a clear improvement of the richness of details. But the lost frequency components (the so

called “missing cone” in the Fourier space) introduce significant artifacts at the upper and lower border of the pollen grain. So a 3D rotation of the pollen in real-world can not be modeled by a 3D rotation of the recorded data set. This makes this type of data not well suited for full 3D rotational invariants. The detailed results can be found in (Schmidt, 2003).

10.5.2 Use of Additional Microscopical Techniques

In collaboration with Rafael Baumgartner it was studied whether the utilization of additional microscopical techniques can provide valuable additional information. The studies were done on pollen taxa that are typically found in honey samples. From manually prepared slides a data set with 1117 pollen grains from 20 taxa was recorded. Instead of the usual two channels (fluorescence stacks and transmitted light image stacks with $2/3$ opened aperture), six channels were recorded (see fig 10.17).

By the combination of these different channels a significant improvement of the recognition rates could be reached (see table 10.17). The used features were multi-channel gray scale invariants with a two-point-kernel and multi-channel features based on mutual information and entropy.

Table 10.17: Recognition results on honey pollen data using additional microscopical channels (from (Baumgartner, 2005)). The best result was found by using four channels: transmitted light (with 2 aperture settings: $1/1$ and $1/4$), dark field and fluorescence

Channels	Mean Recognition Rate
transmitted light only	77.19%
dark field only	80.6%
fluorescence only	87.57%
transmitted light + fluorescence	93.83%
four channels	96.5%

The details can be found in (Baumgartner, 2005).

10.5.3 Automatic Kernel Selection with the GVCD

If the HI-framework shall be applied to a new problem, it is necessary to select appropriate kernel functions. In many cases the possibility to integrate prior knowledge via appropriate kernels is of advantage. But if the user has no prior information of the problem it is desirable to have a method that selects automatically appropriate kernels based on a provided sample data set.

In collaboration with Thorsten Schmidt a feasibility study on an automated kernel selection based on the GVCD (see section 5.2.1) was performed. The main challenge turned out to be already the very first step, the computation of a full GVCD for a given data set. Several fast computation schemes with different approximations were developed. The experiments were done on the confocal data set (described in section 3.1) and a data set containing 1742 manually prepared pollen grains from 12 taxa. The second data set contains 2D images with a transmitted light channel and a fluorescence channel that were recorded with a conventional microscope.

The resulting recognition rates with the automatically selected kernel functions were competitive (e. g., about 90% on the confocal data set), but did not reach the best results, that were obtained with the manually optimized kernels.

To further increase the quality of the automatically selected kernels several possibilities were outlined. One of these is to perform a closer analysis of the selected approximations in the computation of the GVCD's. Possibly the selected approximations distorted valuable information already in this step. Another possibility to improve the results may be to use significant larger training data sets. And

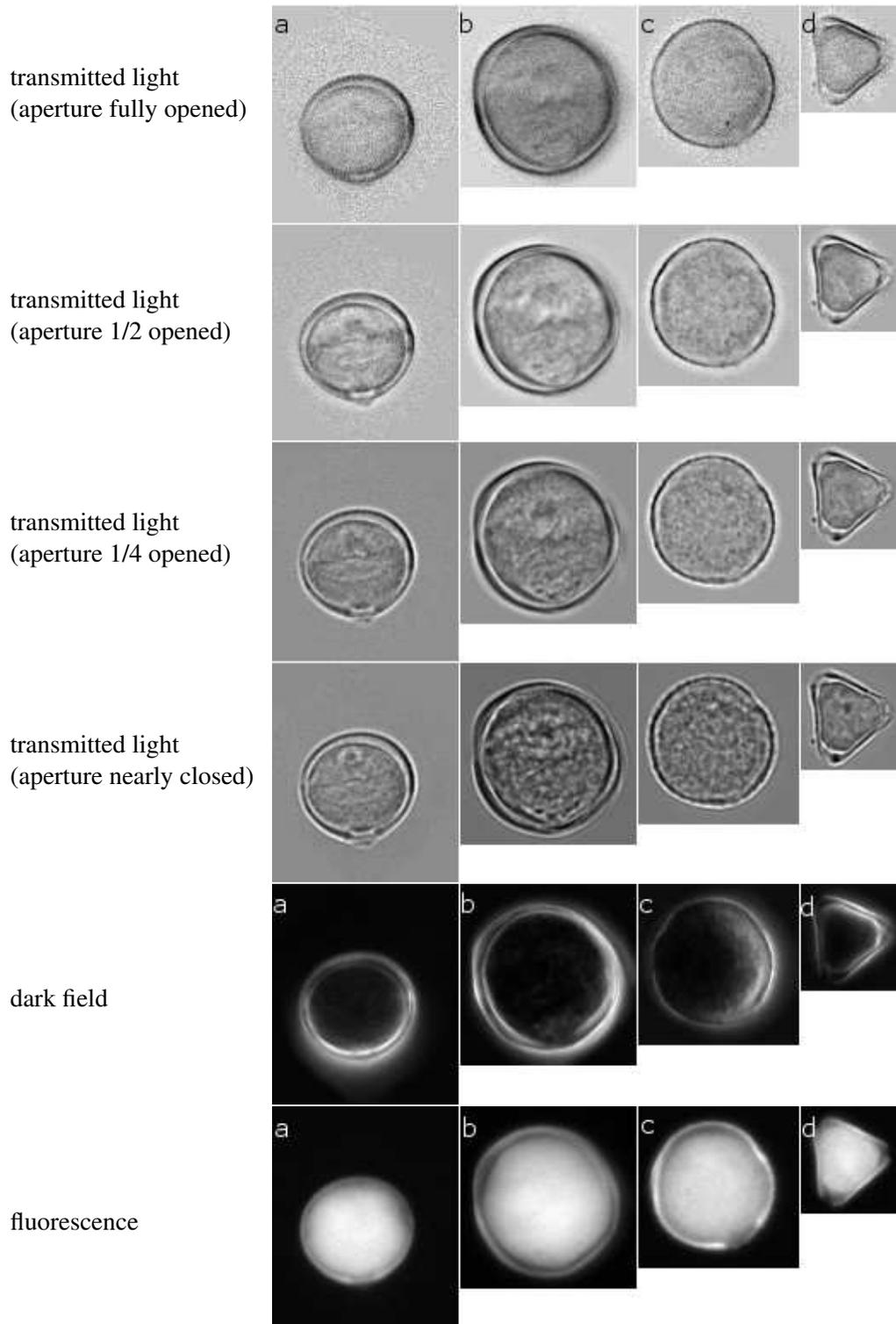


Figure 10.17: Typical pollen grains from honey samples recorded with different microscopic setups (from (Baumgartner, 2005)). a) Roßkastanie, b) Mönchspfeffer, c) Spinnenblume, d) Brautmyrthe

finally a limitation of the permitted kernel function to certain classes of functions might also help to find kernels with better performance. The details of the performed studies are found in [Schmidt \(2004\)](#).

10.5.4 Voxel-wise Invariants for Segmentation and Recognition of Cell Nuclei

The voxel-wise invariants that were introduced in section 6.7 were first applied in collaboration with Janis Fehr to the simultaneous segmentation and recognition of cell nuclei. The used data sets contained confocally recorded tissue samples from the chorioallantoic membrane of chicken embryos. In combination with a voxel-wise classification using support vector machines, the voxel-wise invariants allow a simultaneous segmentation and classification. The training is performed in an intuitive interactive environment:

After the voxel-wise extraction of feature-vectors using two- and three-point kernels, a support-vector machine (SVM) ([Vapnik, 1995](#)) model is trained in an interactive procedure over several iterations: First a small number of training samples (voxels) is manually selected for each class. Second, a SVM model is trained based on the training feature-vectors. In the last step of one iteration, all voxels are classified against the previously trained model. After each iteration new training samples can be added in order to improve segmentation and classification results until the model reaches a "stable" state, e. g. the support-vectors do not change after adding new samples (see figure 10.18 and 10.19 for an example).

The computation time for the segmentation of a volumetric data set with $256 \times 256 \times 100$ (about 6.5 million voxels) was about 2 minutes on a Pentium 4 with 2.8 GHz. Further details on these studies can be found in ([Fehr, 2004](#)), ([Fehr et al., 2005](#)) and ([Ronneberger et al., 2005b](#)).

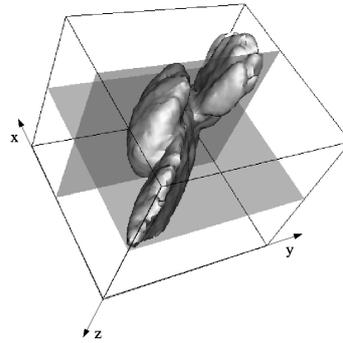


Figure 10.18: 3D isosurface reconstruction of 3 neighboring cells

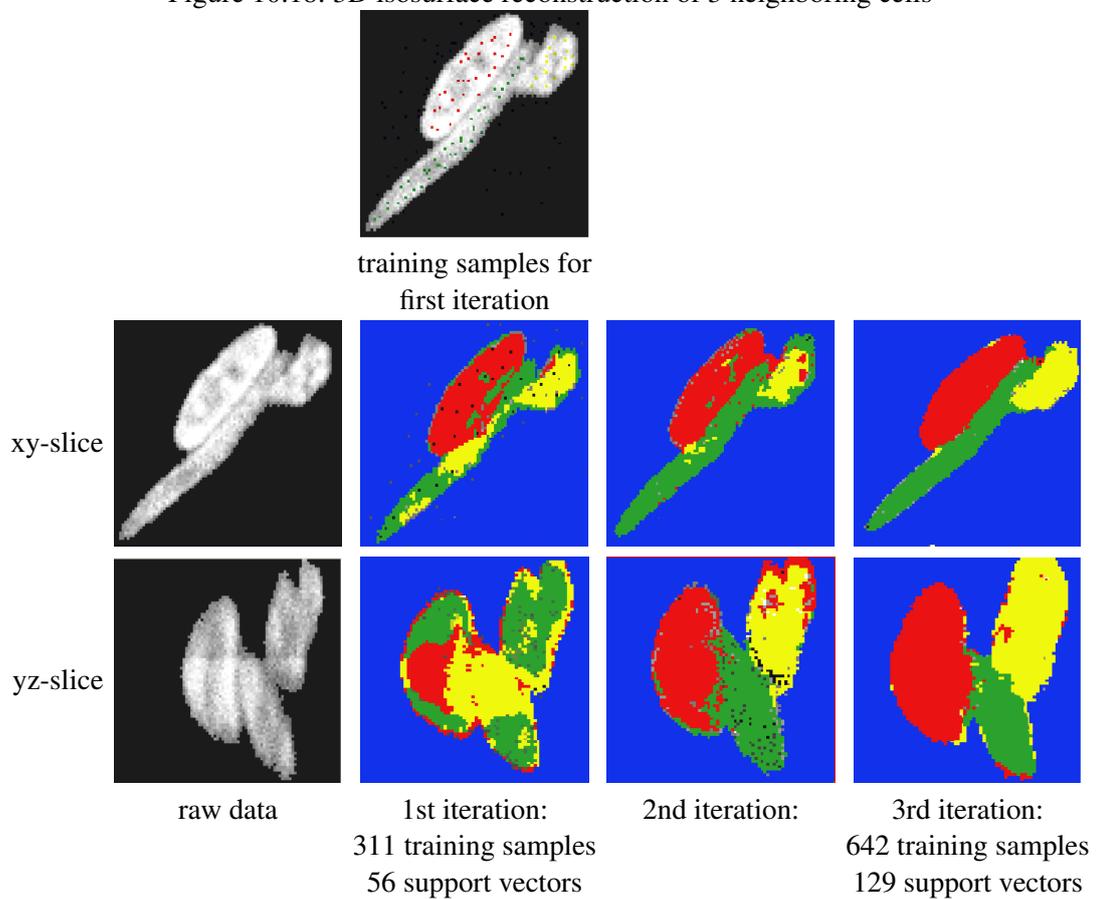


Figure 10.19: The interactive training process - Sections of data as indicated in the 3D reconstruction (figure 10.18) and results after each training step. (from Fehr et al. (2005) and Ronneberger et al. (2005b))

11 Summary, Conclusion and Outlook

11.1 Summary and Conclusions

In this thesis the development of 3D invariants for the recognition of biological structures is described. These invariants are based on the Haar-integration-framework of [Schulz-Mirbach \(1995b\)](#). The invariance properties are reached by an integration over the desired transformation groups, e. g. rotation and translation. Several important aspects of the application of the Haar-integration framework to 3D volumetric data sets of biological structures are described. The most important extension is the introduction of deformation models such that the resulting features are robust to elastic deformations of the structures. Another important aspect is the reached robustness to even non-linear gray-scale transformations, that allow certain variations of the recording parameters between the training and the test objects. The direct computation of these invariants is computationally very expensive. Several new techniques are introduced that allow a fast computation of the invariants by means of the FFT, by the expansion of the integral into spherical harmonics series or by simultaneous computation of multiple invariants based on invertible vectorial kernel functions. Furthermore voxel-wise invariants are introduced for a simultaneous segmentation and recognition of 3D structures. Vectorial invariants are developed for a fast and reliable detection of spherical objects in cluttered environments.

A very challenging application for these invariants is the recognition of pollen grains in air samples. Due to the high number of different pollen taxa and the different types of structures that must be recognized, the pollen can be taken as a good representative for several biological structures that are studied in the biomedical research.

In the given application we use microscopically recorded images to recognize a real-world object. For the correct interpretation of the gray values it is important to understand the different effects within a microscope. In chapter 2 the main four steps, illumination, interaction of the object with the light, transformation of the emitted light, and the recording of the light are explained.

In pollen recognition (as in many other applications) we should differentiate between those results that can be reached within a clean and well-controlled laboratory environment and those results of a real-world routine application. In this thesis two representative data sets for these scenarios are used (see chapter 3): The first one (denoted as “confocal data set” here) is a typical laboratory data set: The pollen were collected directly from the corresponding plants. They were carefully prepared on one slide per taxon and manually recorded as a full 3D volumetric data set with confocal laser scanning microscopy. Due to the high costs of such a system and the time-consuming operation, only a small data set containing 389 pollen grains of 26 different taxa (15 grains per taxon) was recorded. The preparation and recording applied here guarantees a 100% correct labeling of the pollen grains and contains the lowest possible degree of distortion due to optical effects. On the other hand the used samples do not represent all variations within each taxon, such as different genera, species or subspecies, different growth conditions of the plants, etc. Furthermore this data set does not contain deformed, contaminated or agglomerated pollen grains or pollen grains at different levels of degradation and it does not contain the vast amount of other particles that are found in real air samples.

The second data set (denoted as “pollenmonitor data set”) is a typical real-world data set. It was automatically collected, prepared and recorded by the first prototype of the pollenmonitor during the pollen season 2006 in Freiburg and Zürich. It contains about 22,750 pollen grains from 33 taxa together with about 170,000 other spherical airborne particles.

In a first feasibility study invariants based on general-purpose 2-point-kernel functions were applied

to the confocal data set. A fast and exact computation of these invariants was reached by the reformulation of the problem, such that it can be computed by means of the fast Fourier transform (see section 7.4). Already with a very low number of only 14 invariants, a recognition rate of 91.9% was reached on this data set. This result could be further improved to 95.6% by a normalization of the same 14 features and an extensive grid search for the best parameters of the applied support vector machines.

Compared to the results of other groups on such typical laboratory data sets (see section 10.2.4) we already outperformed all other approaches. The only other results to which we cannot provide a fair comparison here are the results of Jones (2000). The used data set in this study was much larger than ours (16,220 pollen grains from 80 taxa) such that we cannot conclude whether the lower recognition rate of 70.7% was caused by a higher difficulty on this data set or a lower performance of the selected approaches.

11.1.1 Conclusions and Intermediate Results that led to the Final MiSP Invariants

The conclusions and intermediate results that led to the final form of the MiSP¹ invariants (see chapter 9) are better described chronologically.

Encouraged by the promising results on the confocal data set, we applied the general-purpose 2-point-kernel functions to several other data sets with increasing similarity to the final real-world data set. E. g., we obtained satisfactory results on real air samples from January, February and March containing the typical three pollen taxa (*Alnus*^(A.2), *Corylus*^(A.5) and *Taxus*^(D.12)) that are found in this season and the typical seasonal dust particles. One month later, when the first *Betula*^(A.4) pollen appeared and an increasing number of other particles of biological origin, the system was not able to distinguish between *Betula*^(A.4) and *Alnus*^(A.2) and the number of false positives increased dramatically.

The most obvious difference of these data compared to the confocal data set was the much lower raw data quality. To retrieve similar data with a conventional fluorescence microscope, deconvolution (see section 10.5.1) and structured light techniques (see section 2.3.3) were tested. Even though they could visibly improve the raw data quality, the reached quality was not high enough to model a real-world 3D rotation of the object by a 3D rotation of the recorded volumetric data set. Only on a honey pollen data set recorded from pure pollen samples, the structured light techniques could slightly improve the results (see section 10.5.2).

Another assumption for the reason of the poor results on the real-world data was that the richness of the selected 2-point-kernel functions is not high enough. An analysis of the amount of information that can be sensed with 2-point kernels (in section 5.2.1) yielded interesting new insights. This analysis was done by splitting the computation of the invariant into the extraction of the gray value co-occurrence distribution (GVCD) and the postponed application of the kernel function. This allows an intuitive understanding of the type of information that is obtained by different kernel functions. Furthermore the presentation of the raw data by means of the GVCD can be used for a rough estimation of the expectable robustness of certain kernel functions on a certain data set. Theoretically it should be possible to find the optimally suited kernel functions for a given problem by the analysis of the GVCD's of the training samples. A first feasibility study on the realization of such an approach returned competitive results (see section 10.5.3), but the automatically selected kernels were not able to outperform the best manually selected kernels.

While further experiments with modified and extended kernel functions did not significantly improve the results, a significant improvement could be reached by an adaption of the transformation group. The usage of the center of the object (that could be easily extracted of the mask computed in the prior segmentation step) allows to replace the afore required translation invariance with a normalization step. Furthermore we draw the conclusions from the deconvolution and structured light experiments.

¹MiSP = Microscopical Spherical Particles

A simulation of a 3D rotation of the real-world object by a rotation of a data set recorded with conventional microscopy was not possible with sufficient quality. The loss of information in z-direction is too high, and the characteristics in z-direction compared to those of the xy-direction are too different (see section 9.1.2). As a consequence we replaced the 3D rotation by a planar rotation around the z-axis and enlarged the training data set, such that enough pollen grains at different orientations relative to the optical axis are available.

The left over transformation group was the group of planar rotations. The resulting restriction of the sensing area of each invariant to a circular region within the object significantly raised the specificity of the extracted features. The recognition rates went up and a lot of dust particles were now correctly classified as non-pollen.

But with these improved invariants still many pollen taxa could not be reliably differentiated. An analysis of the importance of the individual invariants for the final classification gave the crucial hint for a further improvement. A feature selection by the SIMBA method (Gilad-Bachrach et al., 2004) showed that the most discriminative features (e.g. for *Corylus*^(A.5) pollen) were extracted far away from the sharpest layer. In this layer only three blurred bright spots were visible. These spots are produced by the light refracted at the three pores of the pollen grain. So it was not surprising that it is possible to reasonably differentiate between pollen and non-pollen with these invariants, but that the fine differences between morphologically similar pollen taxa can not be resolved.

A closer analysis of the features extracted from the sharpest layer was performed. This layer contains a large fraction of the discriminative information. But due to the circular sensing regions of the features and the significant inner-class variations of the pollen shapes, corresponding structures of two pollen grains were in most cases located in different regions. Obviously, for the recognition of biological objects not only a rotation invariance but also an invariance (or at least a robustness) to deformations must be reached: A radial global deformation (see section 9.1.3) was added to the transformation group and the Haar integral was implicitly limited to finite results by a new approach that uses synthetic channels computed from the segmentation mask (section 6.3.4 and section 9.2.2). With this approach the sensing area of each invariant is restricted to “shells” with a certain distance to the outer border (see section 9.5).

This integration of a deformation model into the HI framework was the most important step towards a reliable recognition of the pollen grains. With this technique the results became significantly better. The number of false positives decreased and the system was able to differentiate between pollen taxa that were not differentiable before.

The only remaining severe drawback was the strong dependency of the features on the selected recording parameters. A slight change in the optical setup before the measurement period in Zürich caused a nearly complete failure of the recognition on the newly recorded pollen grains. It was possible to reach satisfactory results by a retraining of the system with some of the newly recorded pollen grains. An alternative would have been of course a calibration of the optics to some well defined standard parameters.

But for a reliable operation of such a system we would like to have features that are not so sensitive to these intensity and contrast variations in the images. Especially in situations where such variations are caused by contamination within the sample, the variations can not be compensated by a calibration of the optical system.

This problem was solved by replacing the raw gray values in the kernel function by an appropriate combination of gradient magnitude and gradient direction (see section 9.2.1). This extension provided the desired robustness to gray value transformations (see final results on Zürich data set in section 10.4.5). Furthermore it allowed to integrate another deformation model, the “local deformations”, described in section 9.1.4. Without this prior projection of the gray values (according to their gradient direction) into a higher dimensional space the probability-weighted integration over the group of local deformations would just cause a low-pass-filtering of the image (see section 9.5).

11.1.2 Segmentation

In the first feasibility study on the confocal data set no segmentation was needed. By the use of confocal microscopy on pure pollen samples, the pollen grains are already well separated in a data set with black background. The invariants based on general-purpose 2-point kernels and Euclidean transformation do not rely on any further preconditions than a single object on black background.

For real-world samples a segmentation seems to be absolutely necessary. For the first tests we applied standard morphological techniques (see section 8.1). For certain air samples (clear background, not too many particles in the image, no agglomerations of pollen, no occlusions, etc.) these techniques provided satisfactory results. For most real-world samples they performed worse. Some of the objects were only partially segmented while at the same time multiple other objects were segmented as one part.

The first idea to a solution of this problem was to use voxel-wise invariants combined with a voxel-wise classification (section 6.7). Due to the limited discrimination power of centered two-point-kernels for voxel-wise invariants (section 6.7.1), centered three-point-kernels were used. For this type of kernels the fast exact computation via a single convolution (section 7.4) could not be applied. A new class of fast computation algorithms was developed. In these algorithms the kernel function is approximated by an expansion into a Fourier series or a spherical harmonics series (section 7.5ff.). In contrast to alternative approximative computations of a Haar integral (Siggelkow and Schael, 1999) the results obtained with this technique are perfectly invariant to the given transformation group independently of the selected truncation of the series.

With the developed voxel-wise invariants and the voxel-wise classification we obtained encouraging results on confocally recorded cell nuclei (section 10.5.4). But even with several optimizations, the needed processing time was about 2 minutes for a dataset with about 6.5 mega voxels. The available processing time in the online operation of the pollenmonitor is only a few seconds for a data set with about 100 mega voxels. So the proposed method could not be applied.

The solution of the segmentation problem is based on a new class of kernel functions with a very fast computation algorithm (section 6.8 and 7.10). These “sparse vectorial kernel functions” operate on the gradient image and can be used to detect spherical structures in very cluttered environments. This technique was denoted as MiSP detector (section 8.2). The only drawback of this detector for the application to pollen recognition is its limitation to spherical pollen taxa. In Germany and Switzerland all highly, moderately and lowly allergenic pollen are spherically shaped. So for the developed pollenmonitor this limitation has no effects. In other countries or for different applications it might be necessary to extend this detector. After the detection the precise border is obtained by a combination of a modified Canny-Edge Detector, a model-based weighting of the obtained edges, a gradient vector flow and finally a snake approach (see section 8.3). This combination of modified standard algorithms is able to find the correct object borders even under very unfavorable conditions. A minor drawback of this technique is the high number of parameters of these standard algorithms, that had to be properly adapted to the data set. The obtained segmentation borders are displayed as overlay on all pollen grains in the appendices A.1 – D.17.

11.1.3 Application of MiSP Invariants to Pollen Recognition

The final MiSP invariants were applied to the confocal data set as well as to the pollenmonitor data set.

Confocal Data Set

The recognition rate on the confocal data set was very high. In a leave-one-out validation with a simple 1 Nearest Neighbor classifier only 3 of the 389 pollen grains were misclassified (99.2% recognition rate). With a support vector machine only 2 incorrectly classified pollen grains were obtained (99.5%

recognition rate) (see section 10.2.3). In the second case the misclassified pollen were confused with another genus from the same family.

This result outperforms all the results of other groups on pure pollen samples and all our earlier results on the same data set (section 10.2.4). The data was produced in a similar way as in most other studies (pure pollen samples, only one slide per taxon, limited number of plants per taxon). So we can assume that the results are somehow comparable. On the other hand this procedure may lead to positively biased results (see section 1.2.1 and 3.1.4). This criticism applies to our study on this confocal data set as well as to the other studies.

Pollenmonitor Data Set

For pure pollen samples the total recognition rate provides a meaningful statistical measure for the comparison of different studies. In real air samples this total recognition rate has only a limited validity. In such data sets a very high number of non-pollen particles is present, e. g., about 170,000 spherical dust particles compared to 22,750 pollen grains in the pollenmonitor data set. In a thought experiment with a classifier that classifies every particle to the dust class already an impressive recognition rate of 88% can be reached. By decreasing the threshold in the segmentation step we can easily increase the number of found dust particles by a nearly arbitrary factor. This would lead to even higher recognition rates in this experiment.

So we must find statistical measures that are independent of the number of dust particles. Such meaningful statistical measures to compare different pollen recognition systems were identified in section 10.1 by an analysis of the full error of the system for typical particle concentrations in the air. In the pollenmonitor data set the fraction of the most important allergenic pollen taxa (*Poaceae*^(A.6)) to other particles is 1:120, and the fraction of the rarest allergenic pollen taxa (*Artemisia*^(A.1)) is 1:3000. It turned out that the avoidance of false-positives is the most important task. Statistical measures that consider the number of false-positives but are not biased by the number of segmented dust particles are the precision and the introduced “false-positive-concentration”, which relates the number of false positives to the sampled air volume.

A reduction of the number of false-positives is reached by using two techniques. The first is the manual selection of high-quality pollen (not agglomerated, not contaminated, well segmented, etc.) for training. The addition of contaminated pollen to the training data set would increase the risk that an unknown contaminated pollen will be assigned to that pollen with a similar contamination pattern instead of that pollen with the same taxon.

The second technique to avoid false-positives is a restrictive rejection scheme. By the use of the one-vs-rest multi-class approach we can straightforwardly setup a rejection scheme. Only those classification results with exactly one positive decision value are accepted, the others are rejected.

With these two techniques we obtained on a test data set from Freiburg and Zürich with 33 pollen taxa a mean precision of 96.7% (at a recall of 84.3%) for all pollen taxa. The results for the five highly allergenic pollen taxa in this data set (that are monitored for the pollen forecasts) are even better: 98.5% precision at a recall of 86.5%. For certain applications an optimization to a higher recall might be required. This can be reached by a different recombination strategy of the decision values: The object is always assigned to that class with the highest decision value, independently whether it is positive or negative, or whether other positive decision values exist. Using this strategy on the full test data set (Freiburg and Zürich) the obtained precision is decreased from 96.7% to 91.5%, but the recall increases from 84.3% to 91.6%.

These results are better than all other published results on pollen recognition in real air samples (see section 10.4.6). The study with the most similar goals is published in Ranzato et al. (2007). There on 8 pollen taxa a precision of 30% at a recall of 64.9% was reached (using a 10-fold cross validation).

For the application of the pollen recognition system in an online measurement network, the final goal is the correct prediction of the expected allergic stress at the current pollen concentration. These

results are shown for computed daily mean concentrations in section 10.4.5. Here we can clearly see that with the selected rejection scheme in most cases a better prediction of the allergic stress is possible. Especially outside the flowering period of the particular pollen taxa the number of false alarms is significantly lower.

Even though these results are very encouraging, we cannot be sure that these results can be taken as a good prediction for the final recognition rate of an online pollenmonitor in a routine environment. On the one hand this data set may still not represent all possible variations of the pollen grains, because it was recorded only within two locations in one year with the identical machine. This may have led to positively biased results. On the other hand it contains several shortcomings, that can be avoided in future by an improved hardware. These variations (like the low quality of the fluorescence recordings, vibrations during stack recording, non-uniformity of the illumination, misadjusted optics, etc.) are briefly discussed in section 3.2.6. These variations very likely have led to negatively biased results. Furthermore there are still some pollen grains in the data base that were incorrectly labeled by the pollen experts. The author is not a pollen expert but for some of the false positives (e. g. the 4th and 5th of the false positives on the *Alnus*^(A.2)-page) it seems to be obvious that the expert-labeling “No Pollen” is questionable.

11.1.4 High Dimension of the Feature Vectors

Compared to the dimension of feature vectors in a typical pattern recognition application the number of 87,296 MiSP invariants on the pollenmonitor data set may appear extremely high. The most obvious justification of this high number is that these features correspond to the entries in a 4-dimensional feature array with the dimensions $31 \times 11 \times 16 \times 16$. The first two dimensions correspond to the localization of the “shell” relative to the border and the sharpest layer of the object. The other two dimensions correspond to the relative gradient direction and the angle of the kernel function. So a more reasonable interpretation might be a decomposition of the volumetric data into 341 shells and the subsequent extraction of 256 features of each of these shells.

We can see from the depiction of the feature array in figure 9.12 on page 135 that the nonzero-entries in this feature array are quite sparsely distributed. So a feature selection approach applied to the training data set is expected to identify several shells, gradient directions or kernel angles (or combinations of them) which are not needed for the separation of the objects in the training data set. Consequently these shells could be left out from the feature extraction, resulting in a lower dimensional feature vector.

This should be a reasonable strategy if we can ensure that all possible types of particles are incorporated in the training data set. But in the present application, the online pollen monitoring, we have no closed well-known laboratory environment, but collect particles directly from the outdoor environment. So every day new types of particles and new variations of the known particles may appear (and actually did appear within 2006). Therefore the decision to leave out certain regions of the object which are not needed for the discrimination of the training objects, may lead to avoidable false positives. E. g., an unknown particle might have the crucial structures for its rejection exactly within the left out regions.

So we think that it is better to retain all the features and instead to optimize the classifier such that it can deal with such high-dimensional feature vectors. In fact we have developed an optimized support vector machine, that is able to manage such high-dimensional feature vectors. The respective library, libsvmml (“support vector machine template library”), is available as open source from the authors home page <http://lmb.informatik.uni-freiburg.de/people/ronneber/>.

A further optimization which makes the full dimensionality of the feature vector nearly negligible has already been applied to intermediate studies with nearest neighbor searches in this 87,000 dimensional space. For this technique, the features in the feature vector are arranged according to their variance (the features with the highest variance first). The computation of the distance between a test object and a reference object is performed iteratively “feature-by-feature”. If a certain distance threshold is exceeded, this computation is stopped and the distance is set to infinity. This threshold can be

specified in advance or can (even better) be adapted according to the current state of the classification process. E. g., for a 10-nearest-neighbor classifier the distance of the (up to now) 10th neighbor specifies the threshold for the distance computations of the remaining reference objects.

This technique was applied to the pollenmonitor data set, and it has turned out that on average only about 10% of the features need to be taken into account to compute the distances. This approach can also be integrated into certain kernel functions (e. g. the RBF kernel) in a support vector machine. Experiments have not yet been performed in this field, but we expect similar positive effects as in the nearest neighbor classification.

In other words, we suggest to postpone the “feature selection” to the time, when the relevance of a particular feature can be safely judged, i. e. when the new object to be classified has already been recorded and is compared to the existing objects.

11.1.5 Final Conclusions

In this thesis the proposed algorithms were not only validated on small laboratory data sets, but on a very large real-world data set. Pollen recognition in a online environment is an extremely challenging application due to

- the large number of different pollen taxa
- the large intra-class variation within one pollen taxon
- the subtle inter-class differences
- the need for 3D techniques for a reliable differentiation
- the very high number of other particles on an air sample, that must be correctly rejected
- the recognition of structures at the same scale as the wavelength of light
- the timing constraints that must be fulfilled within an online system
- the “open” system. We can never know in advance what kind of other particles will be found on a new sample.

Even though we used a much larger and more difficult data set than all other published approaches to pollen recognition, the obtained precision and recall was still significantly better than the other results on smaller and easier data sets.

This may be taken as a hint that with the development of the proposed MiSP invariants a significant progress in the field of biological object recognition has been done.

11.2 Outlook

The challenges to be tackled next in pollen recognition become clear by a closer analysis of the rejected, the incorrectly classified and the false-positive pollen grains listed in the appendices [A.1](#) – [D.17](#). It is again emphasized that all the displayed pollen grains have automatically been selected from the whole data set without any manual intervention.

In most cases the incorrectly classified pollen grains are partially occluded, significantly contaminated or corrupted. In many cases this has led to an incorrect segmentation of the object and, as a consequence, to a rejection or an incorrect classification. In contrast to our automatic techniques a human is able to identify such distortions in the images and to concentrate on the undistorted parts of the object for the classification. A similar approach could be integrated into the MiSP invariants. At the present state the computation of these invariants can be interpreted as a two stage process. In the

first stage local features are extracted at each voxel, namely the gradient magnitude and the gradient direction. In the second stage these local feature are combined to global features for the description of the whole object. To meet the mentioned requirements we could extend this framework in several ways:

One possibility is to replace the local feature extraction by a more sophisticated approach. E. g., we could exchange the gradient with an approach that is able to emphasize certain pollen-like structure, and to suppress contamination-like structures.

Another possibility is to extend this two stage approach to a multi stage approach. E. g., the local features that are extracted in the first stage can be combined to some regional features in a second step. This regional features could already describe (or even recognize) certain substructures within the object (e. g. porates, certain textures etc.). In a final step this regional features are then recombined to the desired global features.

A third possibility is to leave the classification as it is but to optimize it to a high recall (with a high number of false-positives, though). An additional validation step is responsible then to reject the incorrectly classified objects. This validation has to operate only on a small fraction of the whole data set, and the potential pollen taxon has been determined already in the prior step. Then very expensive techniques up to partially elastic registrations can be used.

For applications other than the recognition of allergenic pollen a detection and segmentation of non-spherical pollen (e. g. the *Pinus*^(D.10) and *Picea*^(D.9)) is desirable. Moreover applications that need to recognize completely different objects are within the reach of the developed techniques. First promising attempts of this kind (2D only) have already been made by [Reisert and Burkhardt \(2007\)](#). In ([Reisert, 2007](#)) these techniques were applied to the 2D detection of fungal spores in real air samples.

From the application point of view the most important drawback of our proposed approaches is the necessity to train the classifier on real air samples. A closer look at table 3.17 on page 65 shows that even in the very large pollenmonitor data set several pollen taxa are so rare that a meaningful training is not possible. So as a further challenge the pattern recognition has to become even more robust to the recording parameters. Such a robustness should allow to train the system on manually prepared samples. Alternatively new preparation methods could be searched by which the manually collected pollen are prepared such that they appear as similar as possible to the pollen within the real air samples. And of course a combination of these two approaches is possible, as well.

An further interesting question is to explain why our approaches were able to outperform all the other hitherto developed approaches on pollen recognition. Some explanations are given already in the previous section. But as we have not only optimized the pattern recognition but also the whole preparation and microscopic recording process (including the usage of full 3D datasets), a direct comparison of our approaches with the other approaches on the identical data set might provide new insights. The authors of ([Ranzato et al., 2007](#)) already kindly provided their pollen data set for such a comparative study. Moreover, the chances are not bad that also the very large data set (16,220 pollen grains from 80 taxa) of [Jones \(2000\)](#) will be made available for this study.

Finally, we think that the new findings on those techniques that work in a real-world application compared to those that perform well only on a limited small data set, will be a valuable foundation for many further developments in the field of the recognition of biological structures in 3D volumetric data sets.

A Highly Allergenic Pollen Taxa

In this and the following appendices a comprehensive overview over all considered pollen taxa is given. Additionally details on the segmentation and classification on the pollenmonitor data set are depicted.

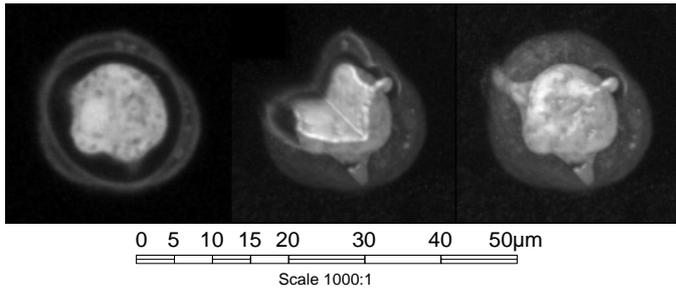
First for each pollen taxon the central layer, a cropped volume rendering and a full volume rendering from one representative of the confocal data set are displayed at scale 1000:1.

This headline is followed by all objects of this taxon that are contained in the confocal data set, displayed as shaded volume rendering at scale 500:1.

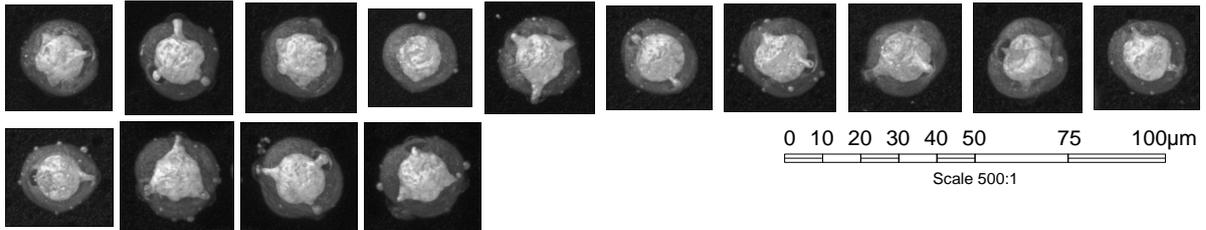
Below this the results and examples of the pollen grains of the pollenmonitor data set are shown at scale 500:1. This table contains a representative subset of the true-positives, the rejected, the false-negatives, and the unknown-positives¹. The pollen grains in each group were sorted by the decision value of the support vector machine, which can be interpreted as a kind of certainty for the classification. From this sorted list every n 'th pollen grain was taken (starting with the “best” down to the “worst”), where n was computed such that the number of retrieved images fits on a double page. The decision values are provided below each pollen grain, and for each incorrectly classified pollen grain the true and the predicted class is listed. (In the electronic version of this document each pollen name is hyper-linked to its page in the appendix). For each pollen grain only the sharpest layer from the 3D stack is displayed together with the found segmentation border. The reduction to only the sharpest layer hides important information in several cases, especially for the incorrectly classified pollen grains, where the “reasons” for the incorrect classification are often located in the other layers. But this 3D information was considered to be not so important as the display of many different individuals of this pollen taxon on the same space.

¹the unknown-positives are those pollen grains which were labeled as “indeterm.” by the pollen experts, which means that they were not able to unequivocally classify them.

A.1 Artemisia (Beifuß)



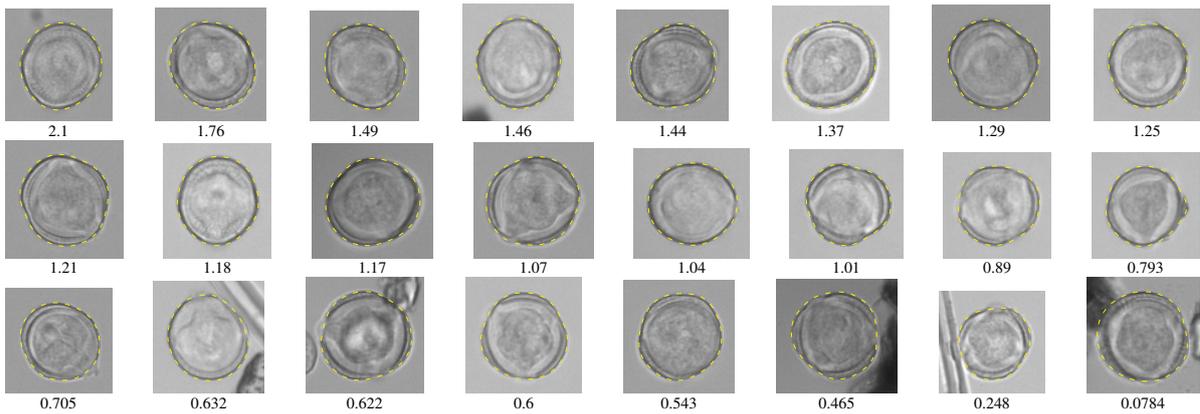
Confocal data set



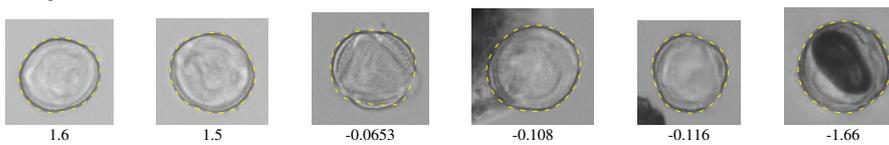
30 *Artemisia* in pollenmonitor test data set (Freiburg + Zürich)

(24 in training data set)

24 correctly recognized as *Artemisia* (80%):



6 rejected (20%):

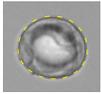


0 classified as other objects (0%):

Other objects that were confused with *Artemisia*

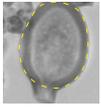
(relative to 28 predicted *Artemisia*)

1 false positives (3.57%):



Urtica (D.15)
→ *Artemisia*
(0.0441)

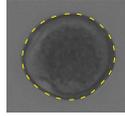
3 unknown positives (10.7%):



Indeterm.
→ *Artemisia*
(0.315)

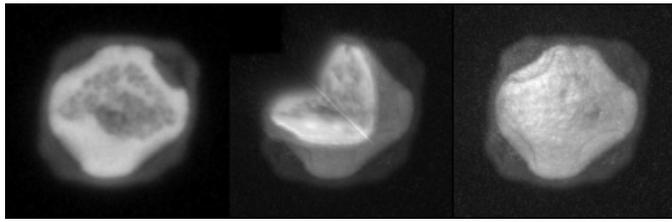


Indeterm.
→ *Artemisia*
(0.122)



Indeterm.
→ *Artemisia*
(0.0416)

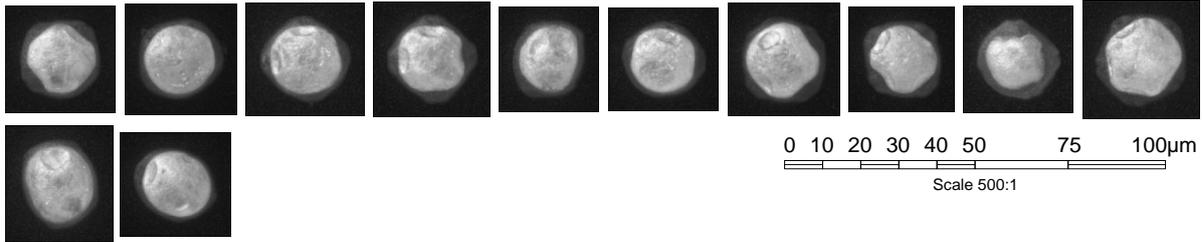
A.2 Alnus (Erle)



0 5 10 15 20 30 40 50µm

Scale 1000:1

Confocal data set



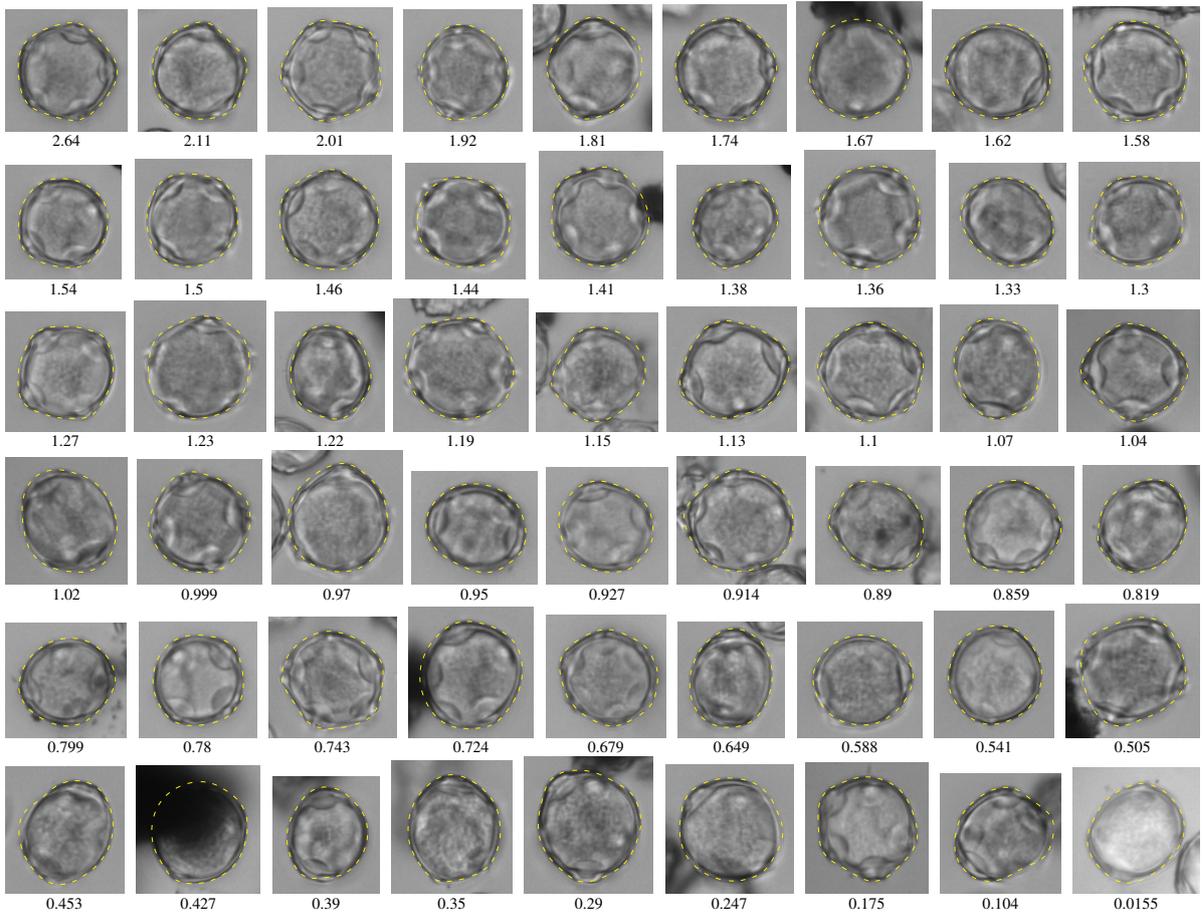
0 10 20 30 40 50 75 100µm

Scale 500:1

890 *Alnus* in pollenmonitor test data set (Freiburg + Zürich)

(145 in training data set)

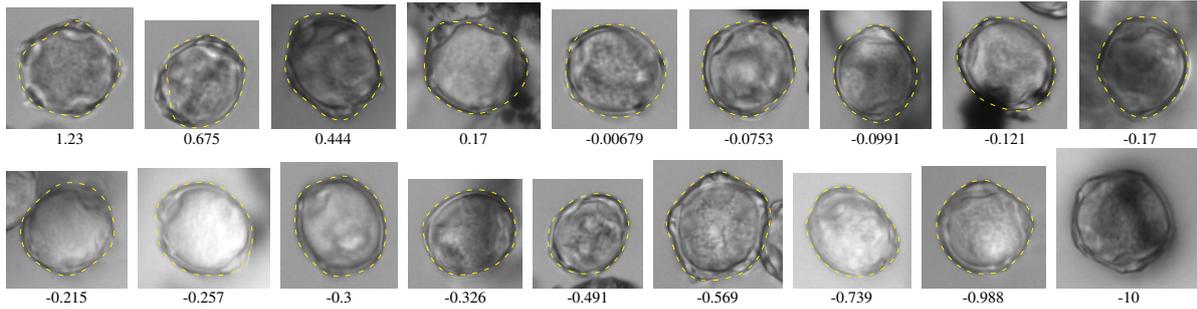
751 correctly recognized as *Alnus* (84.4%):



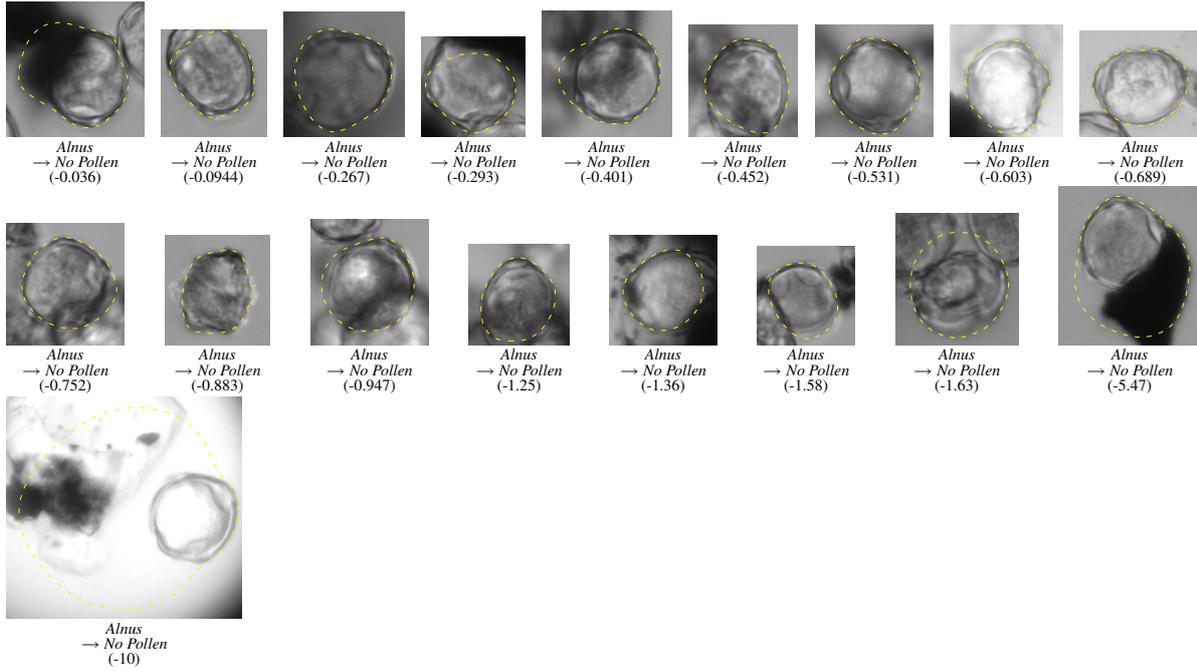
0 10 20 30 40 50 75 100µm

Scale 500:1

96 rejected (10.8%):



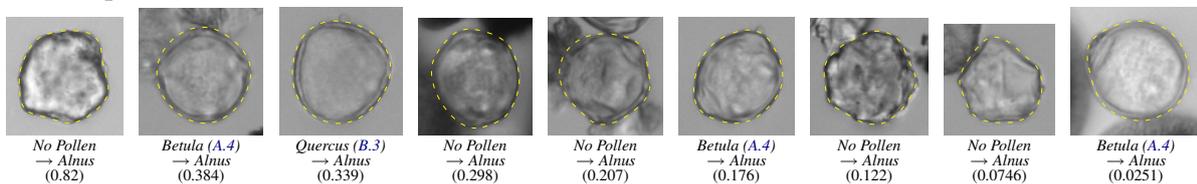
43 classified as other objects (4.83%):



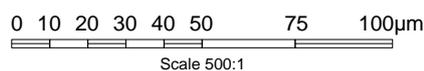
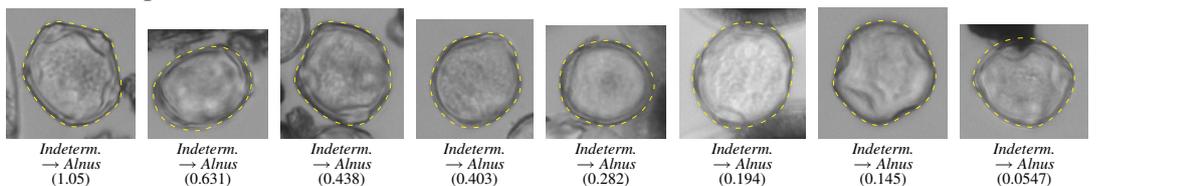
Other objects that were confused with *Alnus*

(relative to 774 predicted *Alnus*)

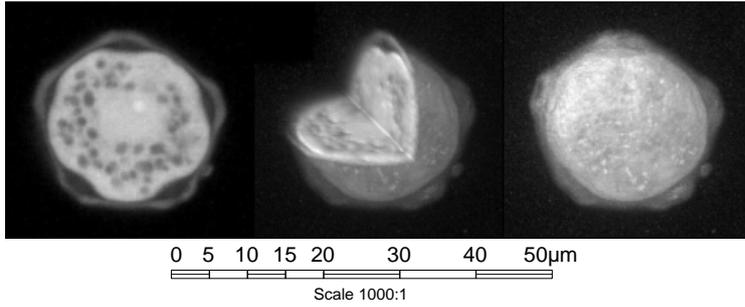
15 false positives (1.94%):



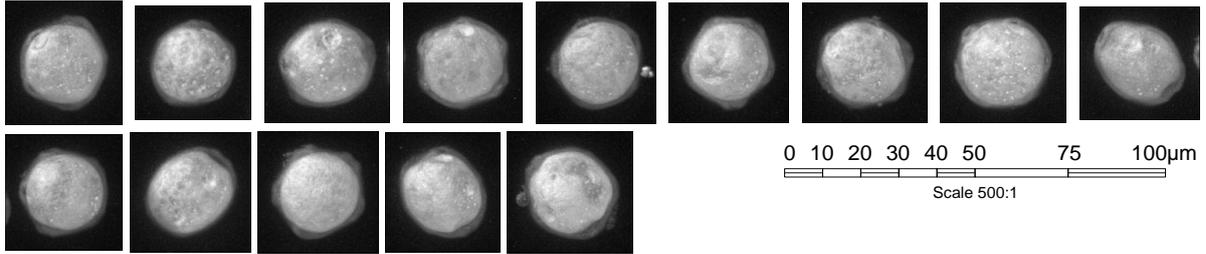
8 unknown positives (1.03%):

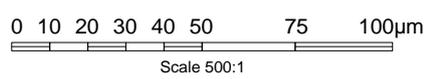


A.3 *Alnus viridis* (Alpen-/ Grün-Erle)

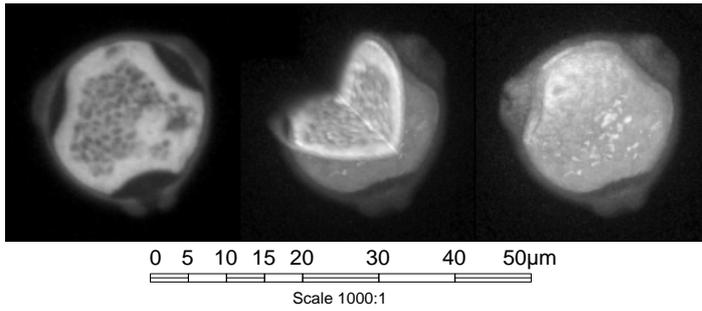


Confocal data set

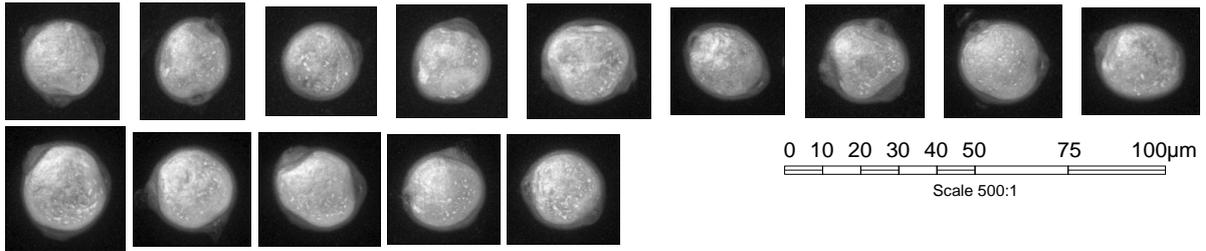




A.4 Betula (Birke)



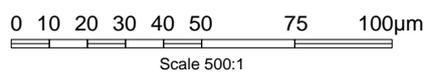
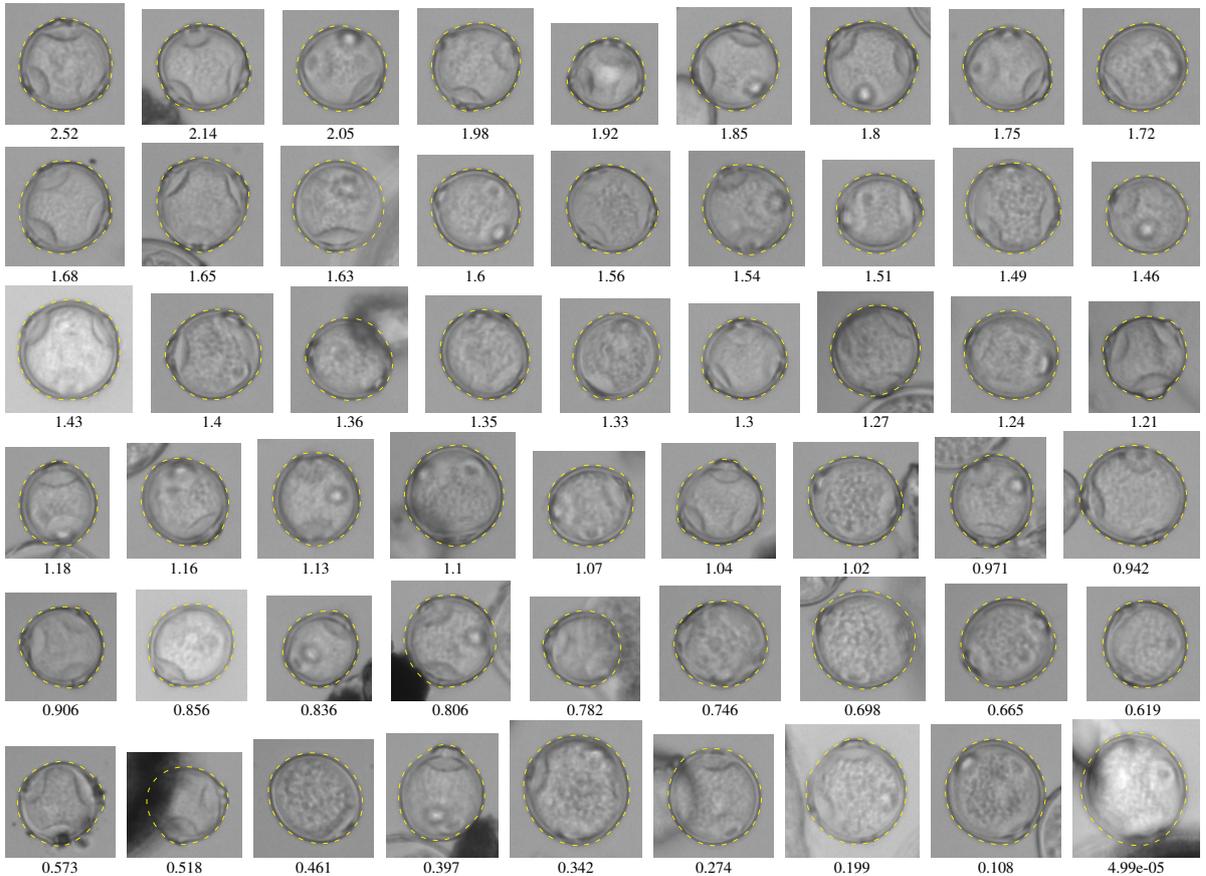
Confocal data set



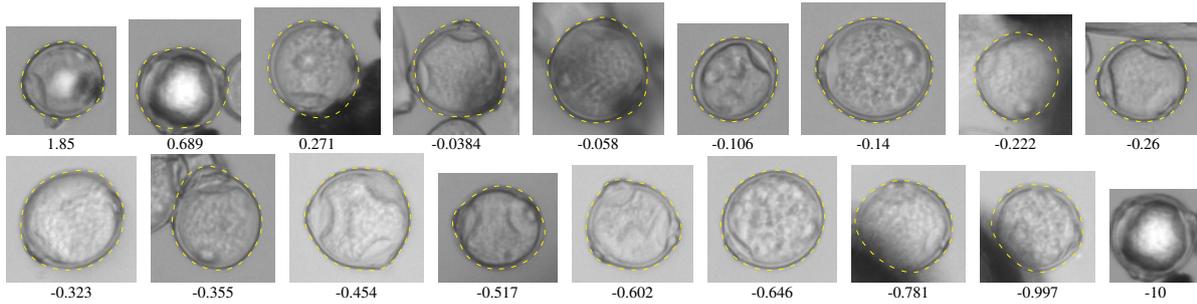
1035 *Betula* in pollenmonitor test data set (Freiburg + Zürich)

(231 in training data set)

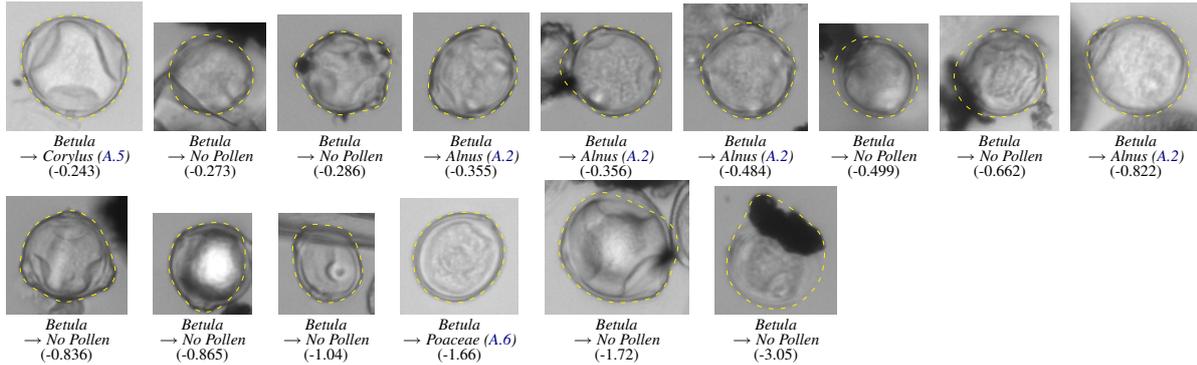
933 correctly recognized as *Betula* (90.1%):



87 rejected (8.41%):



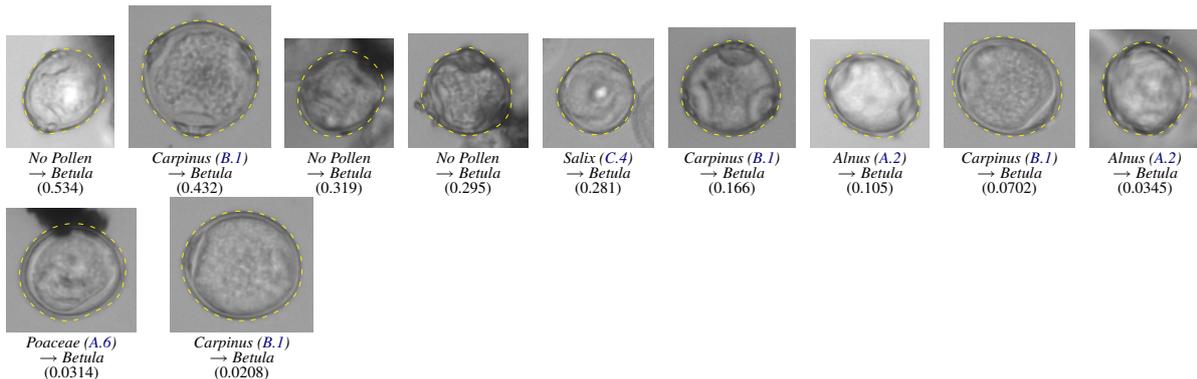
15 classified as other objects (1.45%):



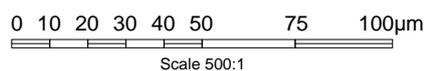
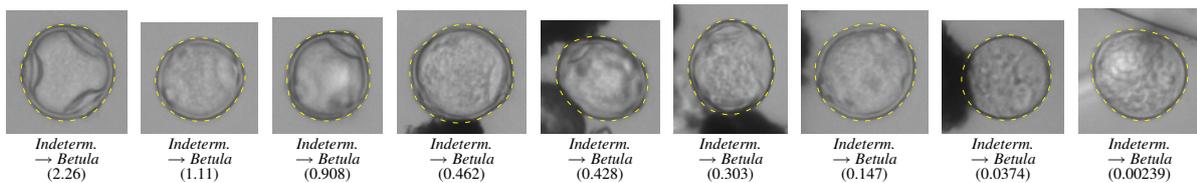
Other objects that were confused with *Betula*

(relative to 967 predicted *Betula*)

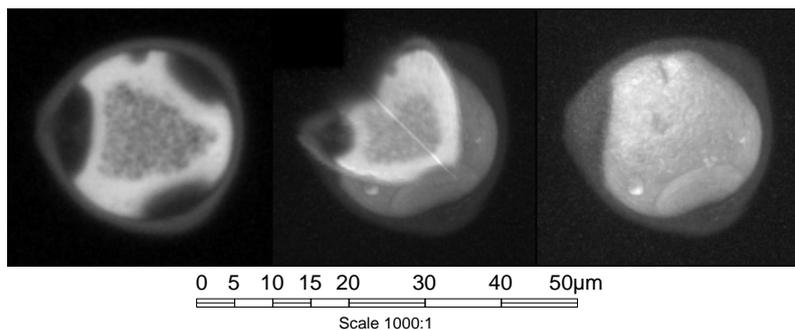
11 false positives (1.14%):



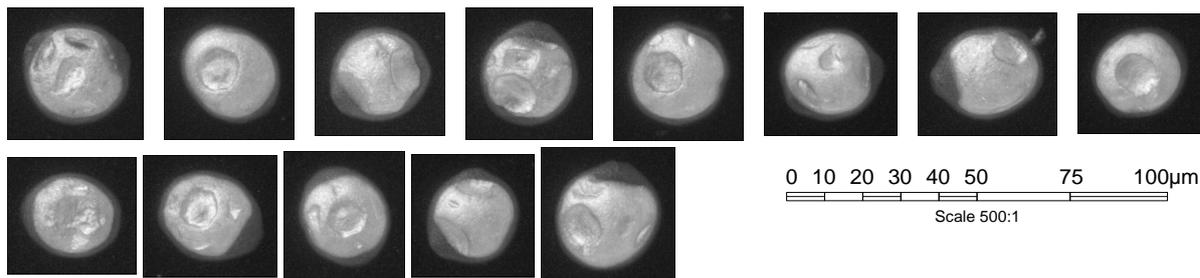
23 unknown positives (2.38%):



A.5 Corylus (Hasel)



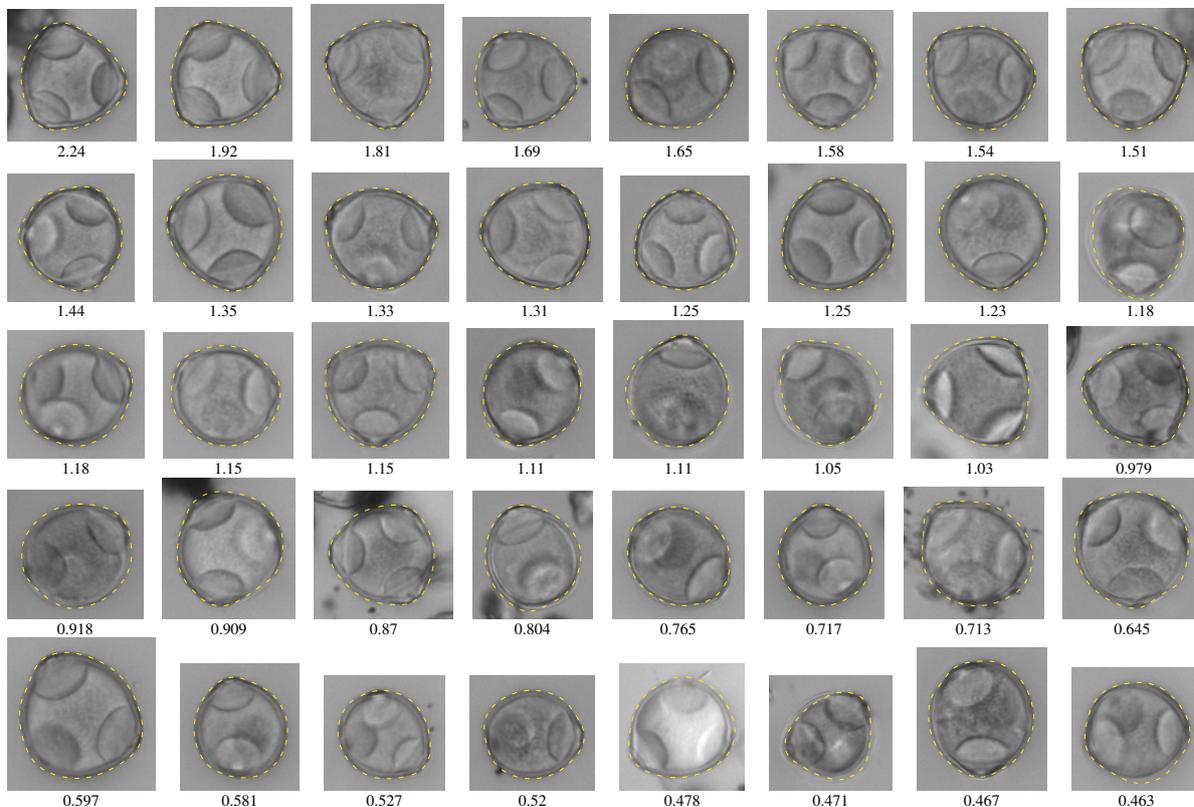
Confocal data set

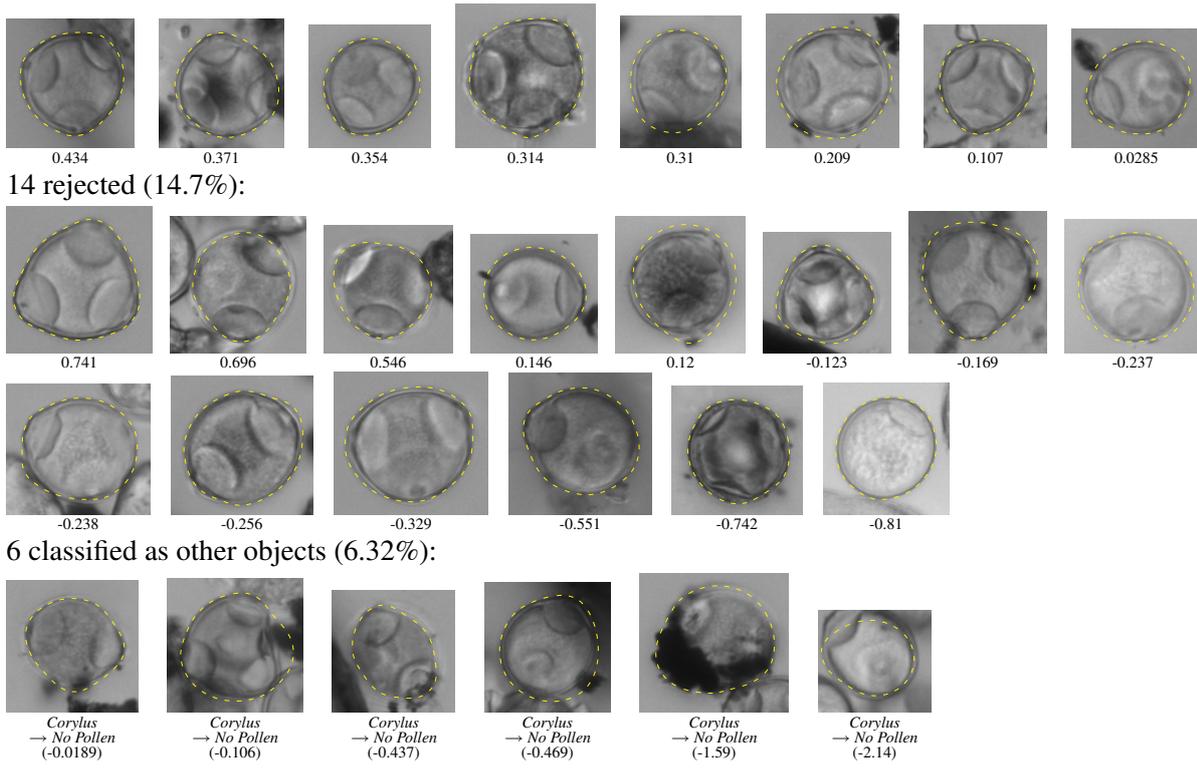


95 Corylus in pollenmonitor test data set (Freiburg + Zürich)

(44 in training data set)

75 correctly recognized as *Corylus* (78.9%):

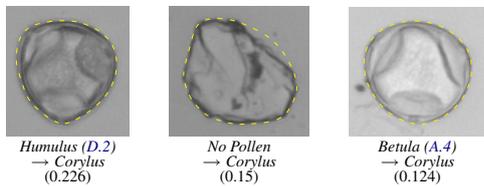




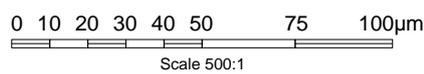
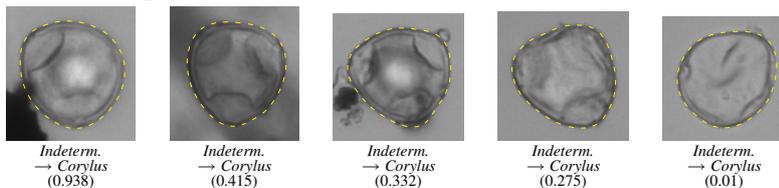
Other objects that were confused with *Corylus*

(relative to 83 predicted *Corylus*)

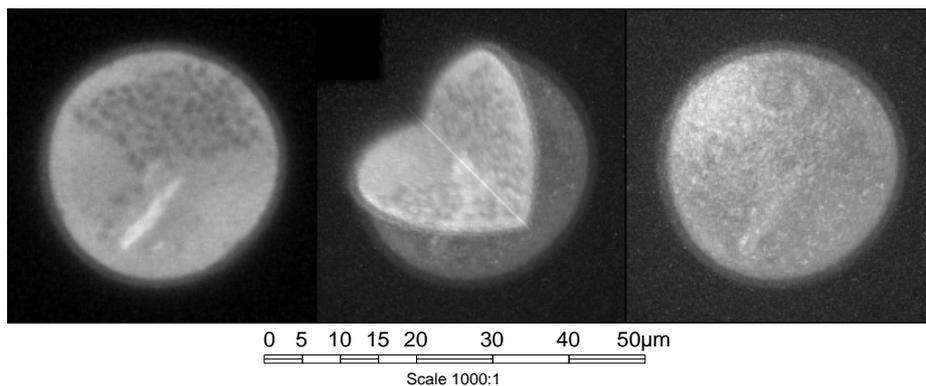
3 false positives (3.61%):



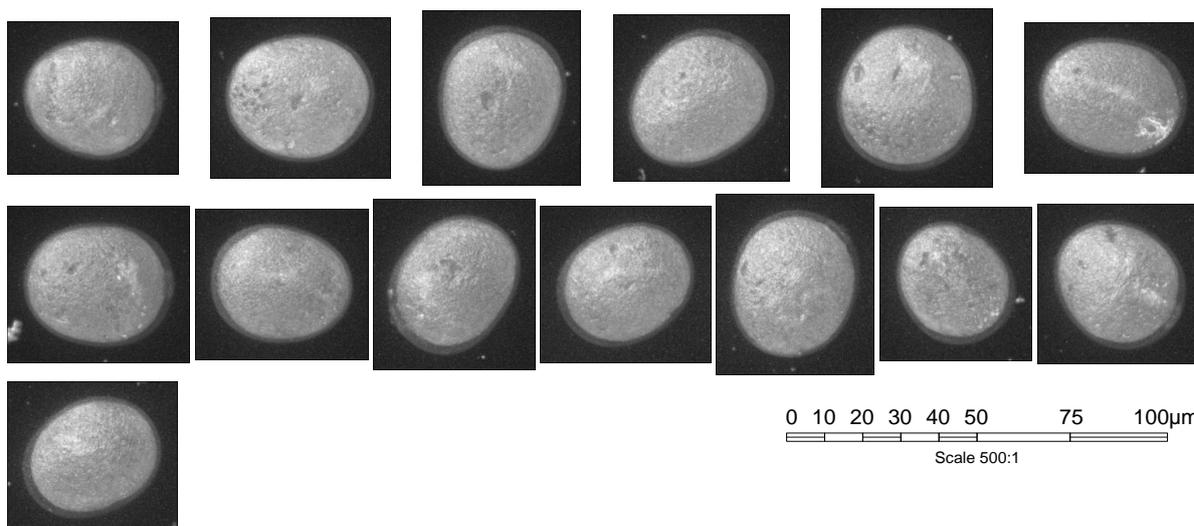
5 unknown positives (6.02%):



A.6 Poaceae (Süßgräser)



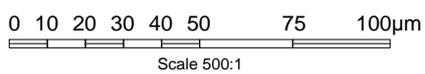
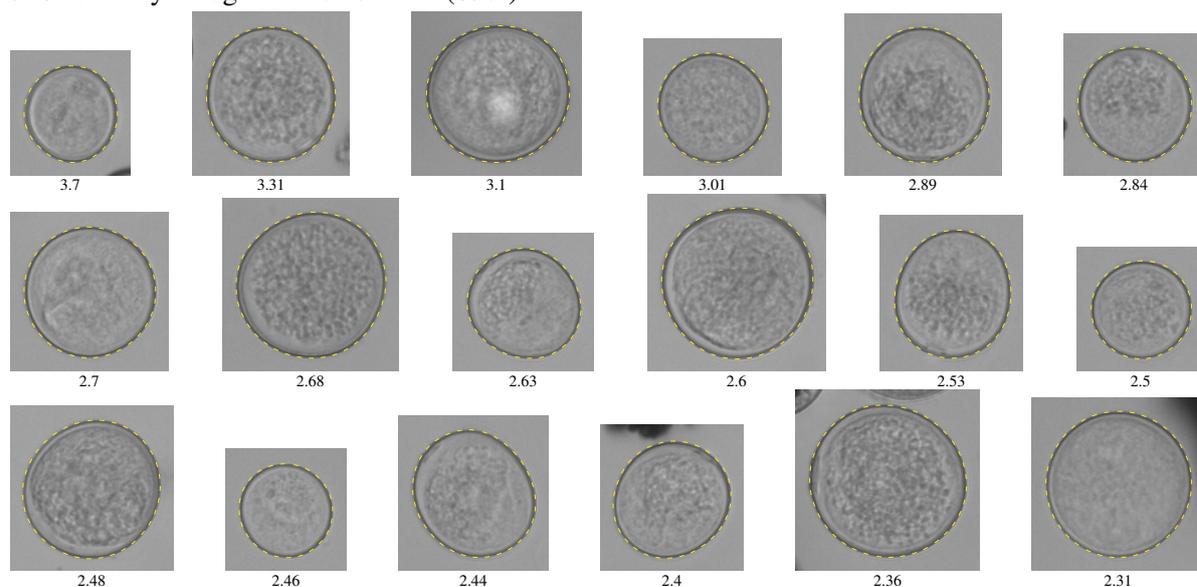
Confocal data set

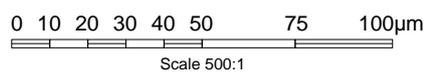
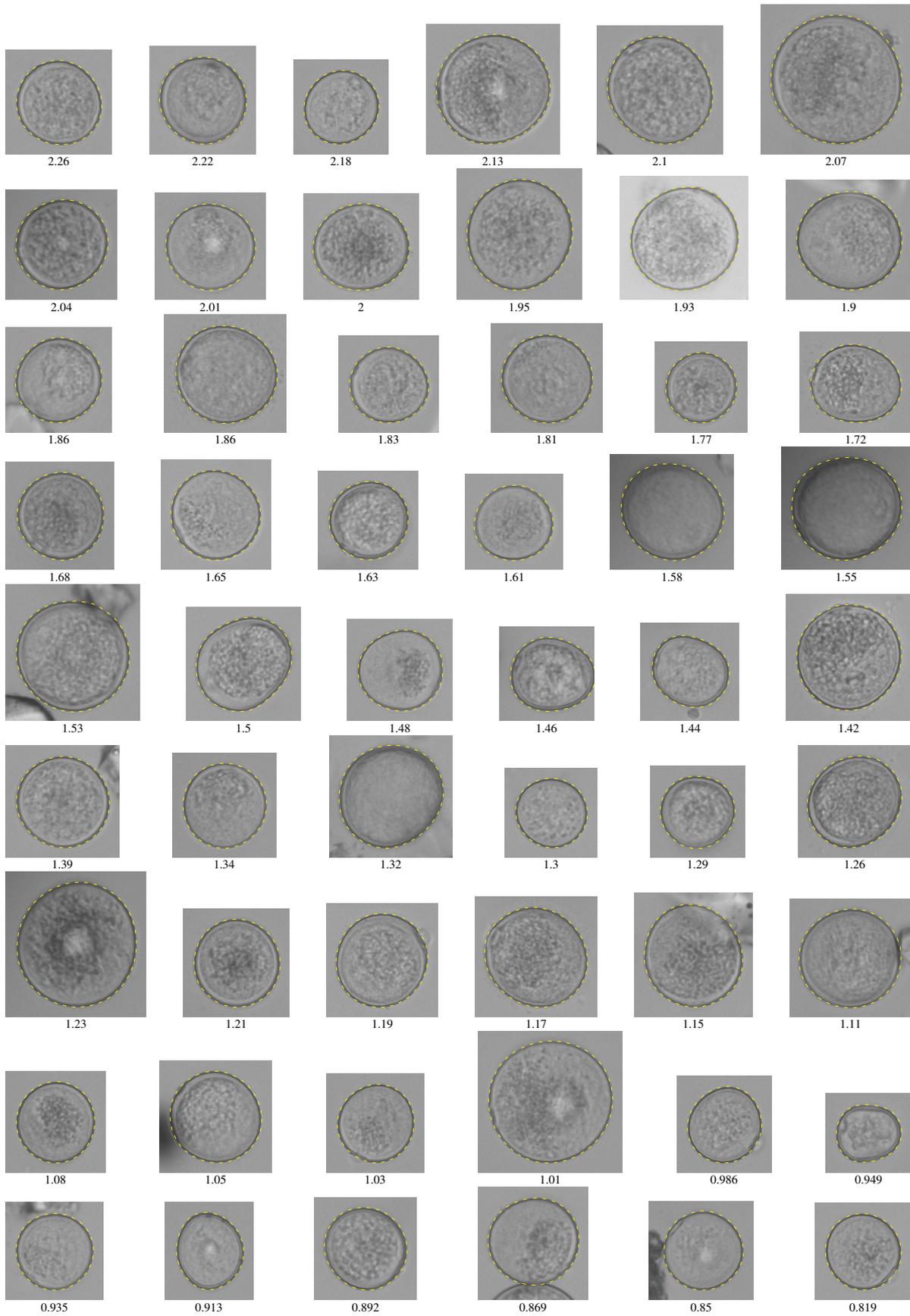


678 Poaceae in pollenmonitor test data set (Freiburg + Zürich)

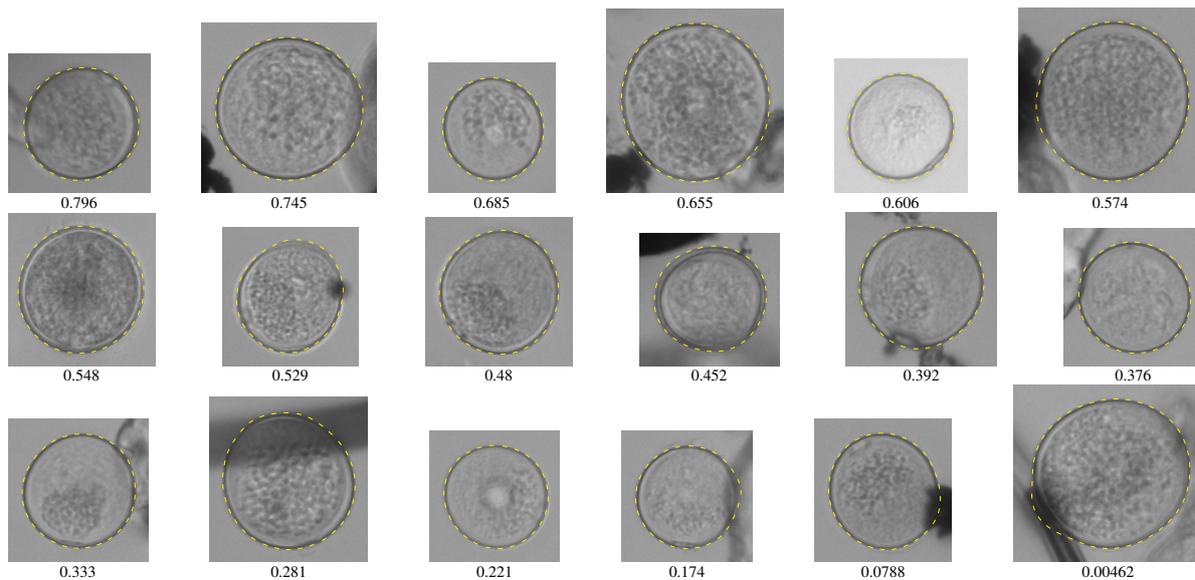
(424 in training data set)

576 correctly recognized as *Poaceae* (85%):

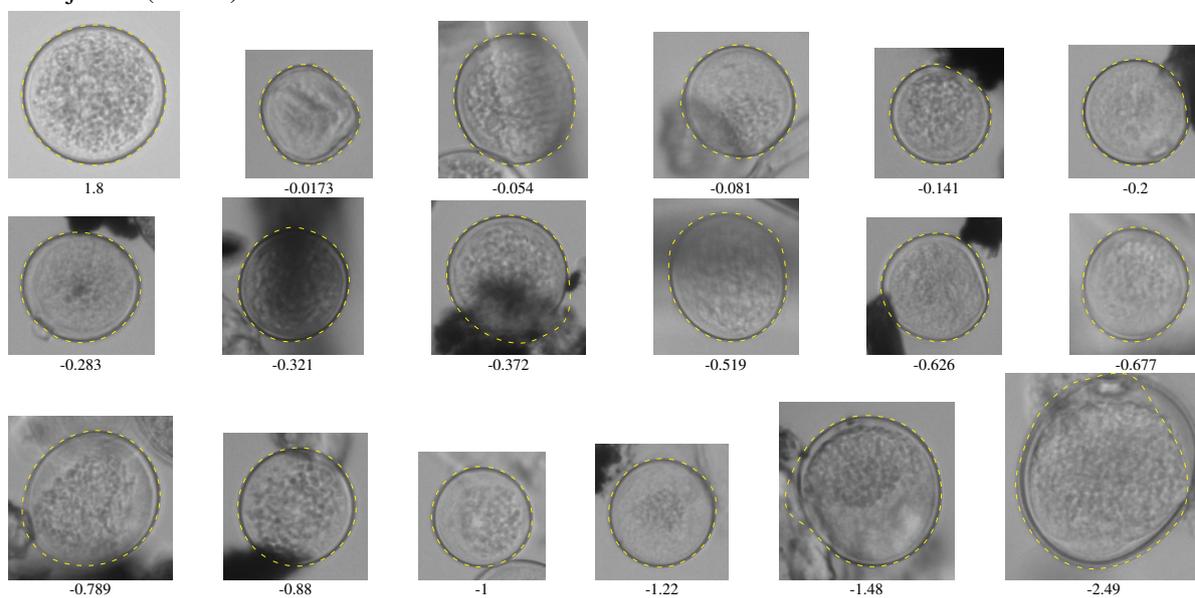




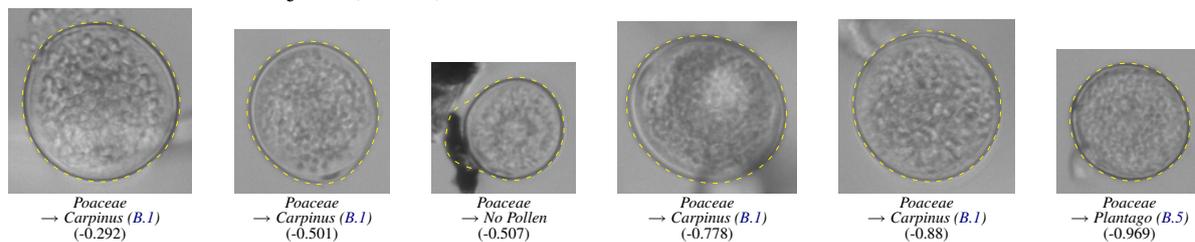
A Highly Allergenic Pollen Taxa

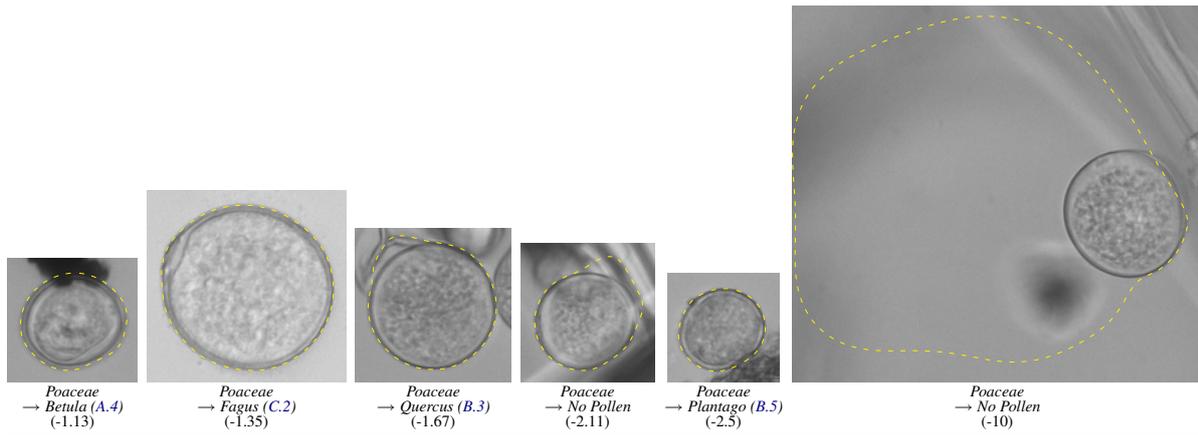


84 rejected (12.4%):



18 classified as other objects (2.65%):

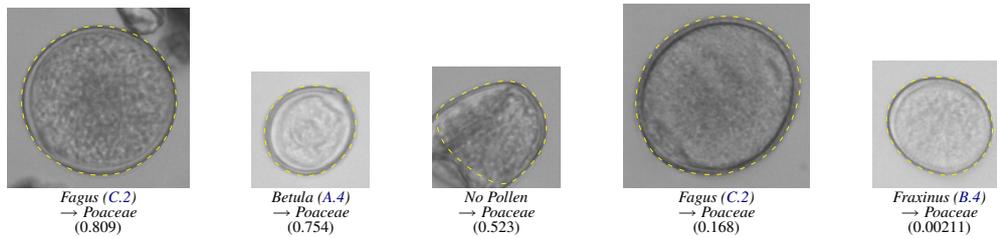




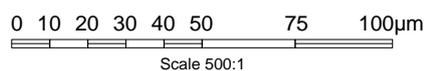
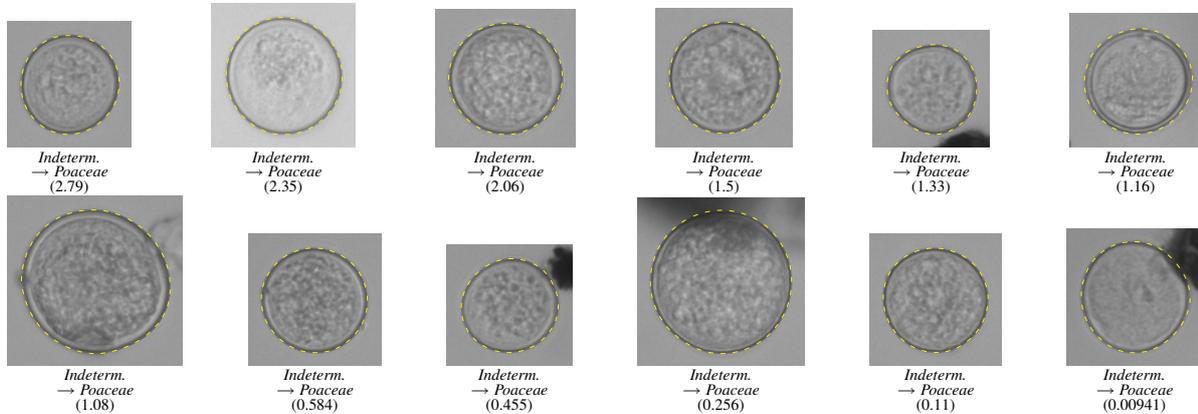
Other objects that were confused with *Poaceae*

(relative to 621 predicted *Poaceae*)

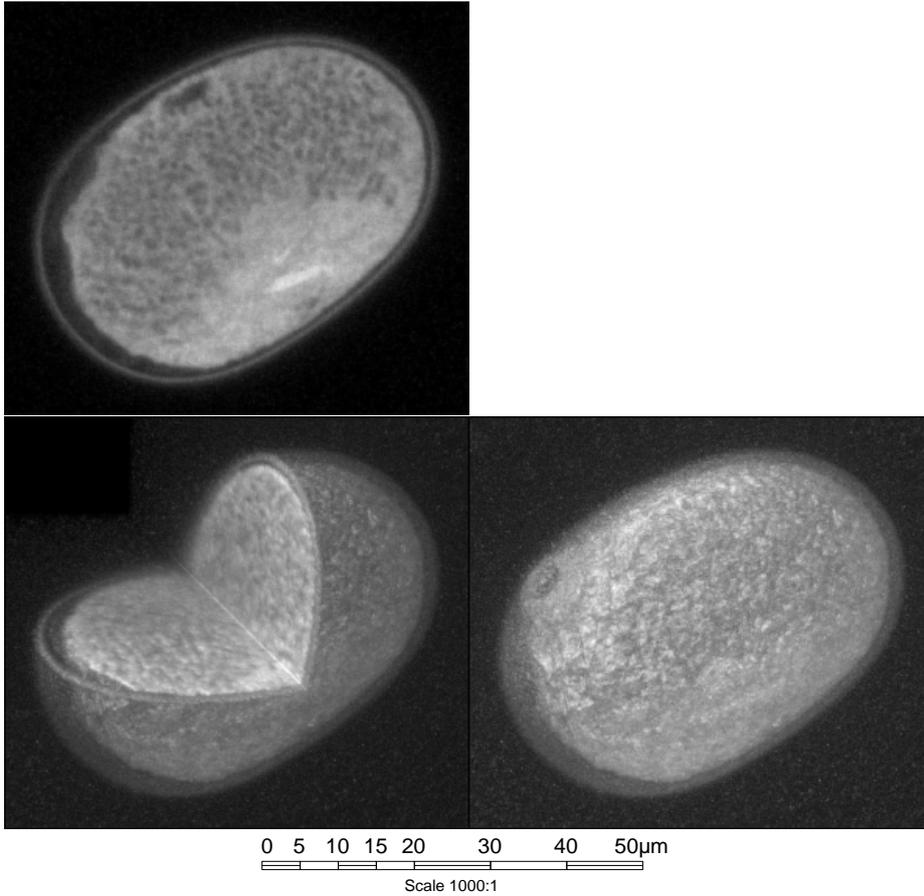
5 false positives (0.805%):



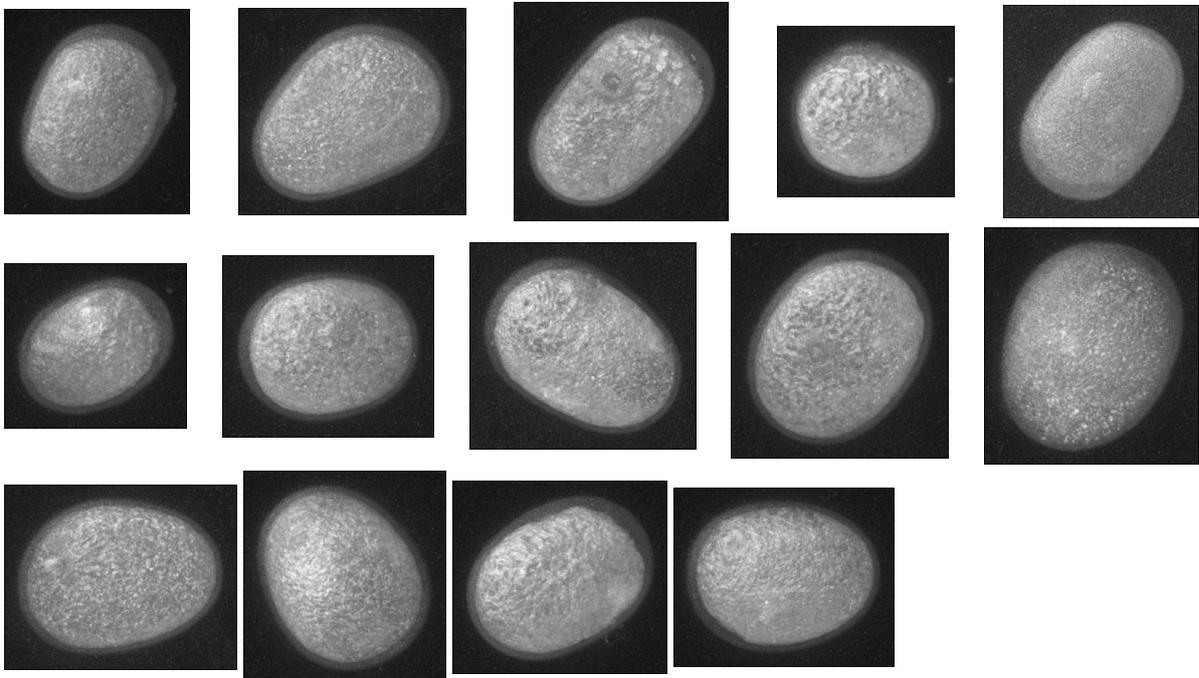
40 unknown positives (6.44%):



A.7 Secale (Roggen)



Confocal data set

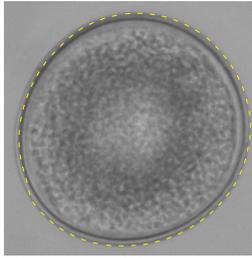


1 Secale in pollenmonitor test data set (Freiburg + Zürich)

(0 in training data set)

0 correctly recognized as *Secale* (0%):

1 rejected (100%):



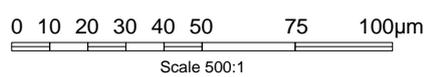
0 classified as other objects (0%):

Other objects that were confused with *Secale*

(relative to 0 predicted *Secale*)

0 false positives (nan%):

0 unknown positives (nan%):



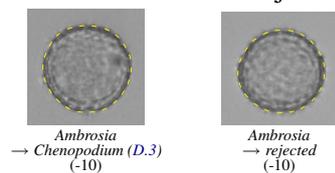
A.8 Ambrosia (Traubenkraut)

2 Ambrosia in pollenmonitor test data set (Freiburg + Zürich) (0 in training data set)

0 correctly recognized as *Ambrosia* (0%):

0 rejected (0%):

2 classified as other objects (100%):

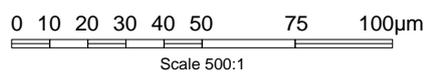


Other objects that were confused with *Ambrosia* (relative to 0 predicted *Ambrosia*)

0 false positives (nan%):

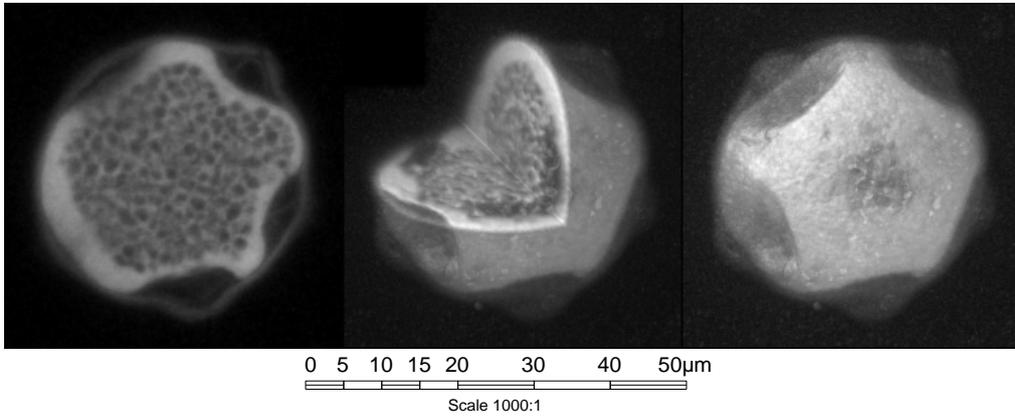
0 unknown positives (nan%):

new

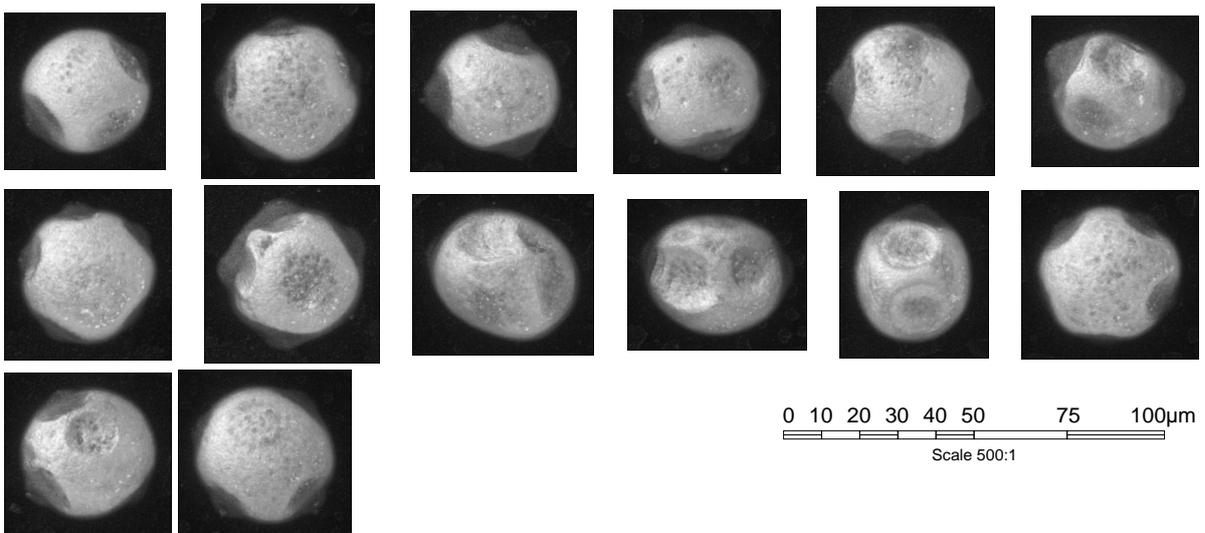


B Moderately Allergenic Pollen Taxa

B.1 Carpinus (Hainbuche)



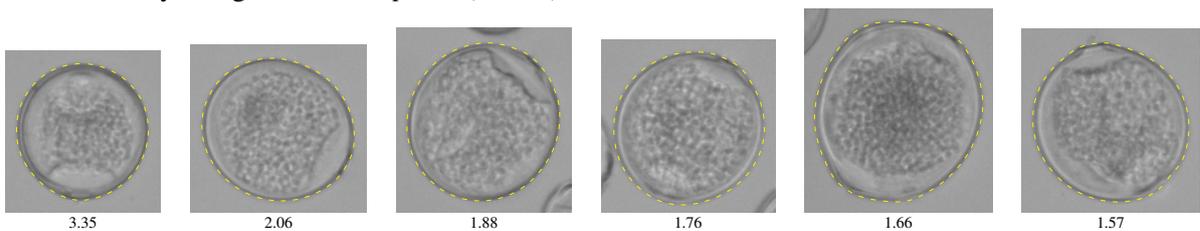
Confocal data set

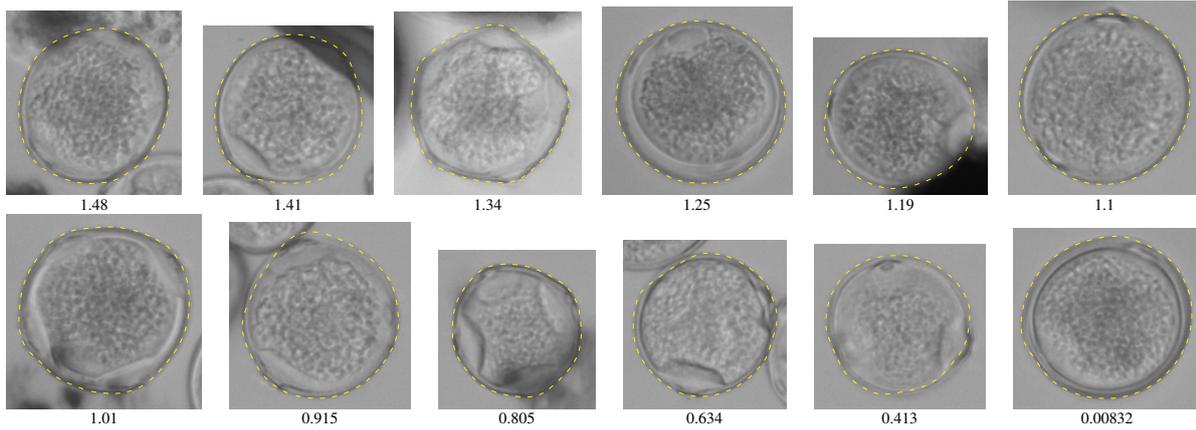


1131 *Carpinus* in pollenmonitor test data set (Freiburg + Zürich)

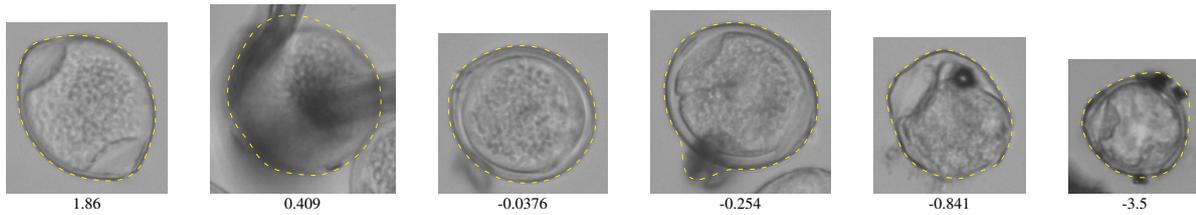
(193 in training data set)

1060 correctly recognized as *Carpinus* (93.7%):

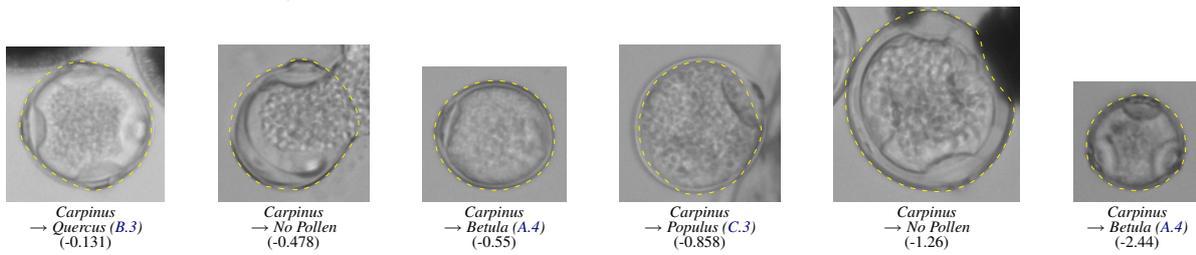




52 rejected (4.6%):



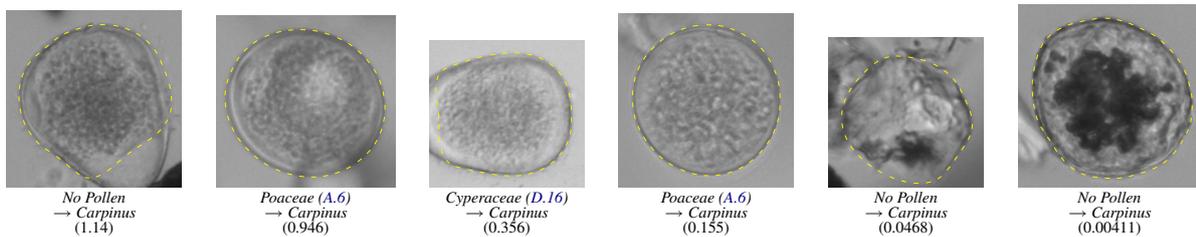
19 classified as other objects (1.68%):



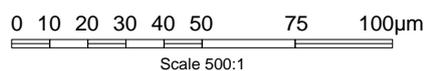
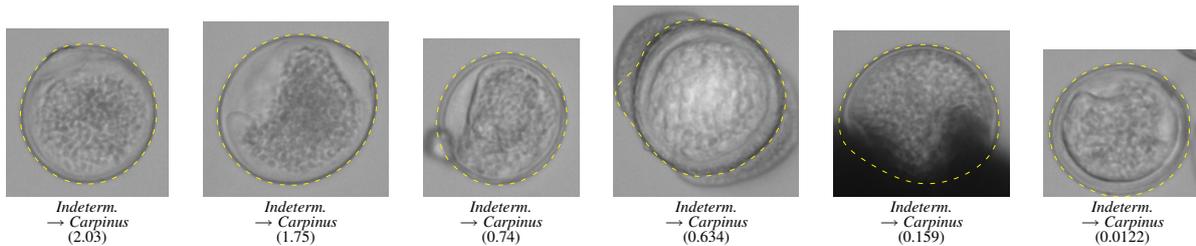
Other objects that were confused with *Carpinus*

(relative to 1084 predicted *Carpinus*)

11 false positives (1.01%):



13 unknown positives (1.2%):



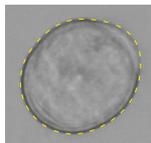
B.2 *Ostrya Carpinifolia* (Hopfenbuche)

1 *Ostr. Carp.* in pollenmonitor test data set (Freiburg + Zürich) (0 in training data set)

0 correctly recognized as *Ostr. Carp.* (0%):

0 rejected (0%):

1 classified as other objects (100%):

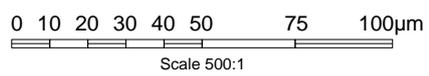


Ostr. Carp.
→ *Quercus* (B.3)
(-10)

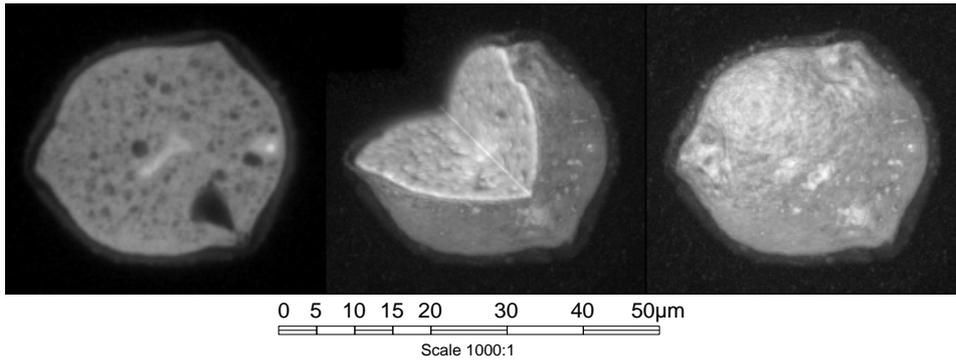
Other objects that were confused with *Ostr. Carp.* (relative to 0 predicted *Ostr. Carp.*)

0 false positives (nan%):

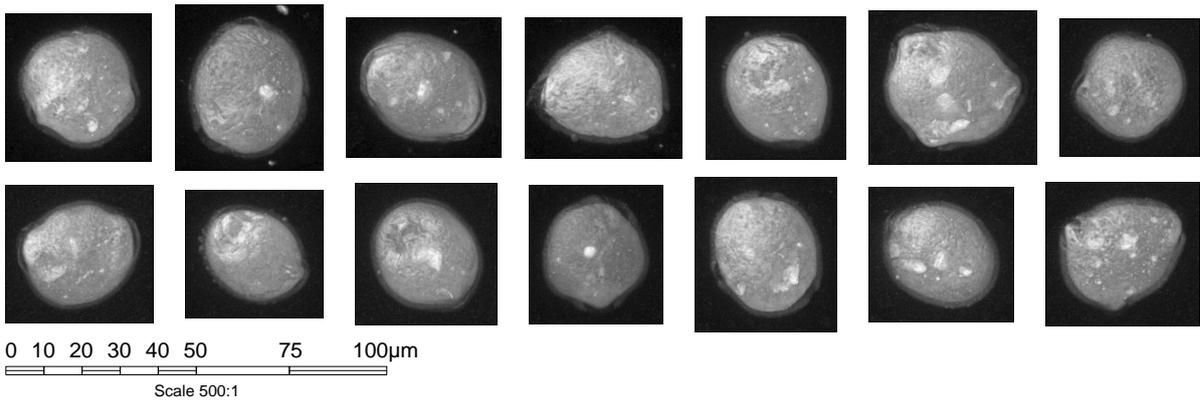
0 unknown positives (nan%):



B.3 Quercus (Eiche)



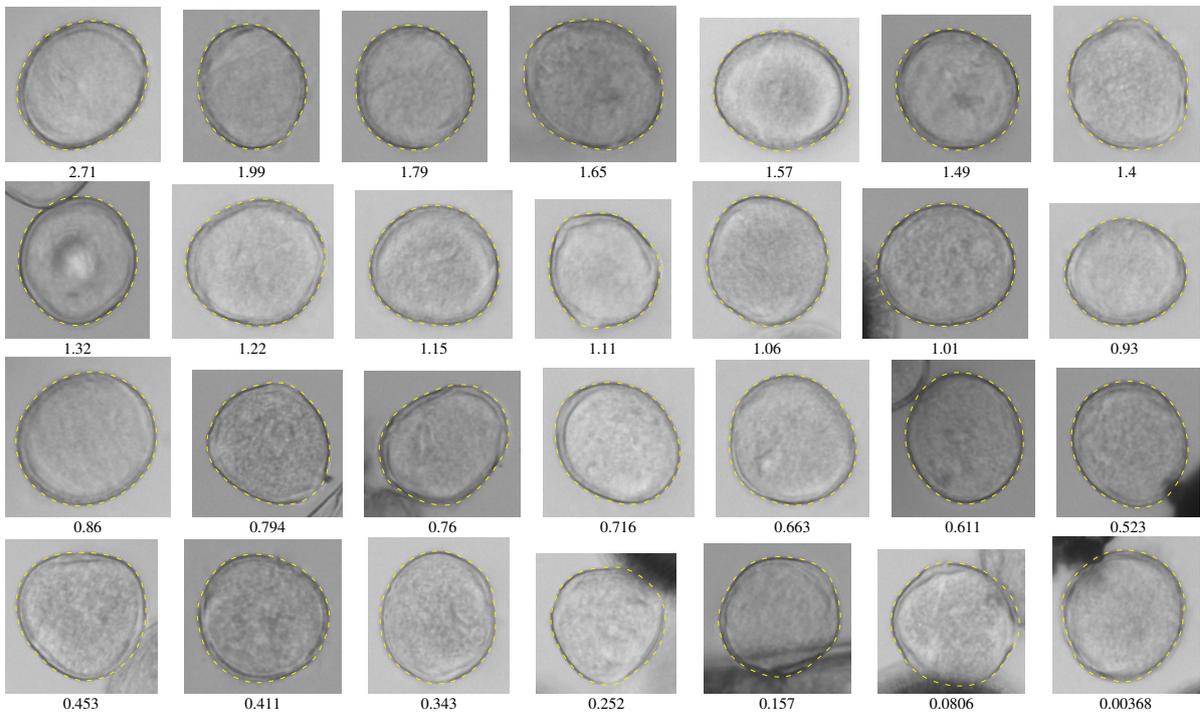
Confocal data set



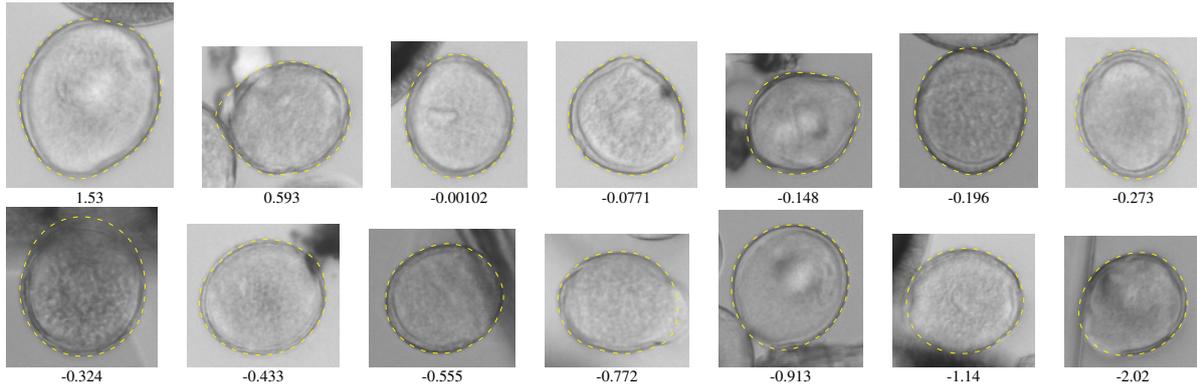
732 *Quercus* in pollenmonitor test data set (Freiburg + Zürich)

(59 in training data set)

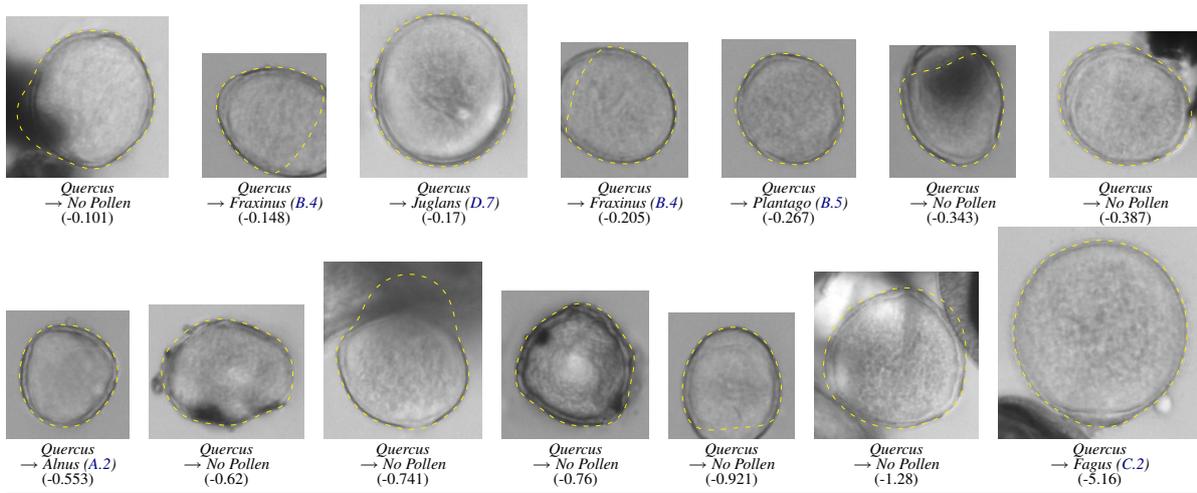
555 correctly recognized as *Quercus* (75.8%):



147 rejected (20.1%):



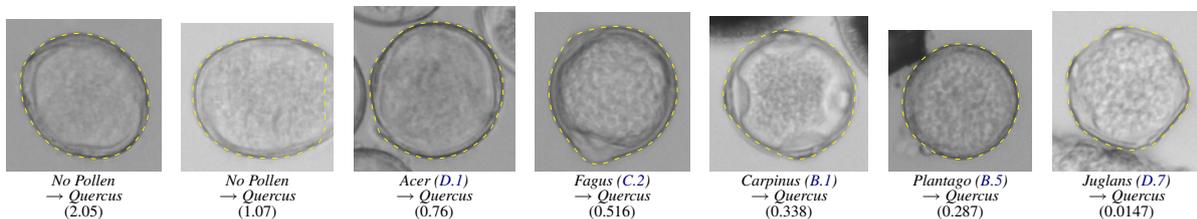
30 classified as other objects (4.1%):



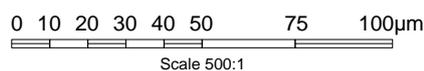
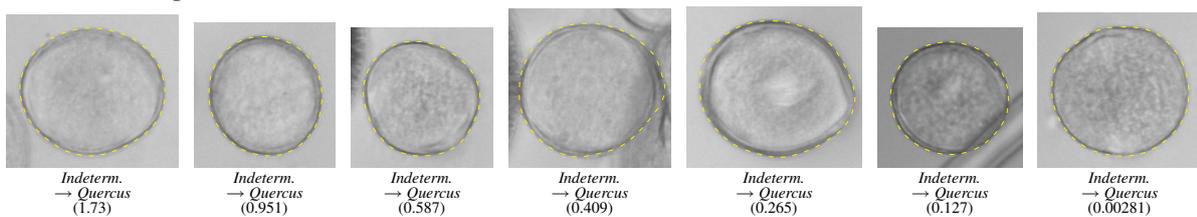
Other objects that were confused with *Quercus*

(relative to 716 predicted *Quercus*)

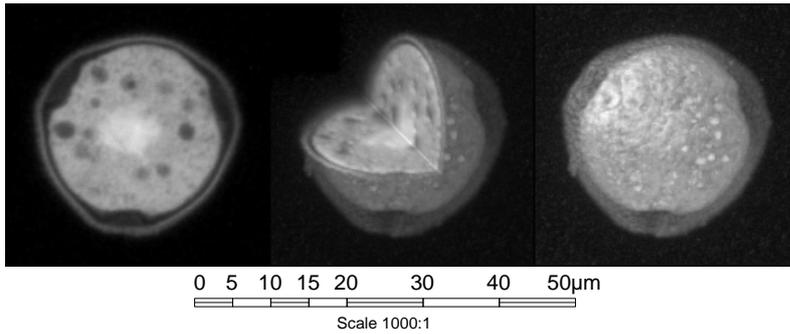
46 false positives (6.42%):



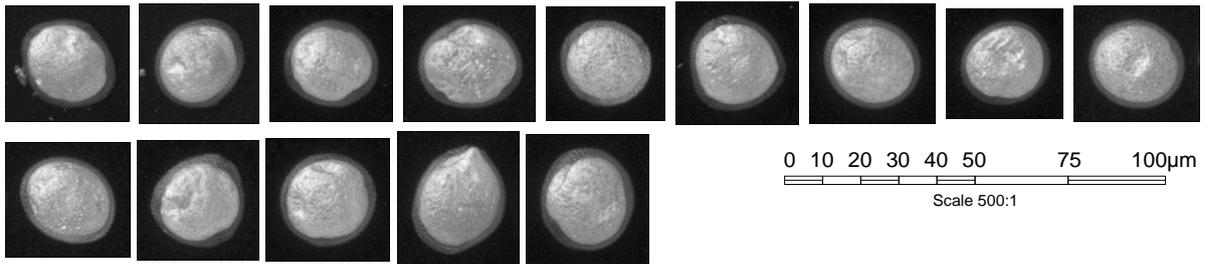
115 unknown positives (16.1%):



B.4 Fraxinus (Esche)



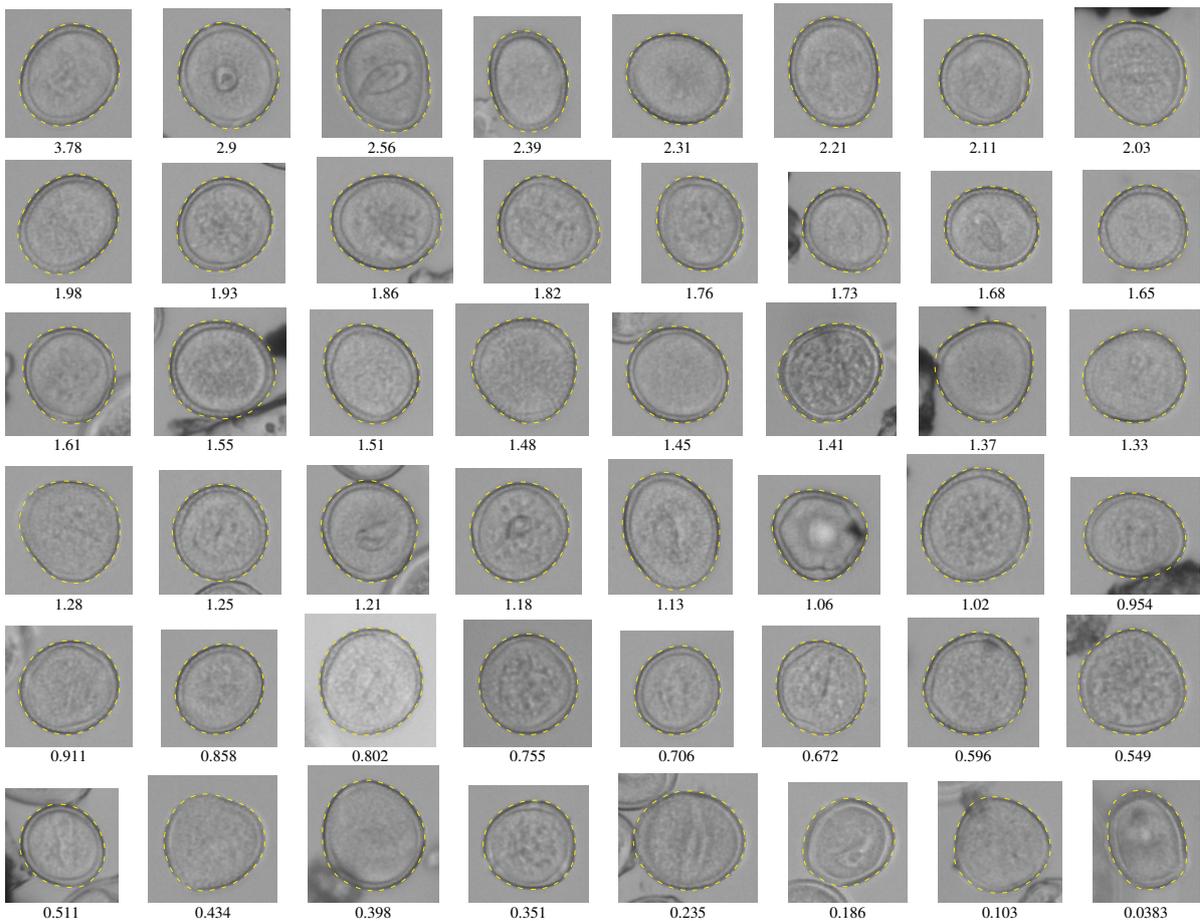
Confocal data set



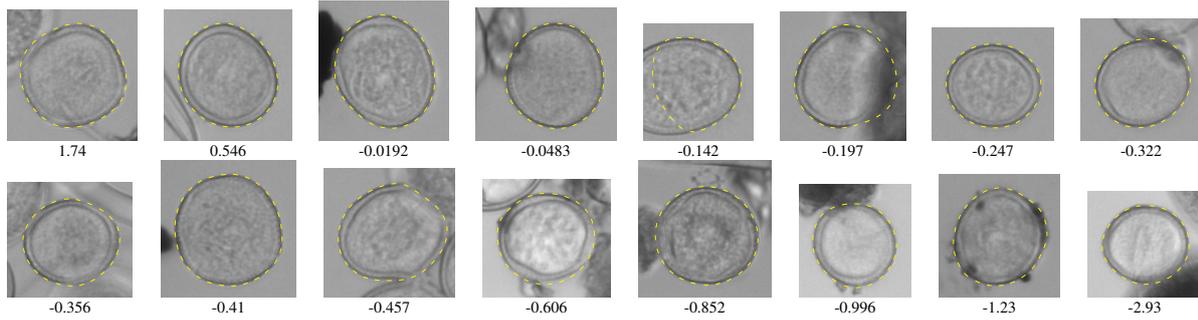
563 *Fraxinus* in pollenmonitor test data set (Freiburg + Zürich)

(160 in training data set)

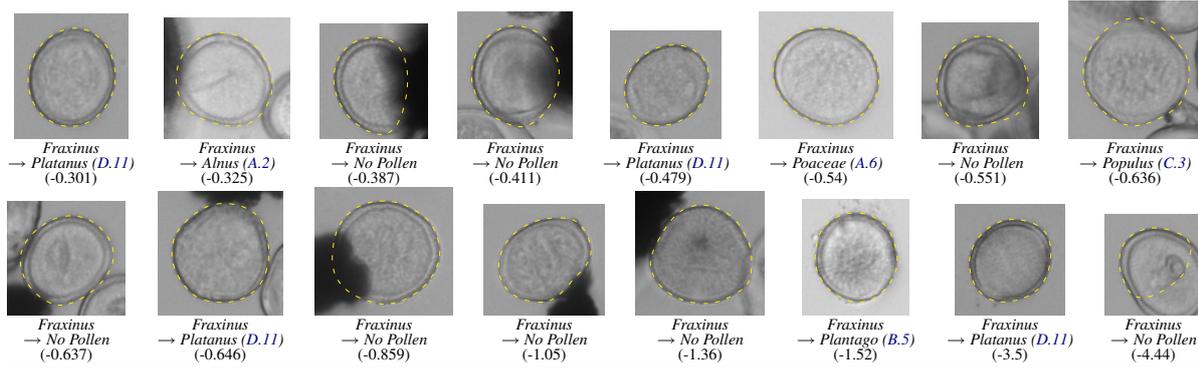
463 correctly recognized as *Fraxinus* (82.2%):



82 rejected (14.6%):



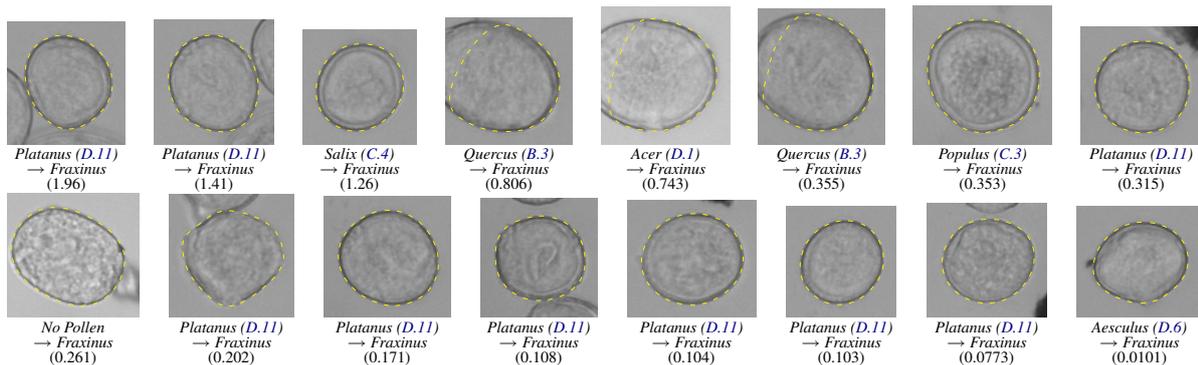
18 classified as other objects (3.2%):



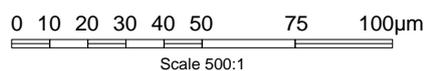
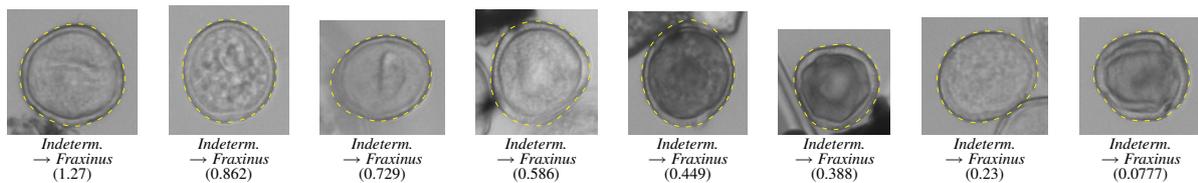
Other objects that were confused with *Fraxinus*

(relative to 502 predicted *Fraxinus*)

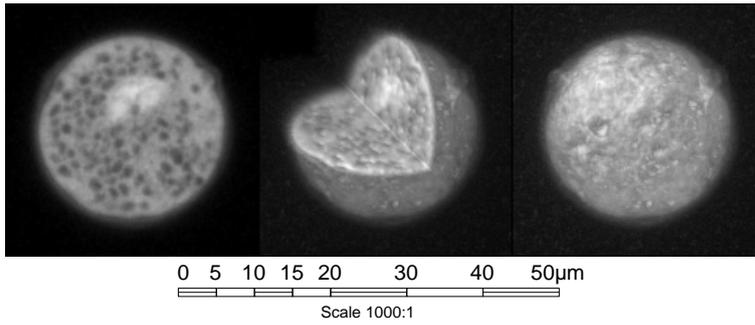
21 false positives (4.18%):



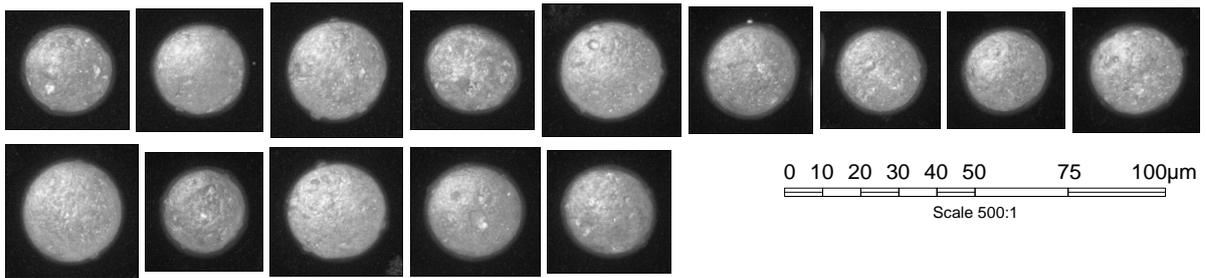
18 unknown positives (3.59%):



B.5 Plantago (Wegerich)



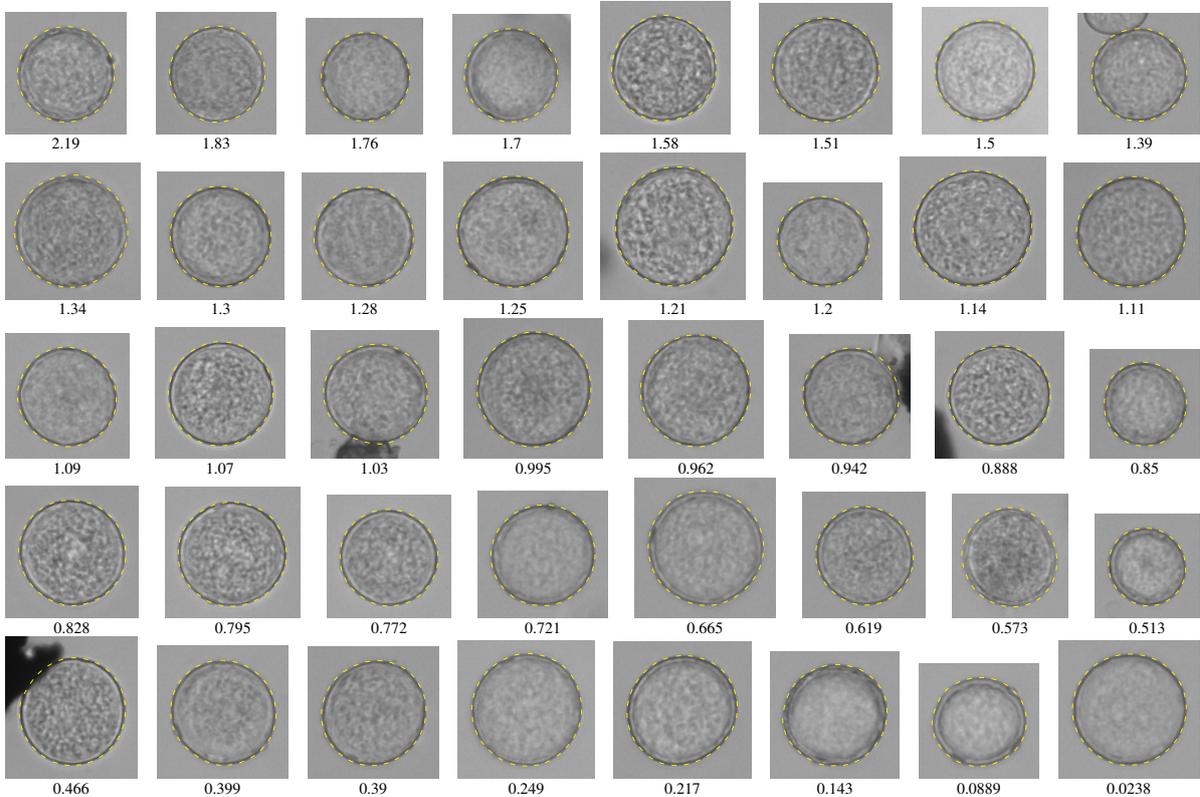
Confocal data set



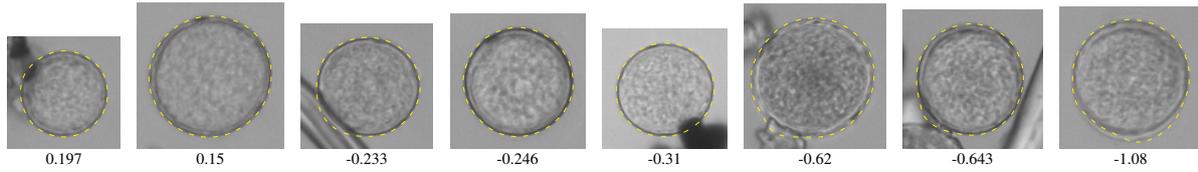
110 *Plantago* in pollenmonitor test data set (Freiburg + Zürich)

(63 in training data set)

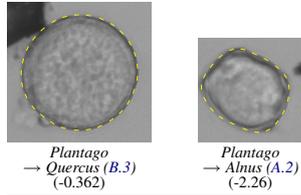
100 correctly recognized as *Plantago* (90.9%):



8 rejected (7.27%):



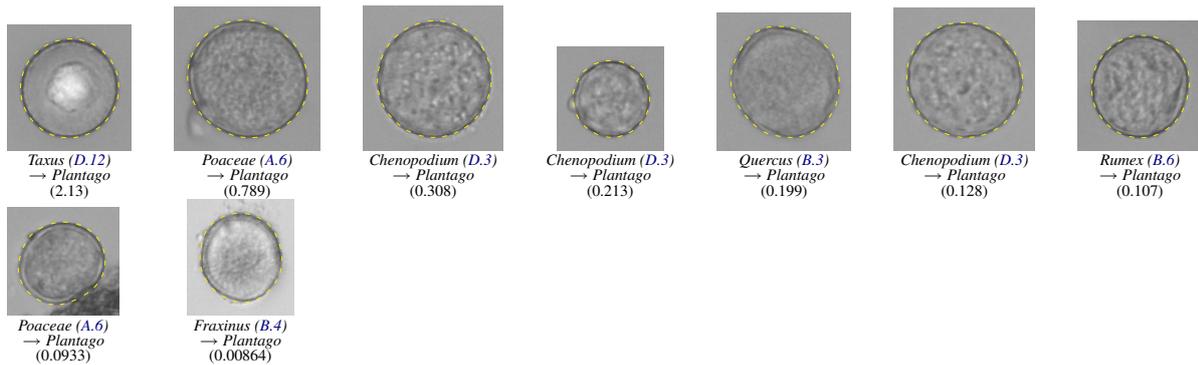
2 classified as other objects (1.82%):



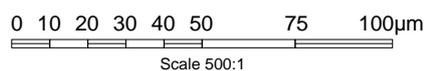
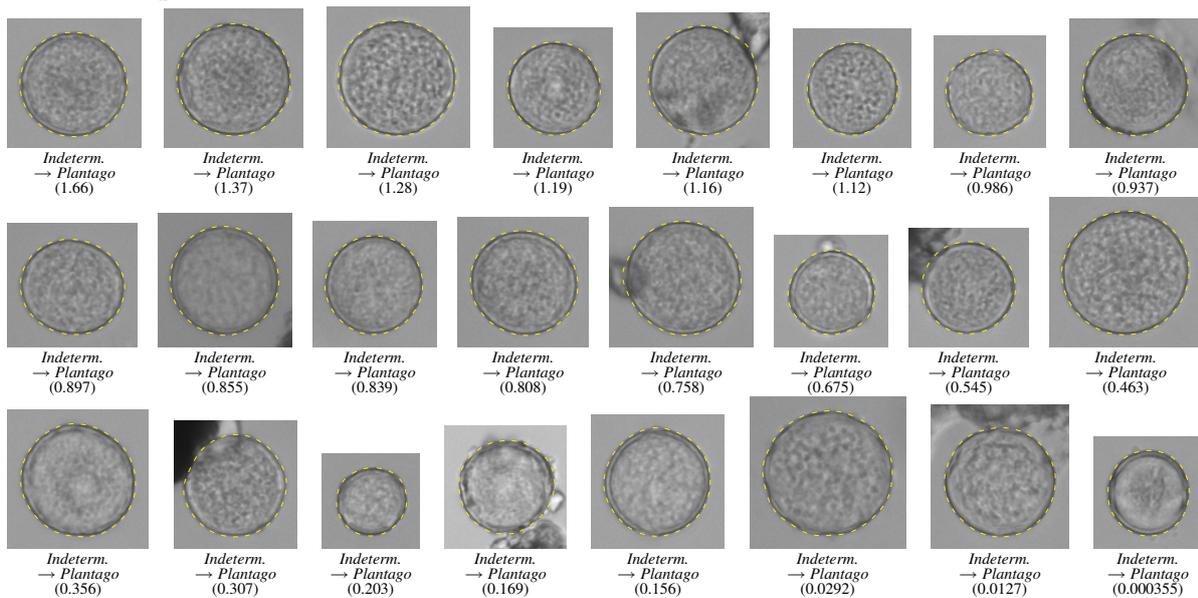
Other objects that were confused with *Plantago*

(relative to 158 predicted *Plantago*)

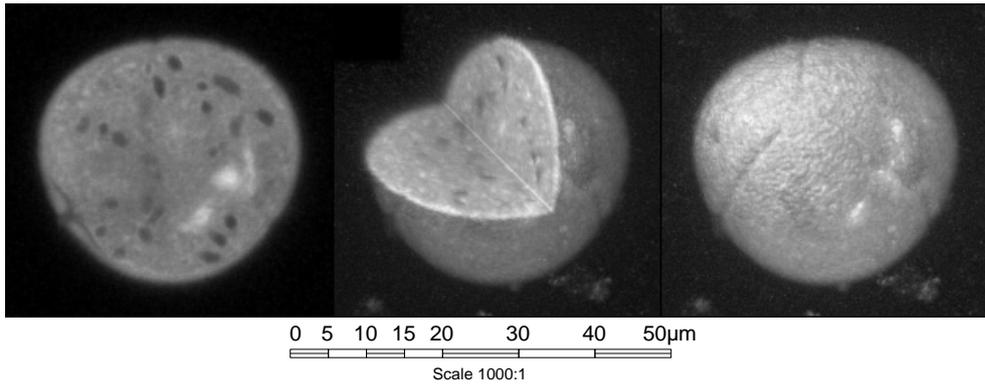
9 false positives (5.7%):



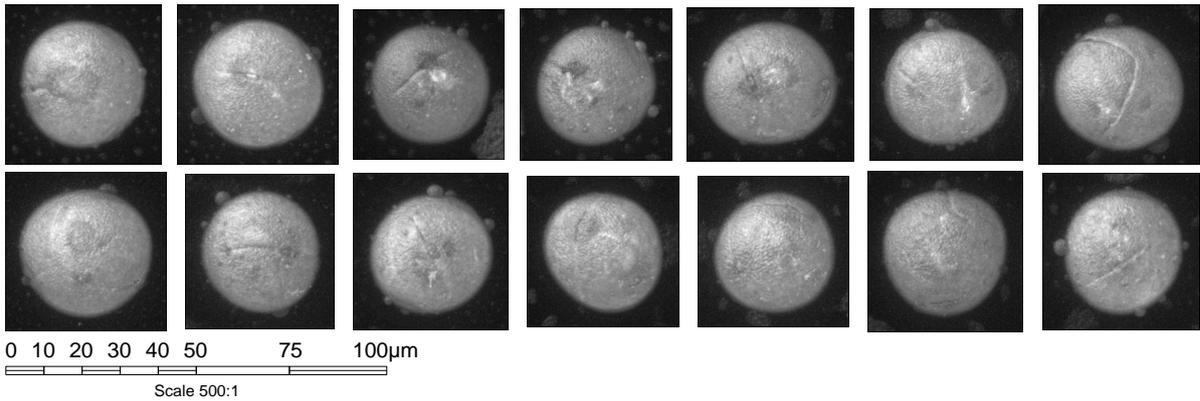
49 unknown positives (31%):



B.6 Rumex (Ampfer)



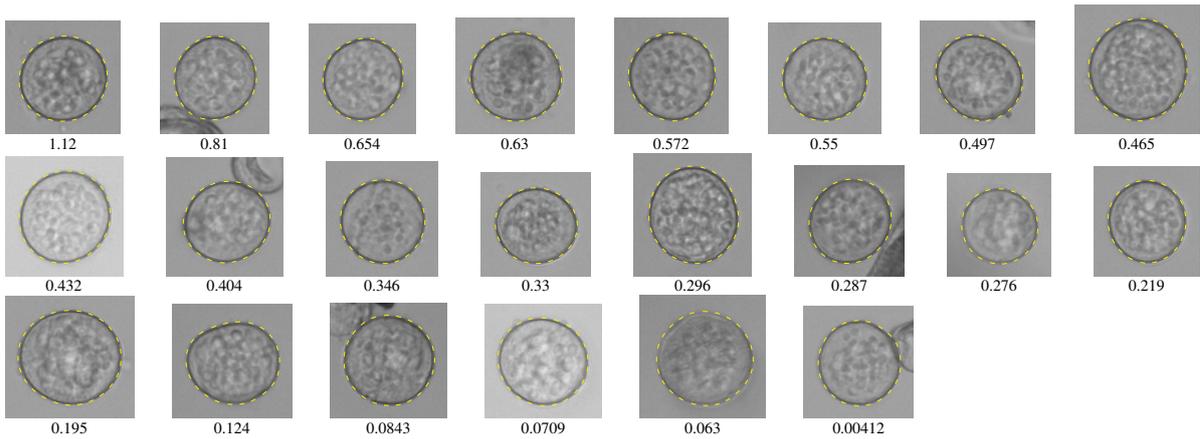
Confocal data set



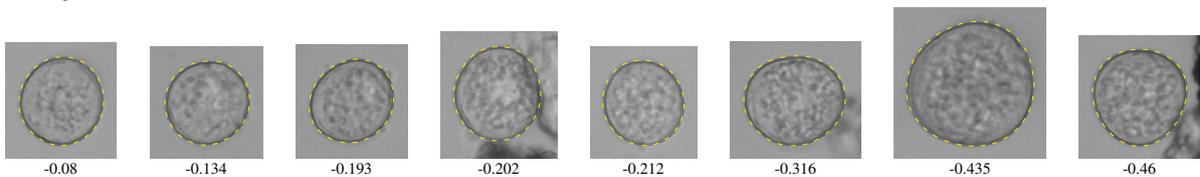
38 Rumex in pollenmonitor test data set (Freiburg + Zürich)

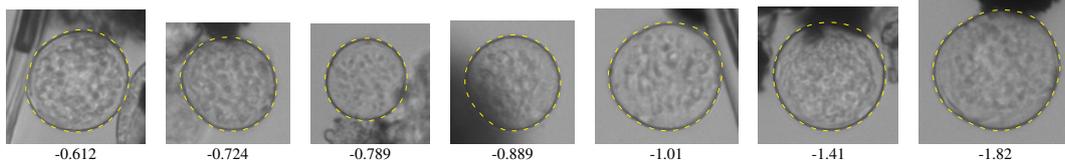
(14 in training data set)

22 correctly recognized as *Rumex* (57.9%):

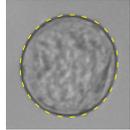


15 rejected (39.5%):





1 classified as other objects (2.63%):

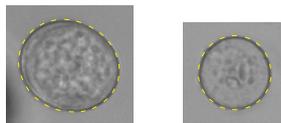


Rumex
→ Plantago (B.5)
(-1.02)

Other objects that were confused with Rumex

(relative to 40 predicted Rumex)

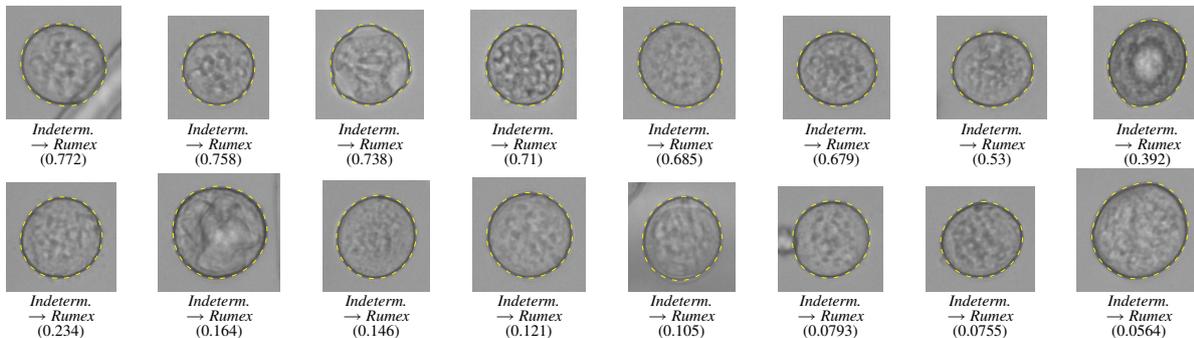
2 false positives (5%):



Poaceae (A.6)
→ Rumex
(0.122)

Urtica (D.15)
→ Rumex
(0.0347)

16 unknown positives (40%):



Indeterm.
→ Rumex
(0.772)

Indeterm.
→ Rumex
(0.758)

Indeterm.
→ Rumex
(0.738)

Indeterm.
→ Rumex
(0.71)

Indeterm.
→ Rumex
(0.685)

Indeterm.
→ Rumex
(0.679)

Indeterm.
→ Rumex
(0.53)

Indeterm.
→ Rumex
(0.392)

Indeterm.
→ Rumex
(0.234)

Indeterm.
→ Rumex
(0.164)

Indeterm.
→ Rumex
(0.146)

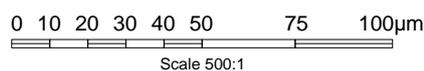
Indeterm.
→ Rumex
(0.121)

Indeterm.
→ Rumex
(0.105)

Indeterm.
→ Rumex
(0.0793)

Indeterm.
→ Rumex
(0.0755)

Indeterm.
→ Rumex
(0.0564)



Scale 500:1

C Lowly Allergenic Pollen Taxa

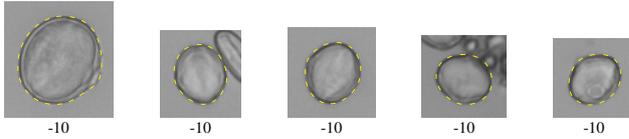
C.1 Castanea (Eßkastanie)

11 *Castanea* in pollenmonitor test data set (Freiburg + Zürich)

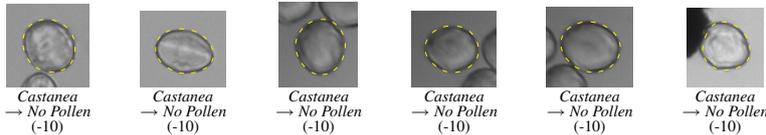
(0 in training data set)

0 correctly recognized as *Castanea* (0%):

5 rejected (45.5%):



6 classified as other objects (54.5%):

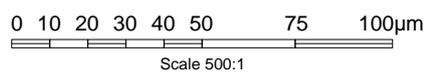


Other objects that were confused with *Castanea*

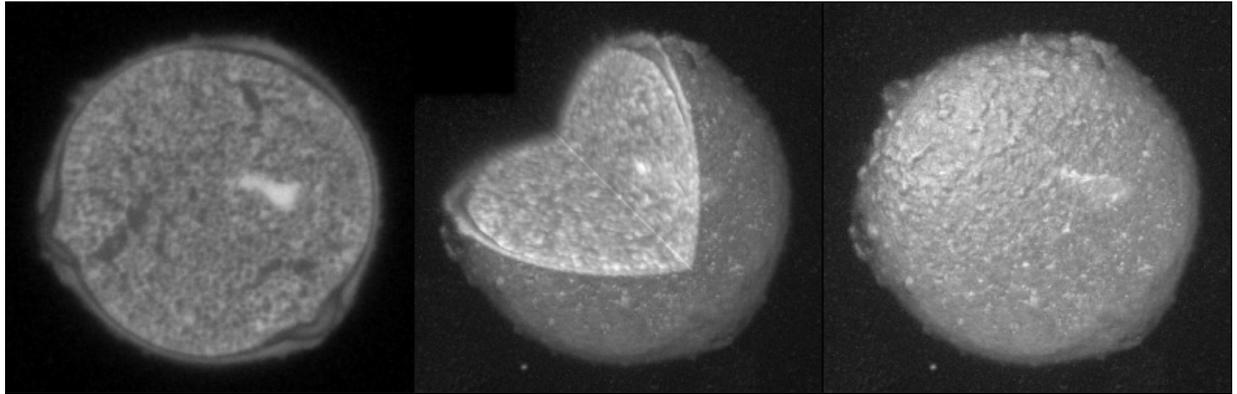
(relative to 0 predicted *Castanea*)

0 false positives (nan%):

0 unknown positives (nan%):

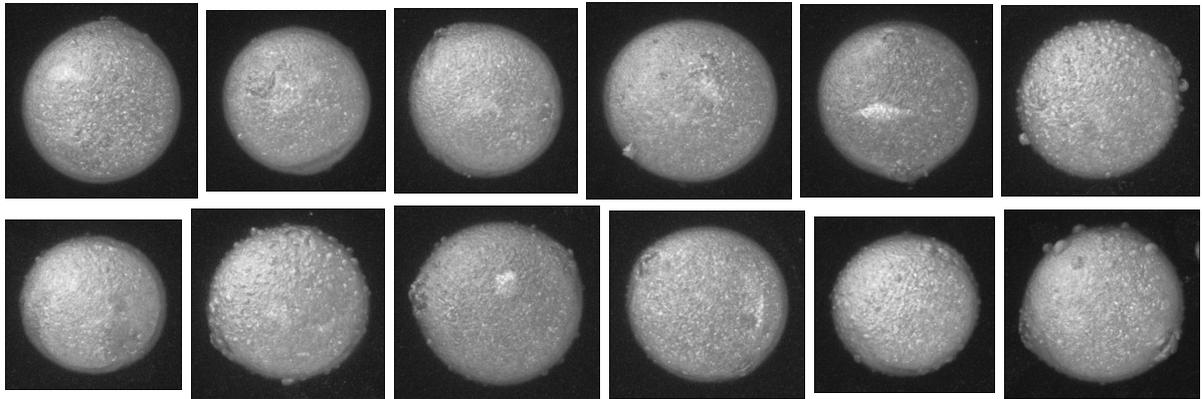


C.2 Fagus (Rotbuche)



0 5 10 15 20 30 40 50µm
Scale 1000:1

Confocal data set (left out two pollen due to limited space)

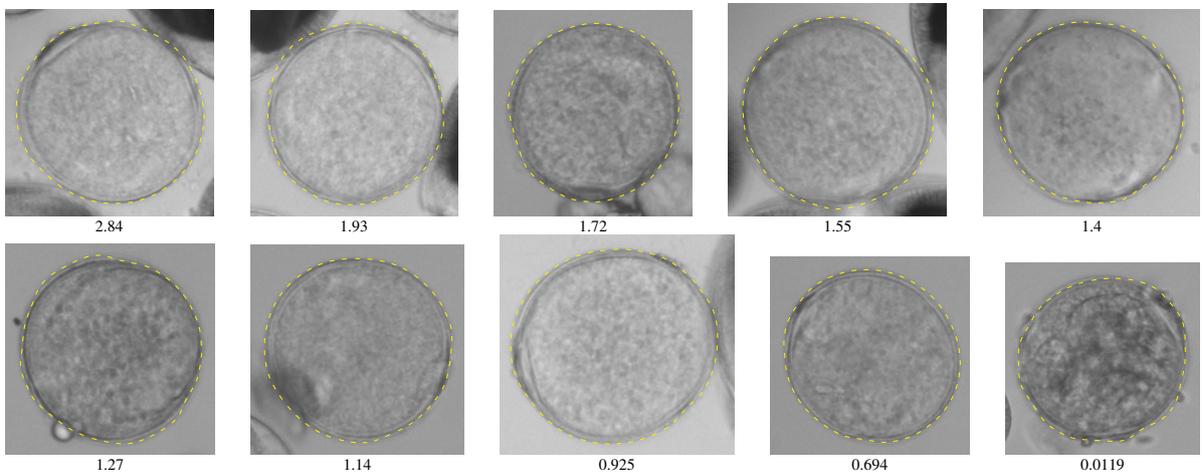


0 10 20 30 40 50 75 100µm
Scale 500:1

1102 *Fagus* in pollenmonitor test data set (Freiburg + Zürich)

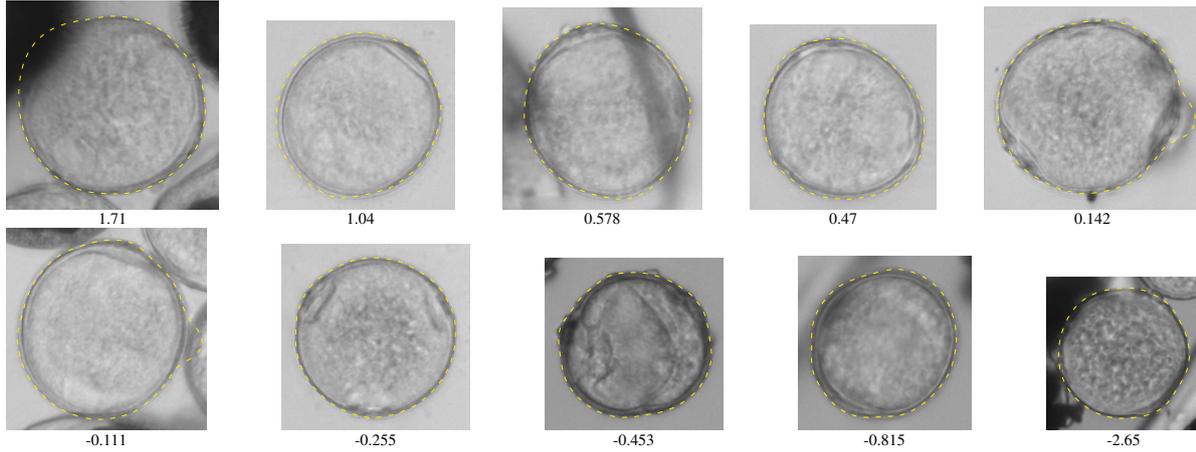
(135 in training data set)

1028 correctly recognized as *Fagus* (93.3%):

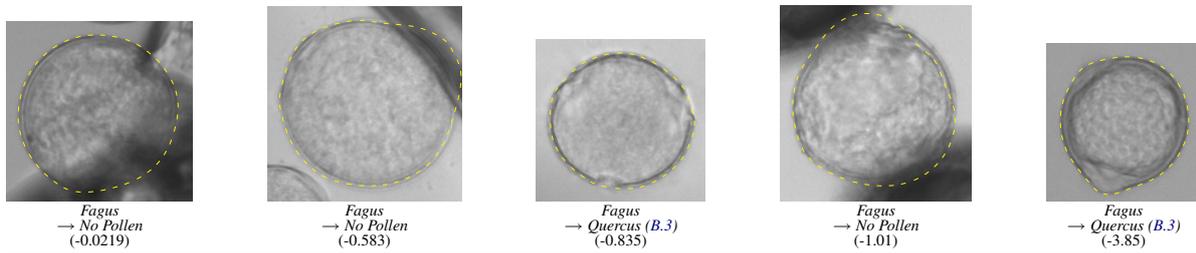


0 10 20 30 40 50 75 100µm
Scale 500:1

62 rejected (5.63%):



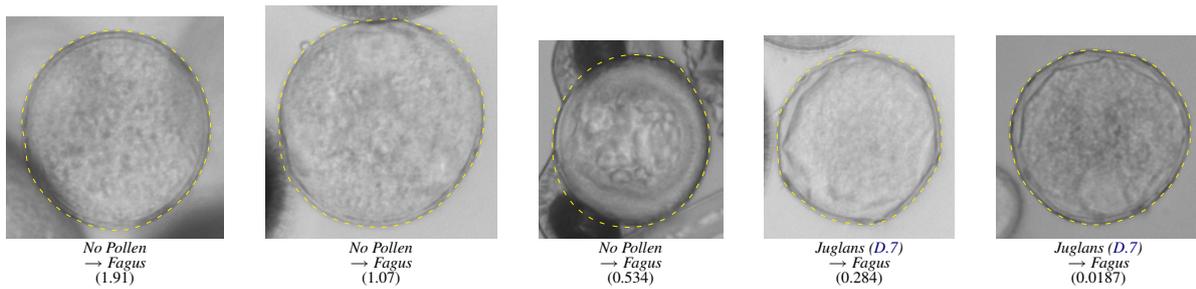
12 classified as other objects (1.09%):



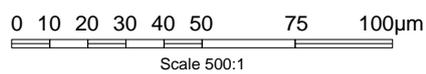
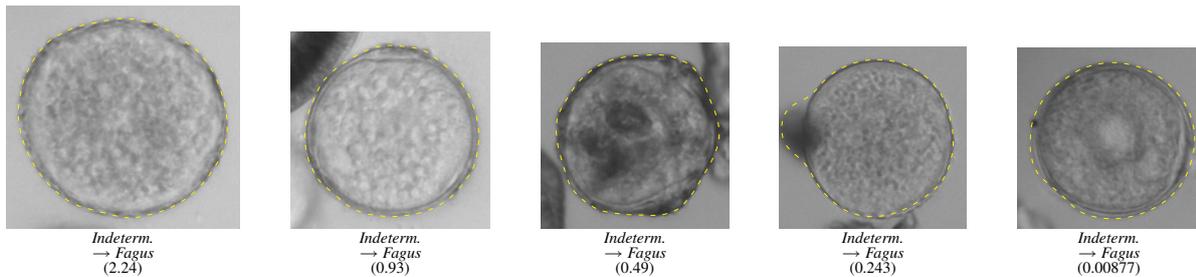
Other objects that were confused with *Fagus*

(relative to 1115 predicted *Fagus*)

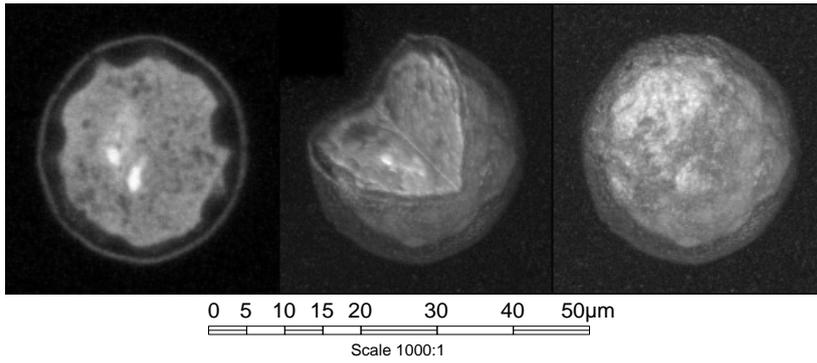
39 false positives (3.5%):



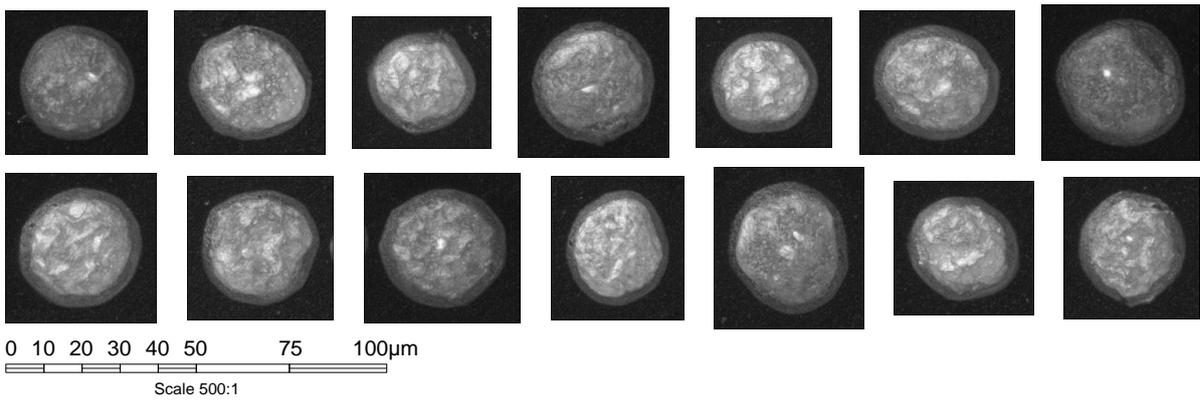
48 unknown positives (4.3%):



C.3 Populus (Pappel)



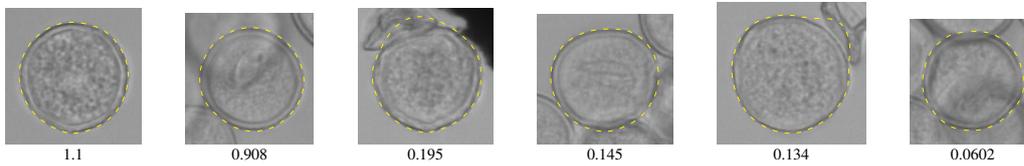
Confocal data set



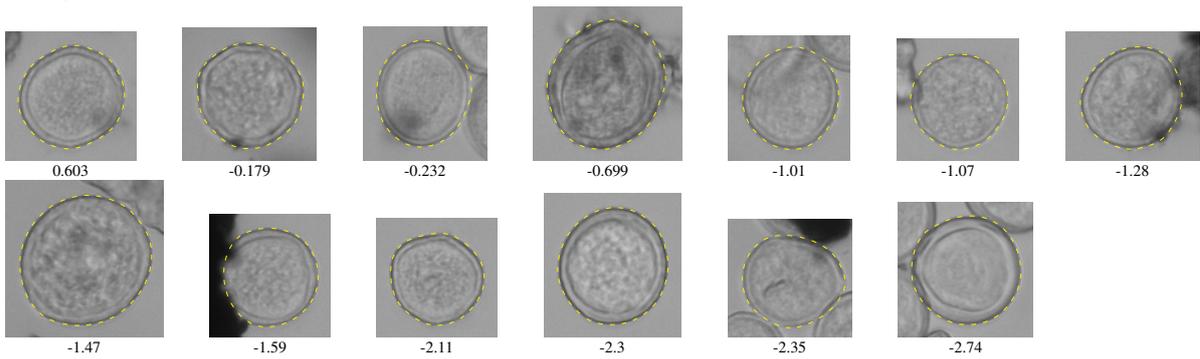
23 Populus in pollenmonitor test data set (Freiburg + Zürich)

(13 in training data set)

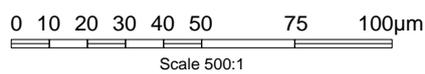
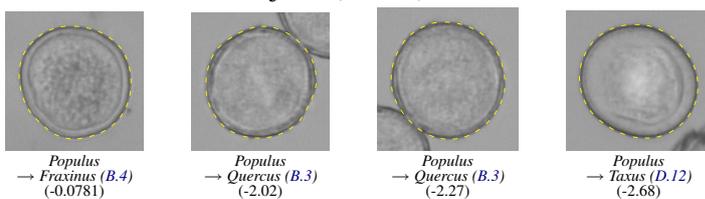
6 correctly recognized as *Populus* (26.1%):



13 rejected (56.5%):



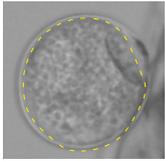
4 classified as other objects (17.4%):



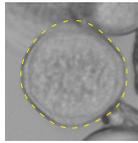
Other objects that were confused with *Populus*

(relative to 11 predicted *Populus*)

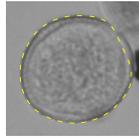
3 false positives (27.3%):



Carpinus (B.1)
→ *Populus*
(0.387)

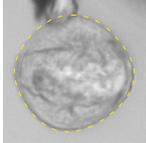


Fraxinus (B.4)
→ *Populus*
(0.0532)

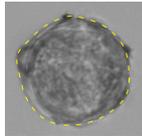


Fraxinus (B.4)
→ *Populus*
(0.0439)

2 unknown positives (18.2%):

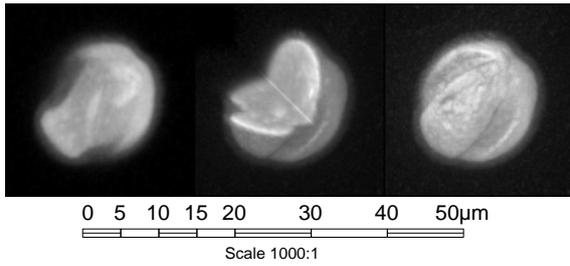


Indeterm.
→ *Populus*
(0.168)

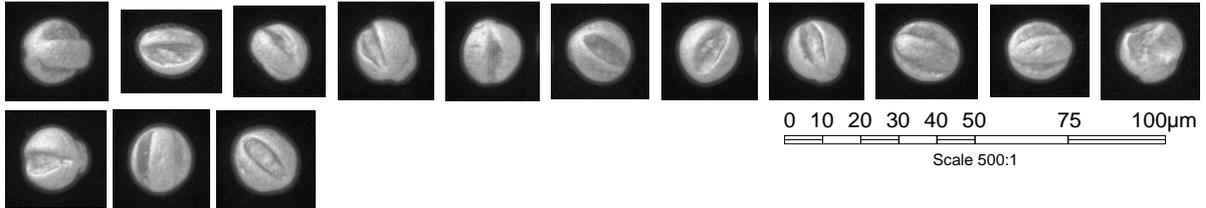


Indeterm.
→ *Populus*
(0.0387)

C.4 Salix (Weide)



Confocal data set

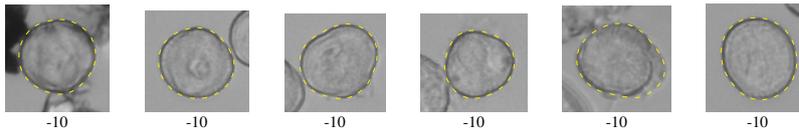


13 Salix in pollenmonitor test data set (Freiburg + Zürich)

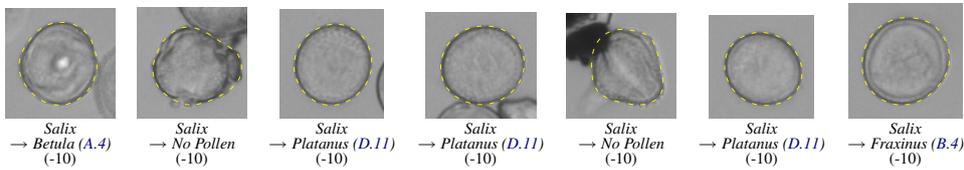
(0 in training data set)

0 correctly recognized as *Salix* (0%):

6 rejected (46.2%):



7 classified as other objects (53.8%):

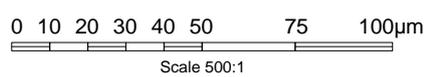


Other objects that were confused with *Salix*

(relative to 0 predicted *Salix*)

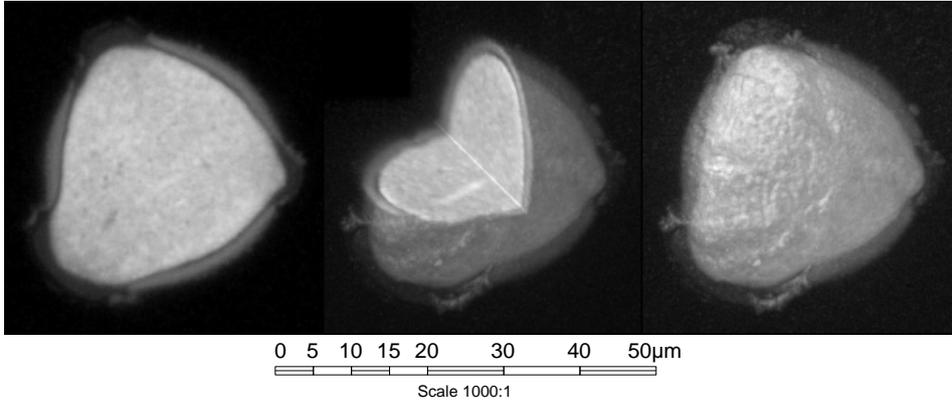
0 false positives (nan%):

0 unknown positives (nan%):

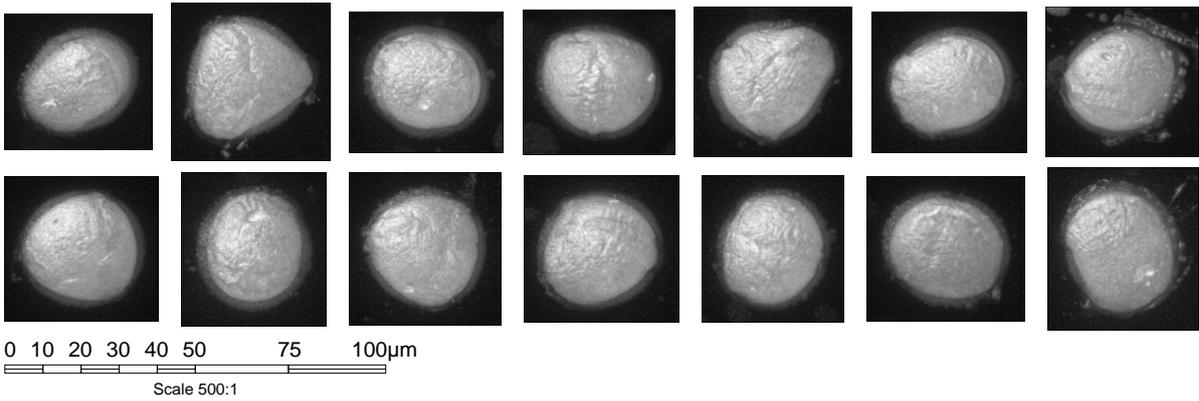


D Non-Allergenic Pollen Taxa

D.1 Acer (Ahorn)



Confocal data set

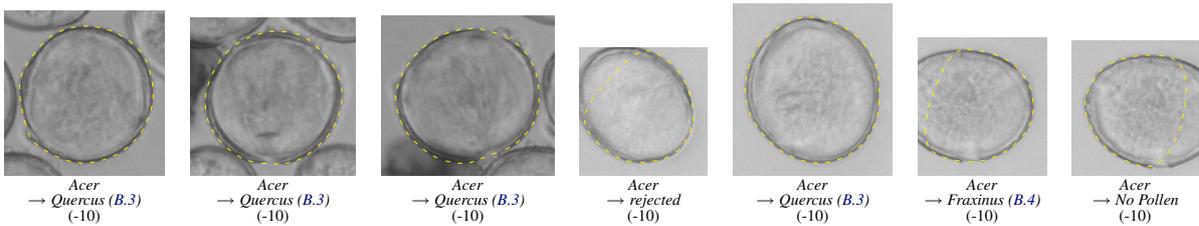


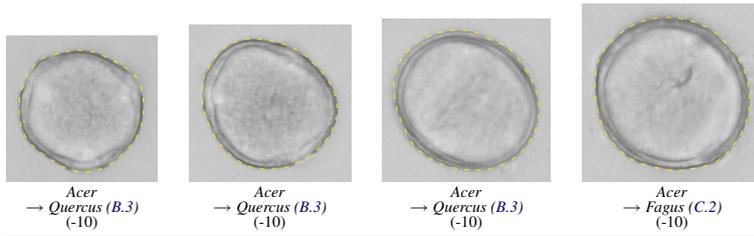
11 *Acer* in pollenmonitor test data set (Freiburg + Zürich) (0 in training data set)

0 correctly recognized as *Acer* (0%):

0 rejected (0%):

11 classified as other objects (100%):



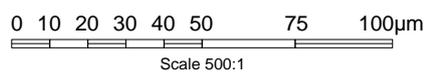


Other objects that were confused with *Acer*

(relative to 0 predicted *Acer*)

0 false positives (nan%):

0 unknown positives (nan%):



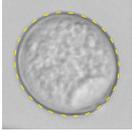
D.2 Humulus (Hopfen)

2 *Humulus* in pollenmonitor test data set (Freiburg + Zürich)

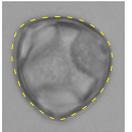
(0 in training data set)

0 correctly recognized as *Humulus* (0%):

1 rejected (50%):



1 classified as other objects (50%):



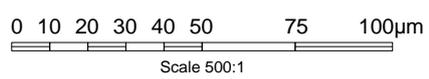
Humulus
→ Corylus (A.5)
(-10)

Other objects that were confused with *Humulus*

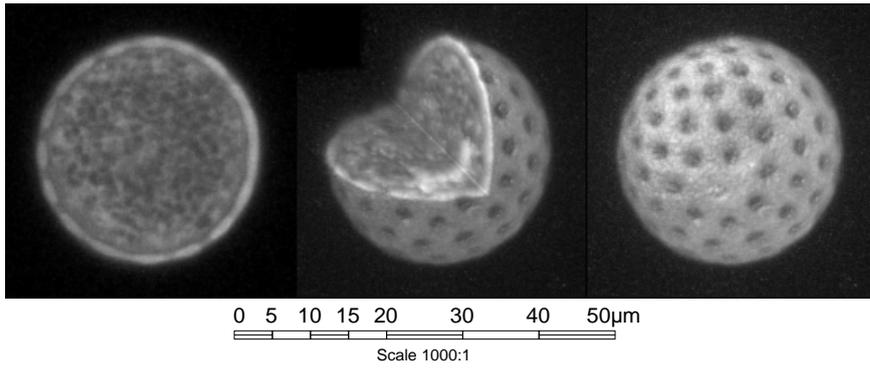
(relative to 0 predicted *Humulus*)

0 false positives (nan%):

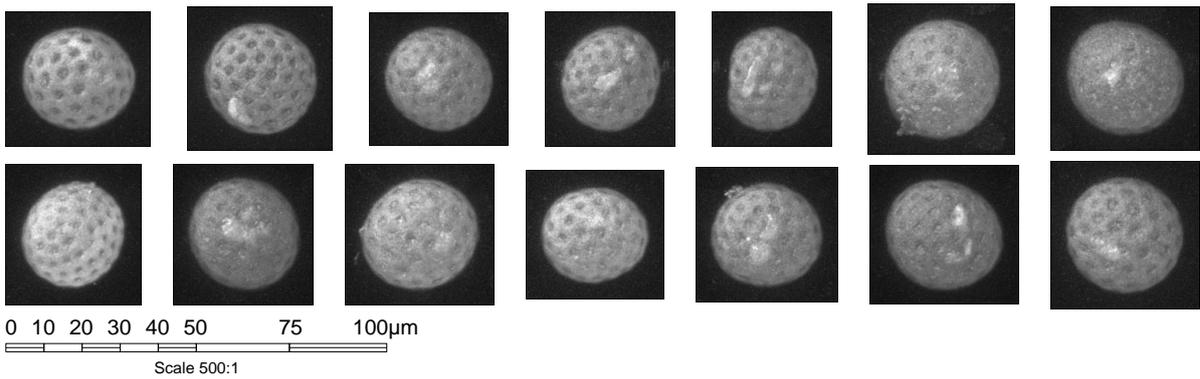
0 unknown positives (nan%):



D.3 Chenopodium (Gänsefuß)



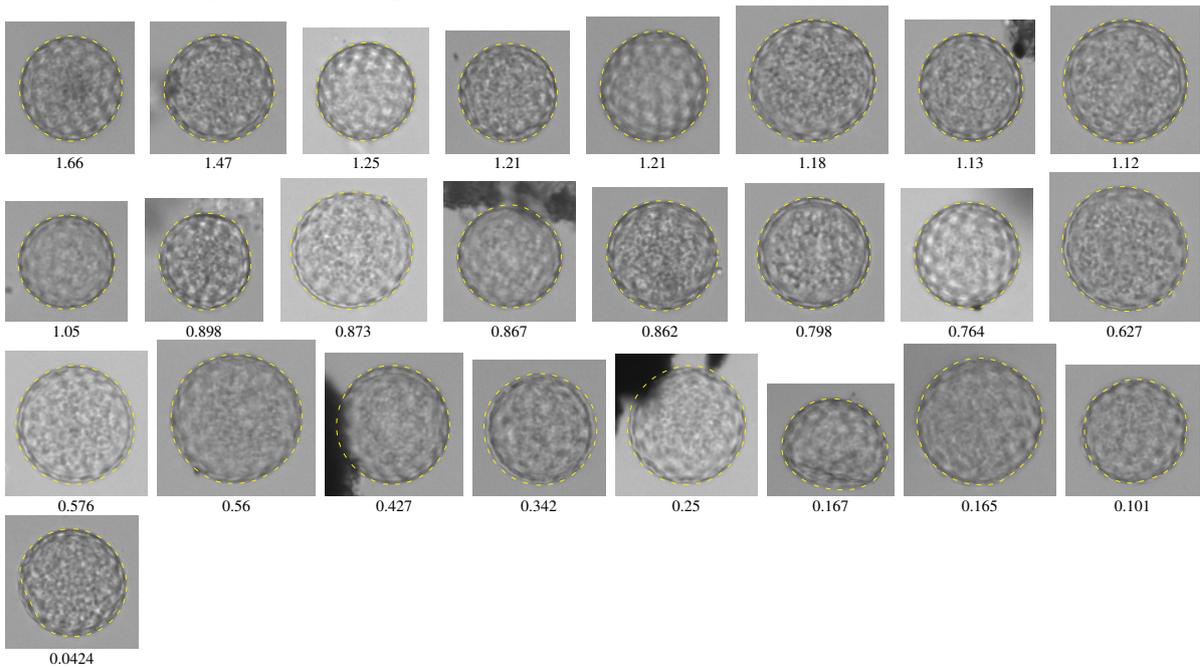
Confocal data set



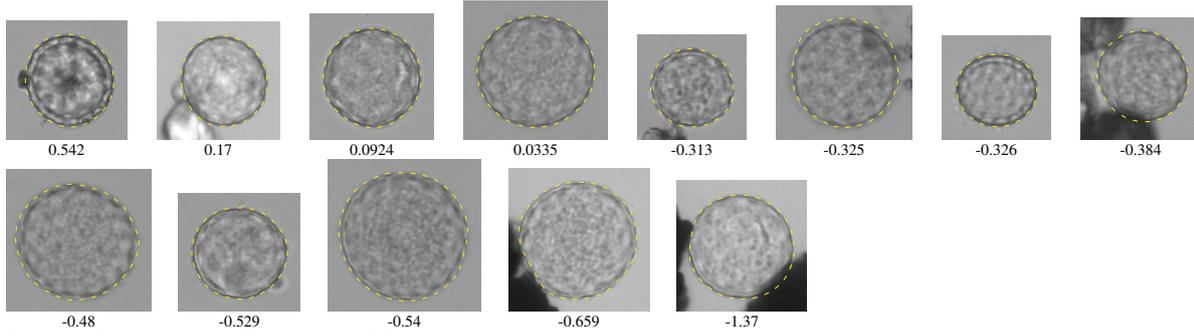
41 *Chenopodium* in pollenmonitor test data set (Freiburg + Zürich)

(17 in training data set)

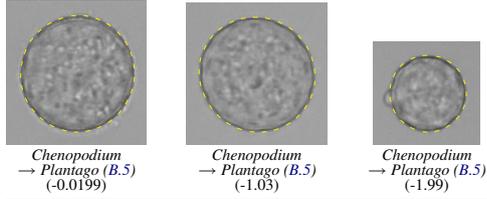
25 correctly recognized as *Chenopodium* (61%):



13 rejected (31.7%):

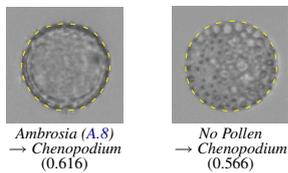


3 classified as other objects (7.32%):

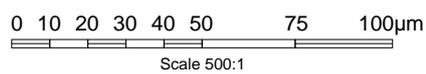
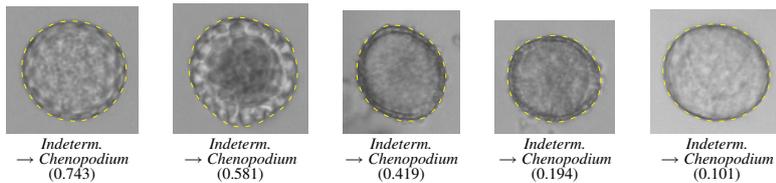


Other objects that were confused with *Chenopodium* (relative to 32 predicted *Chenopodium*)

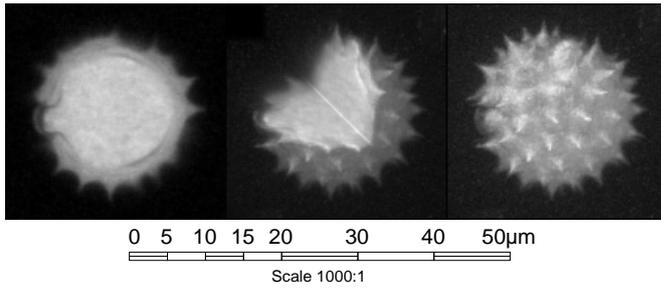
2 false positives (6.25%):



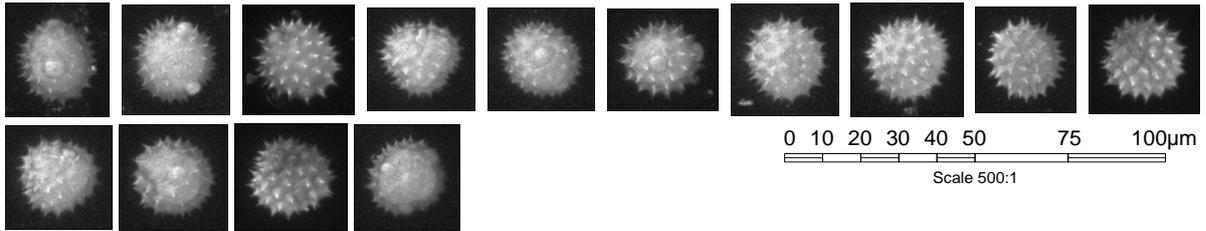
5 unknown positives (15.6%):



D.4 Compositae (Korbblüter)



Confocal data set



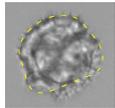
1 Compositae in pollenmonitor test data set (Freiburg + Zürich)

(0 in training data set)

0 correctly recognized as *Compositae* (0%):

0 rejected (0%):

1 classified as other objects (100%):



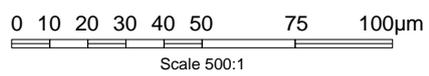
Compositae
→ *No Pollen*
(-10)

Other objects that were confused with *Compositae*

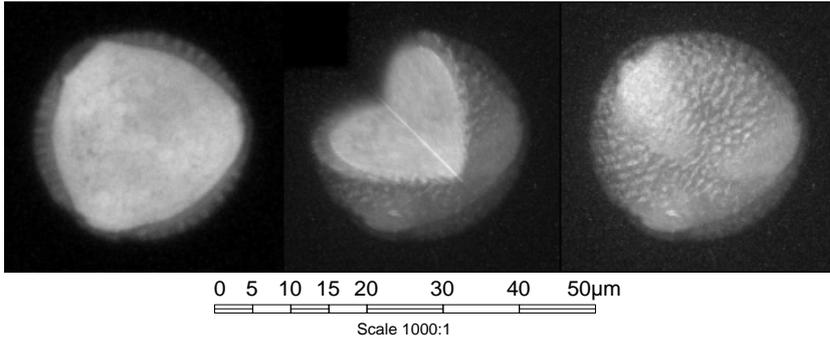
(relative to 0 predicted *Compositae*)

0 false positives (nan%):

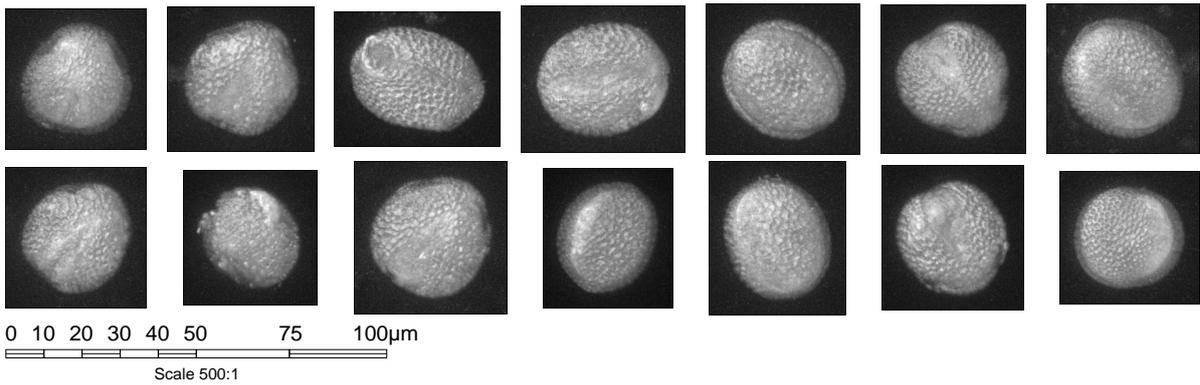
0 unknown positives (nan%):

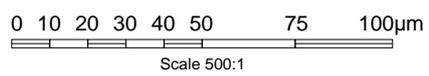


D.5 Cruciferae (Kreuzblütengewächse)

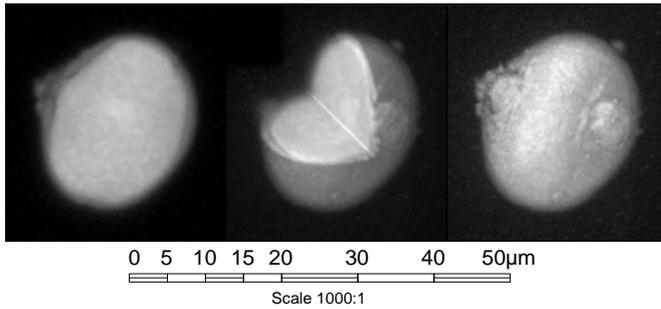


Confocal data set

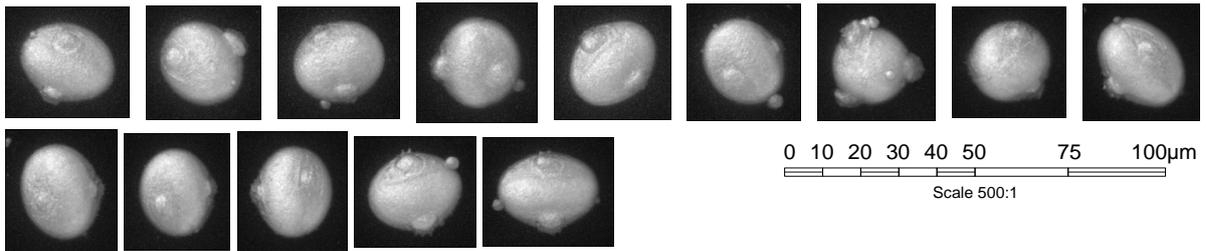




D.6 Aesculus (Roßkastanie)



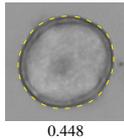
Confocal data set



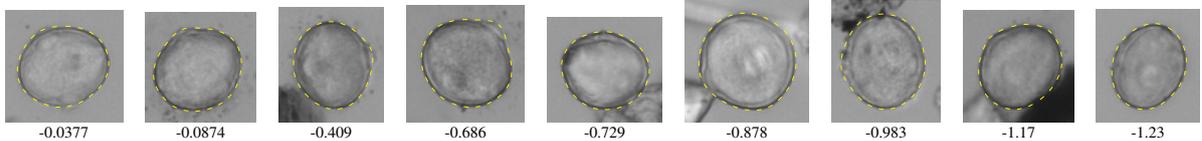
13 *Aesculus* in pollenmonitor test data set (Freiburg + Zürich)

(2 in training data set)

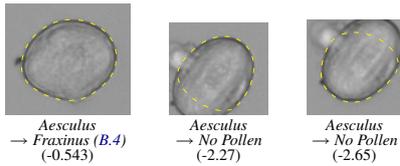
1 correctly recognized as *Aesculus* (7.69%):



9 rejected (69.2%):



3 classified as other objects (23.1%):

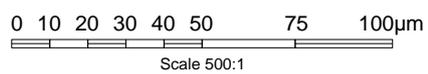


Other objects that were confused with *Aesculus*

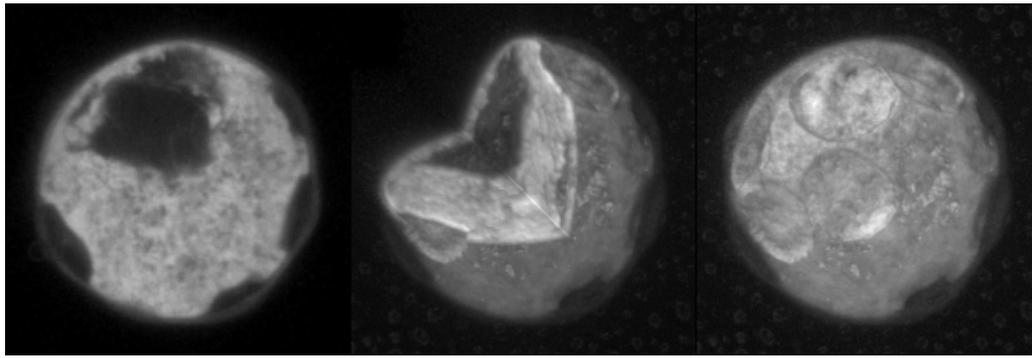
(relative to 1 predicted *Aesculus*)

0 false positives (0%):

0 unknown positives (0%):

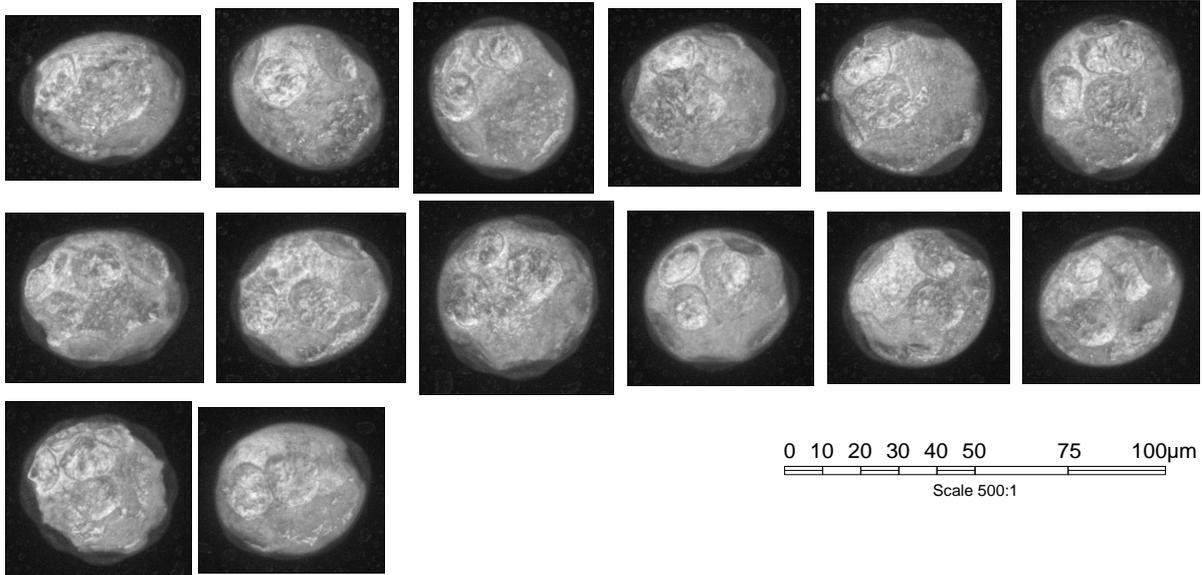


D.7 Juglans (Walnuß)



0 5 10 15 20 30 40 50µm
Scale 1000:1

Confocal data set

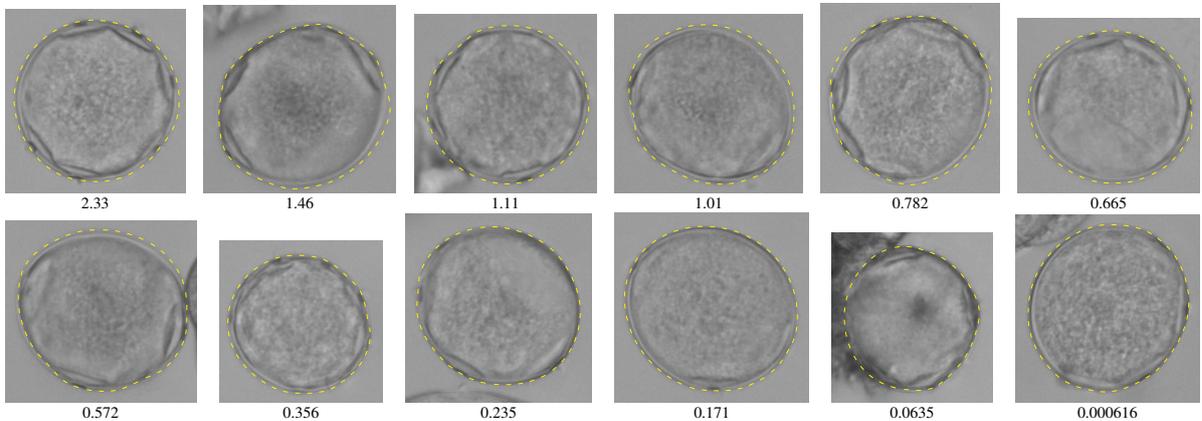


0 10 20 30 40 50 75 100µm
Scale 500:1

142 *Juglans* in pollenmonitor test data set (Freiburg + Zürich)

(25 in training data set)

72 correctly recognized as *Juglans* (50.7%):



2.33

1.46

1.11

1.01

0.782

0.665

0.572

0.356

0.235

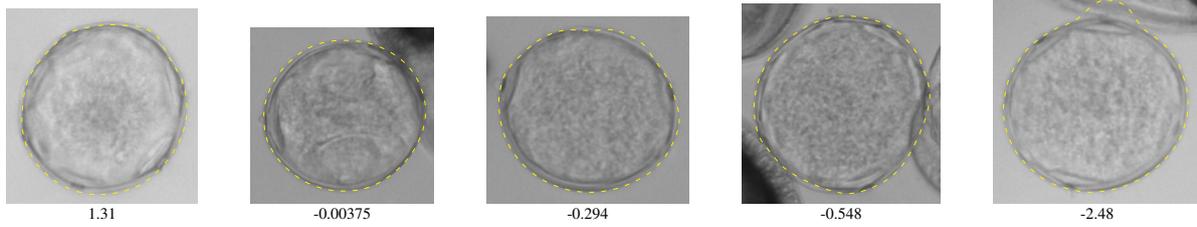
0.171

0.0635

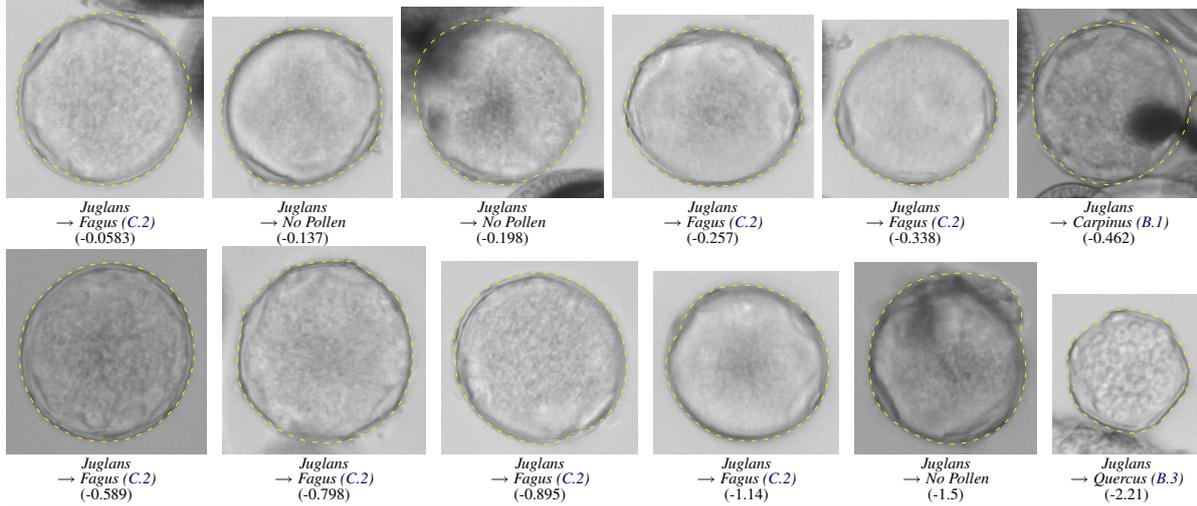
0.000616

0 10 20 30 40 50 75 100µm
Scale 500:1

48 rejected (33.8%):



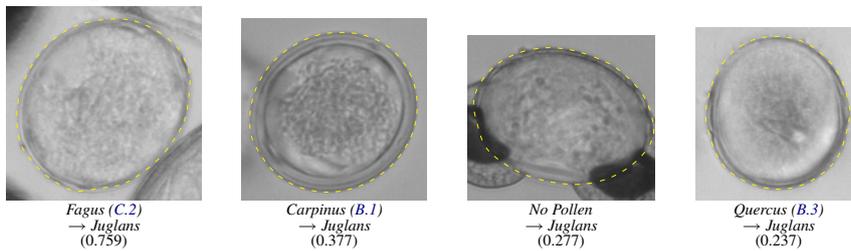
22 classified as other objects (15.5%):



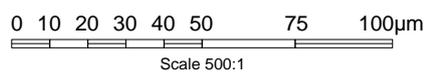
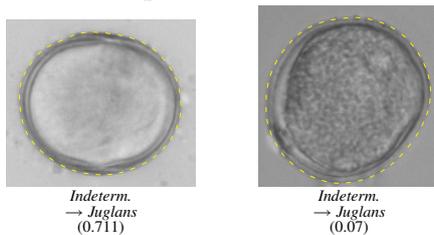
Other objects that were confused with *Juglans*

(relative to 78 predicted *Juglans*)

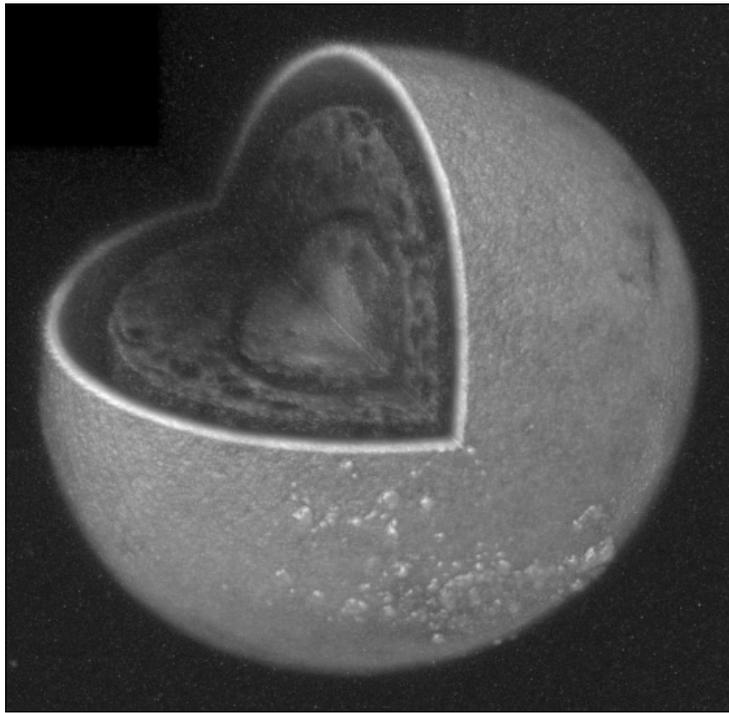
4 false positives (5.13%):



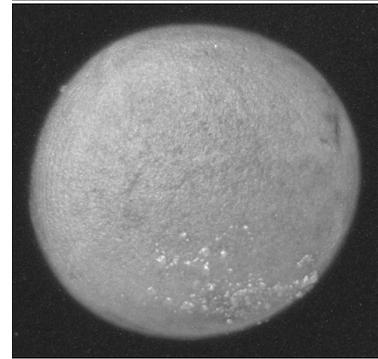
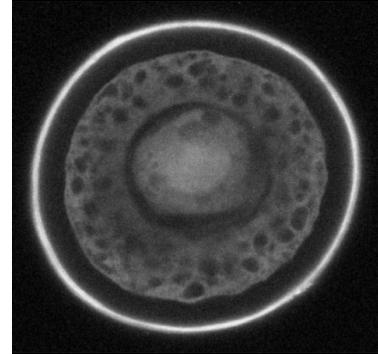
2 unknown positives (2.56%):



D.8 Larix (Lärche)



0 5 10 15 20 30 40 50µm
Scale 1000:1

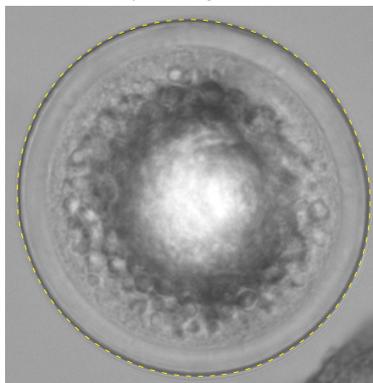


0 10 20 30 40 50 75 100µm
Scale 500:1

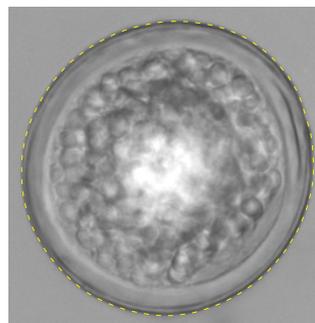
29 Larix in pollenmonitor test data set (Freiburg + Zürich)

(3 in training data set)

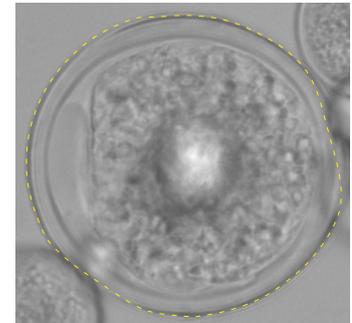
13 correctly recognized as *Larix* (44.8%):



2.64



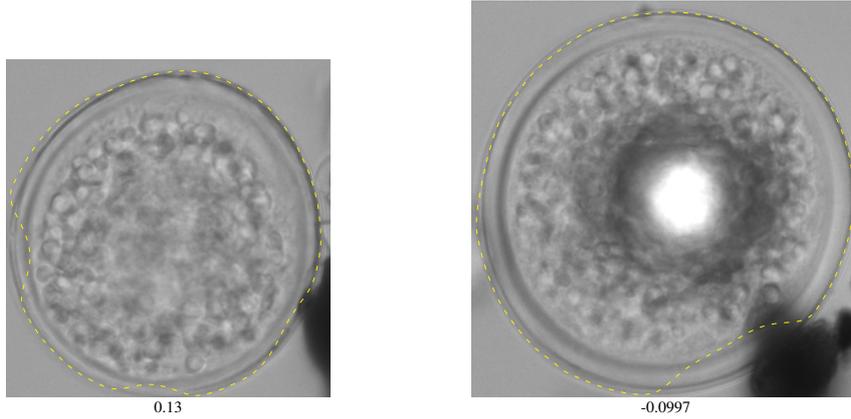
0.598



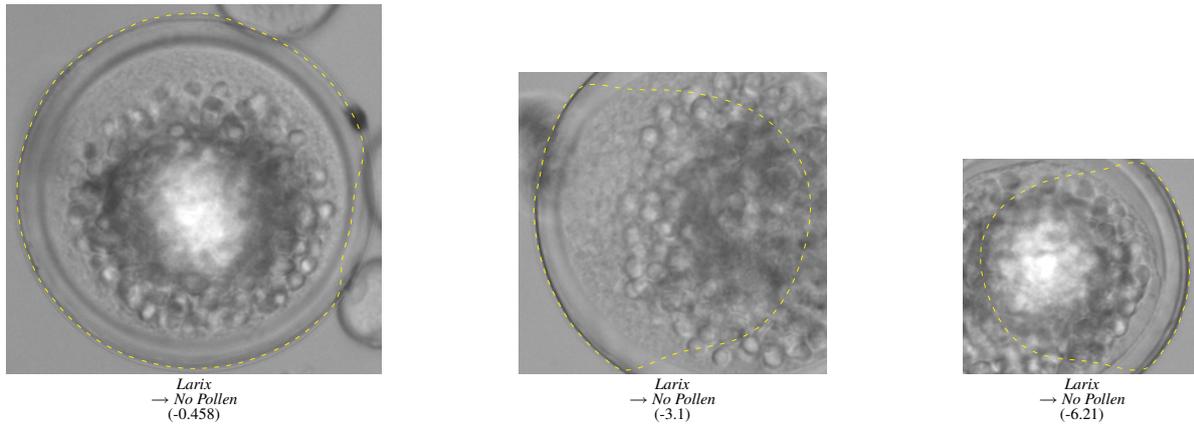
0.0122

0 10 20 30 40 50 75 100µm
Scale 500:1

2 rejected (6.9%):



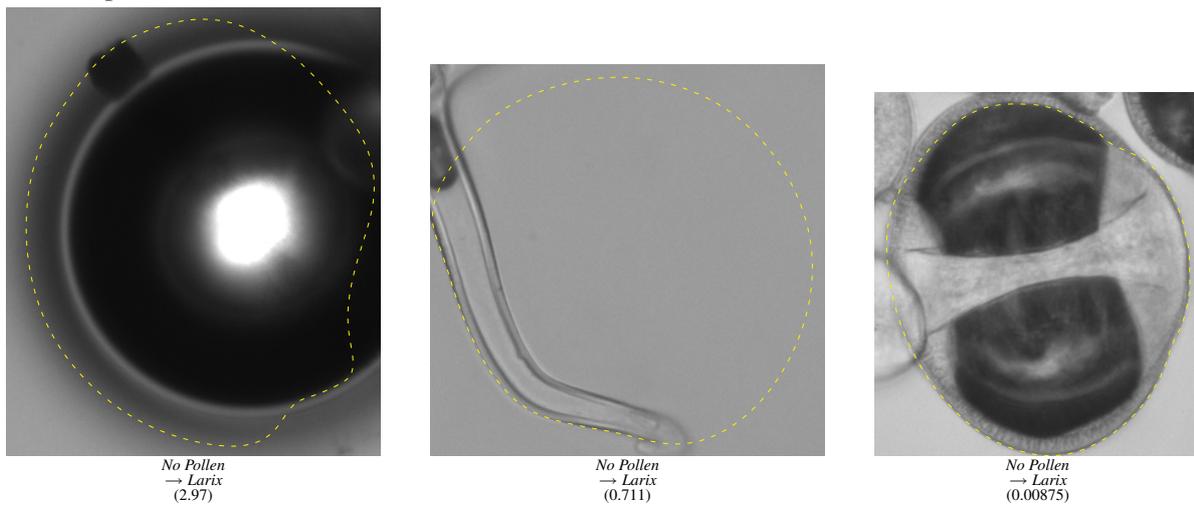
14 classified as other objects (48.3%):



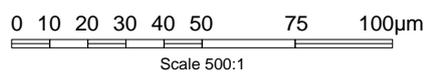
Other objects that were confused with *Larix*

(relative to 57 predicted *Larix*)

44 false positives (77.2%):



0 unknown positives (0%):



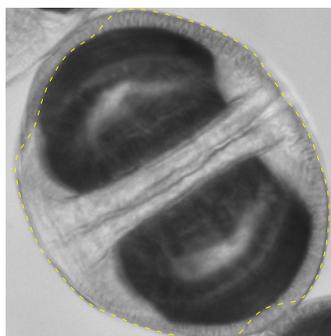
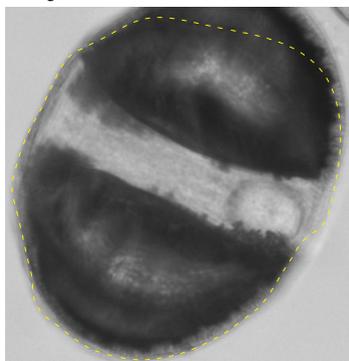
D.9 Picea (Fichte)

40 *Picea* in pollenmonitor test data set (Freiburg + Zürich)

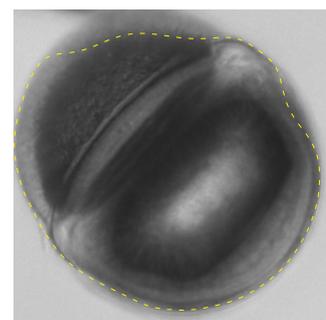
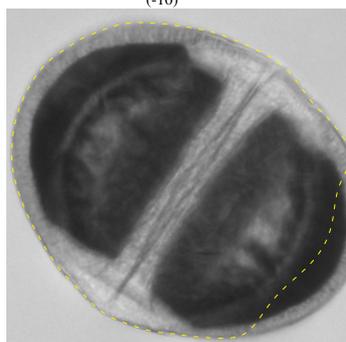
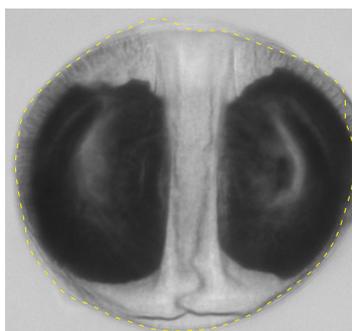
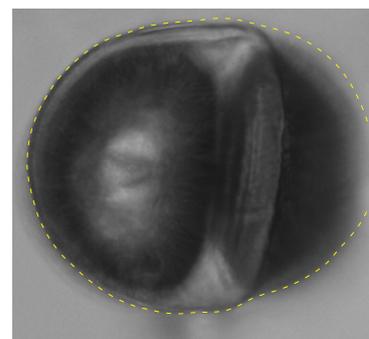
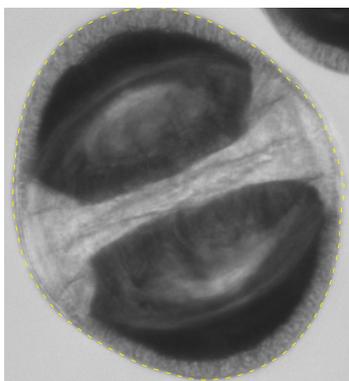
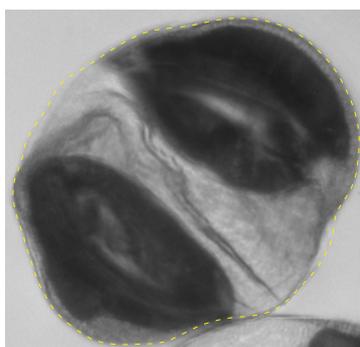
(0 in training data set)

0 correctly recognized as *Picea* (0%):

2 rejected (5%):



38 classified as other objects (95%):

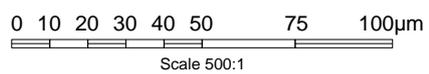


Other objects that were confused with *Picea*

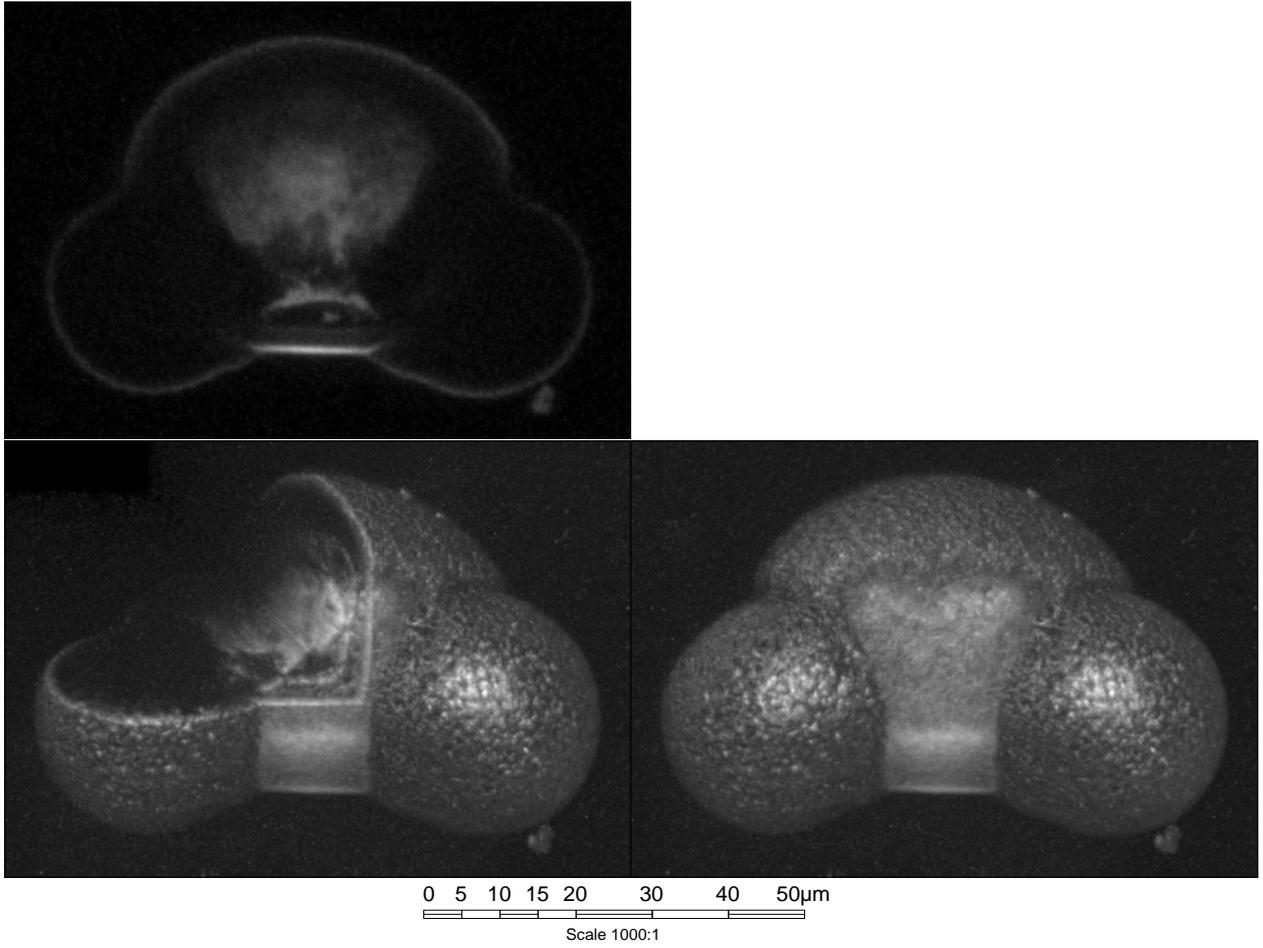
(relative to 0 predicted *Picea*)

0 false positives (nan%):

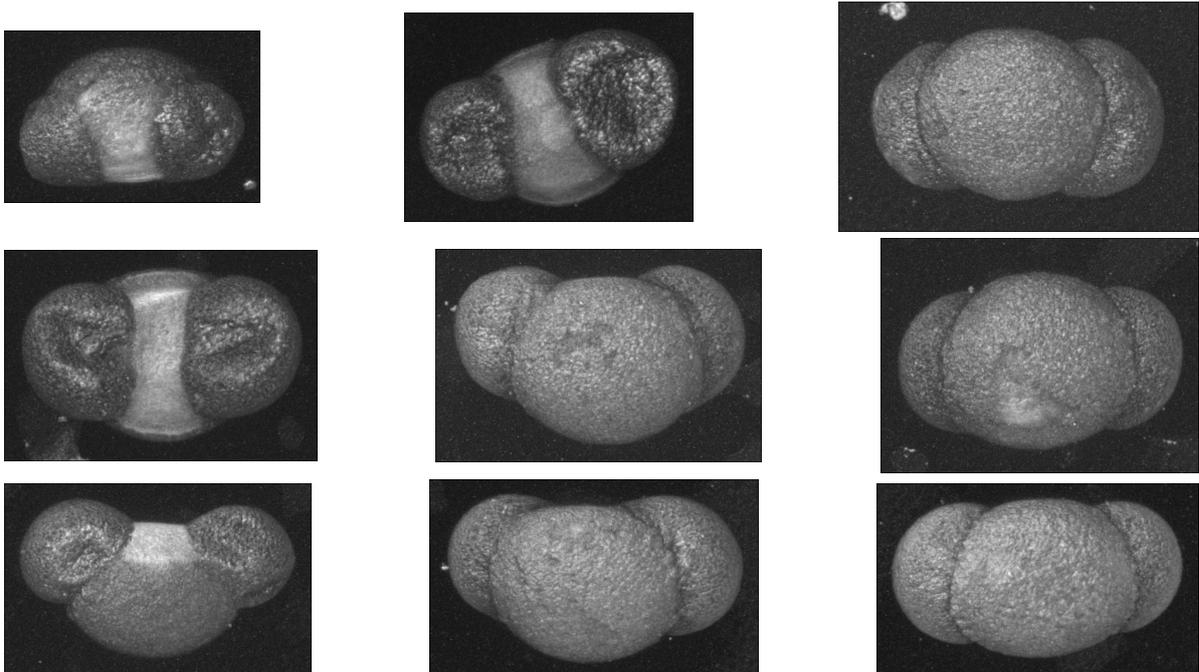
0 unknown positives (nan%):

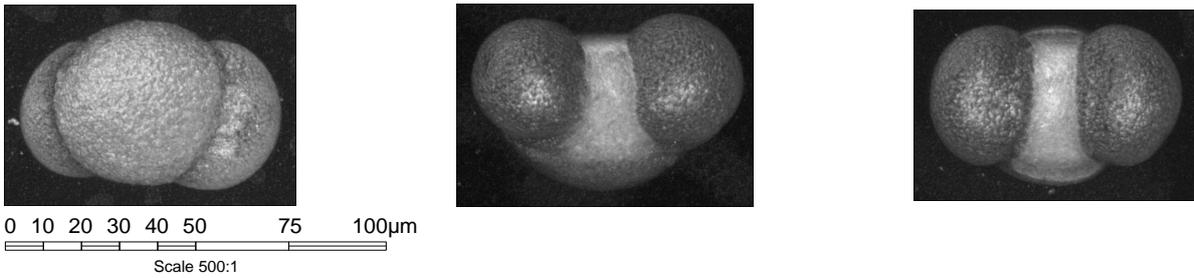


D.10 Pinus (Kiefer)



Confocal data set





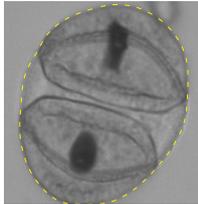
1 *Pinus* in pollenmonitor test data set (Freiburg + Zürich)

(0 in training data set)

0 correctly recognized as *Pinus* (0%):

0 rejected (0%):

1 classified as other objects (100%):



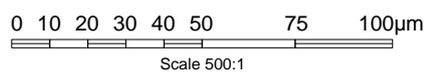
Pinus
→ No Pollen
(-10)

Other objects that were confused with *Pinus*

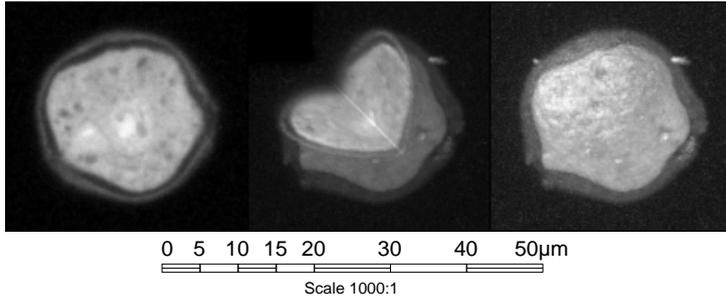
(relative to 0 predicted *Pinus*)

0 false positives (nan%):

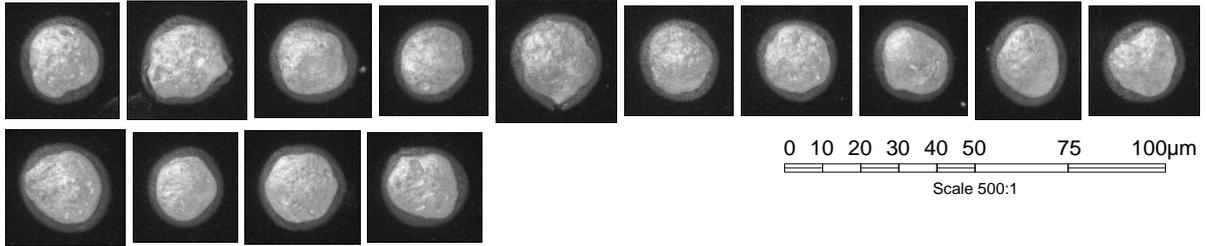
0 unknown positives (nan%):



D.11 Platanus (Platane)



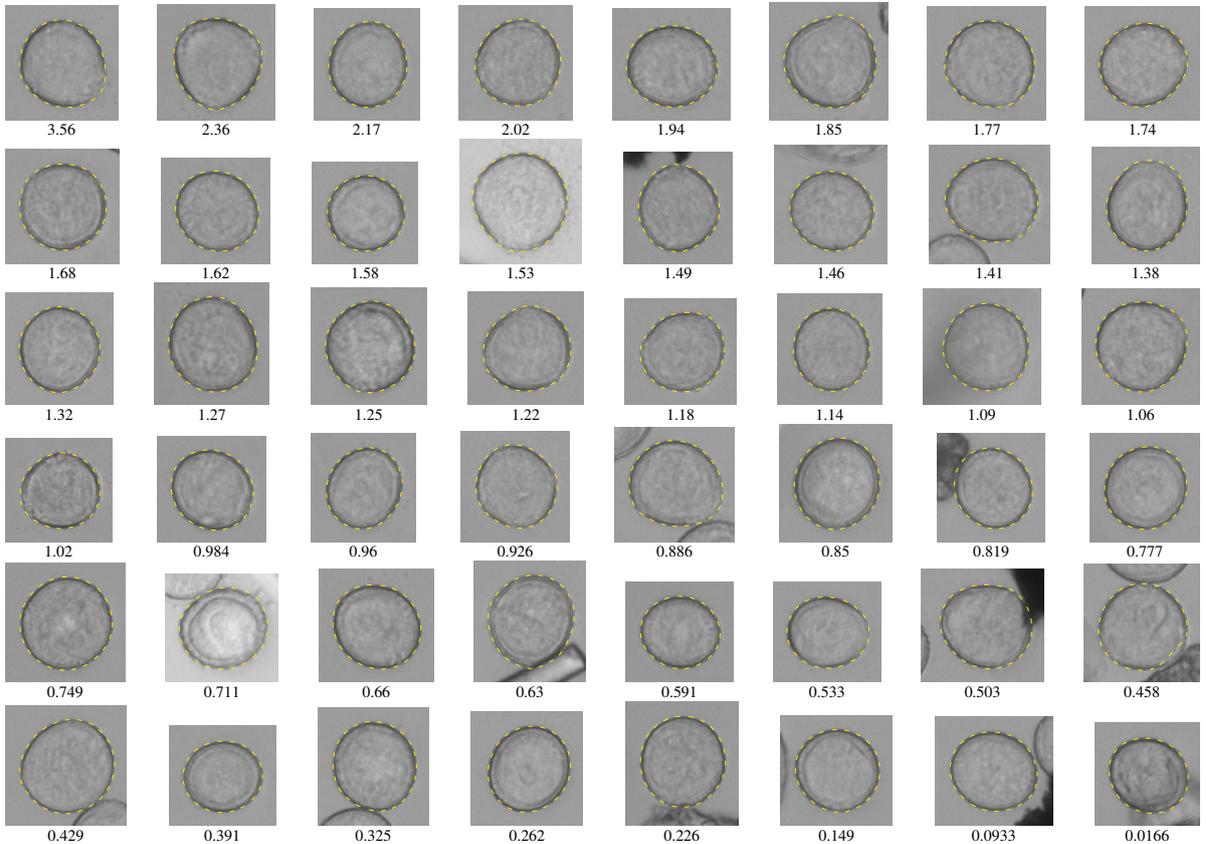
Confocal data set



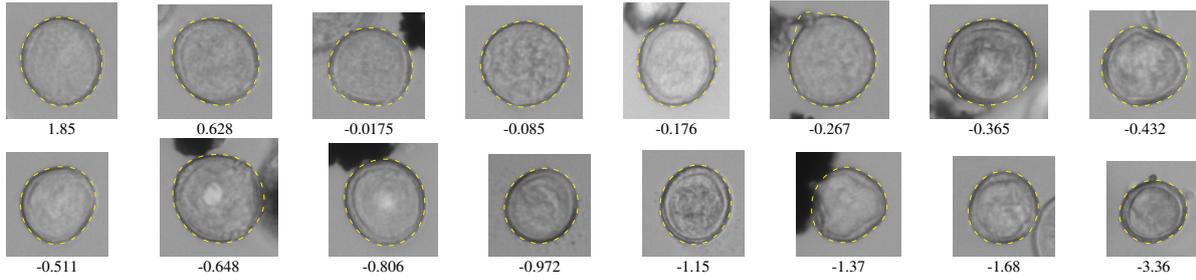
907 *Platanus* in pollenmonitor test data set (Freiburg + Zürich)

(161 in training data set)

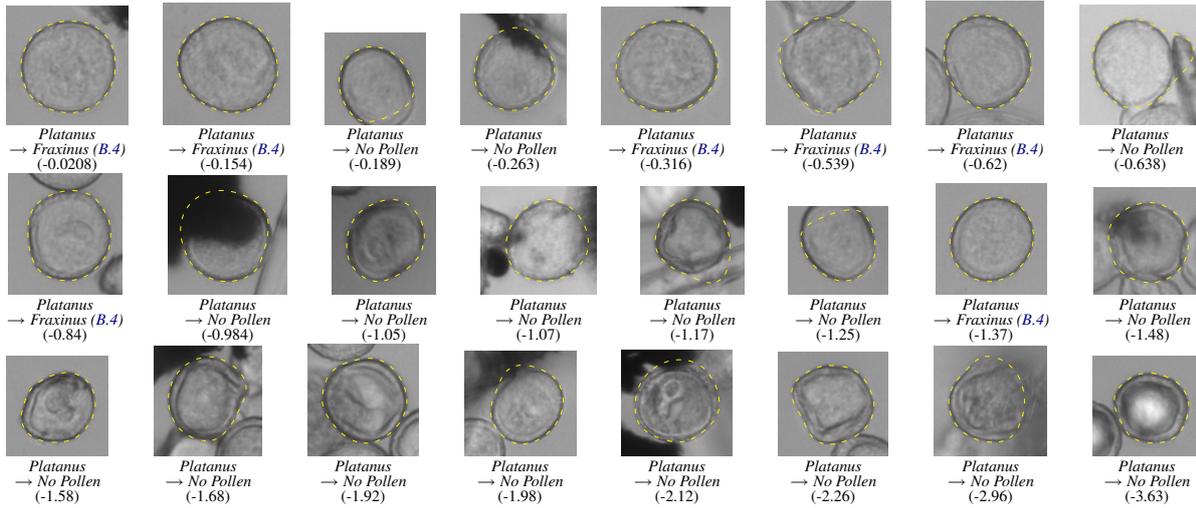
633 correctly recognized as *Platanus* (69.8%):



239 rejected (26.4%):



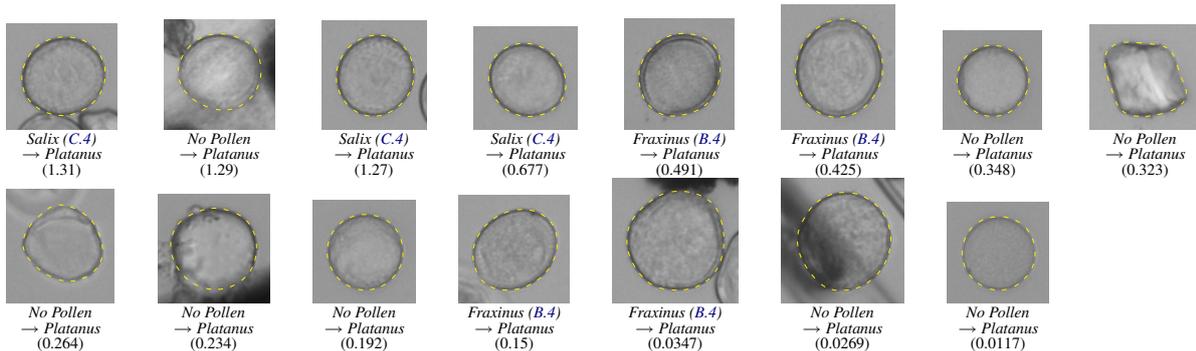
35 classified as other objects (3.86%):



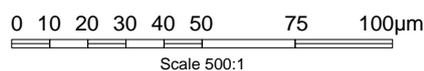
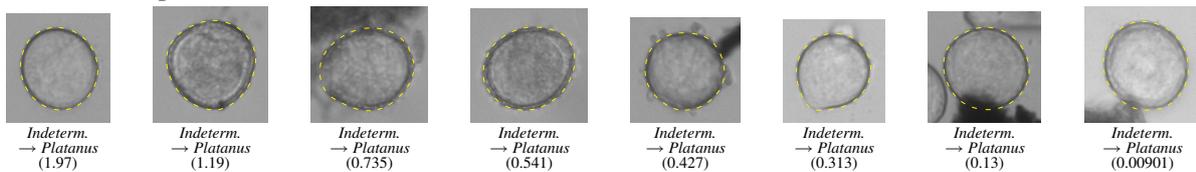
Other objects that were confused with *Platanus*

(relative to 708 predicted *Platanus*)

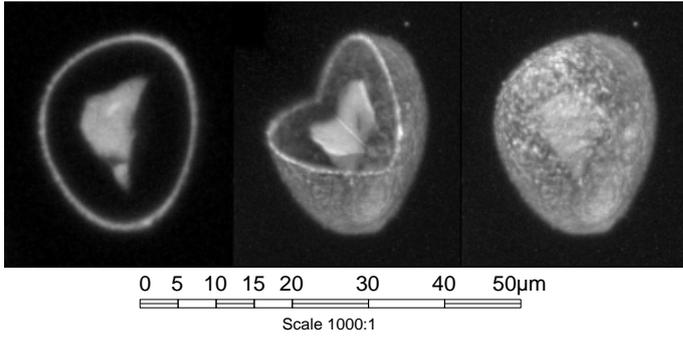
15 false positives (2.12%):



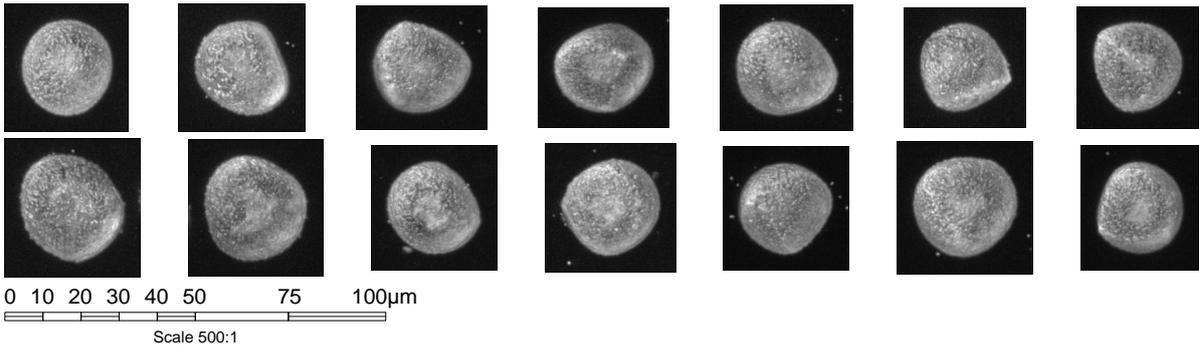
60 unknown positives (8.47%):



D.12 Taxus (Eibe)



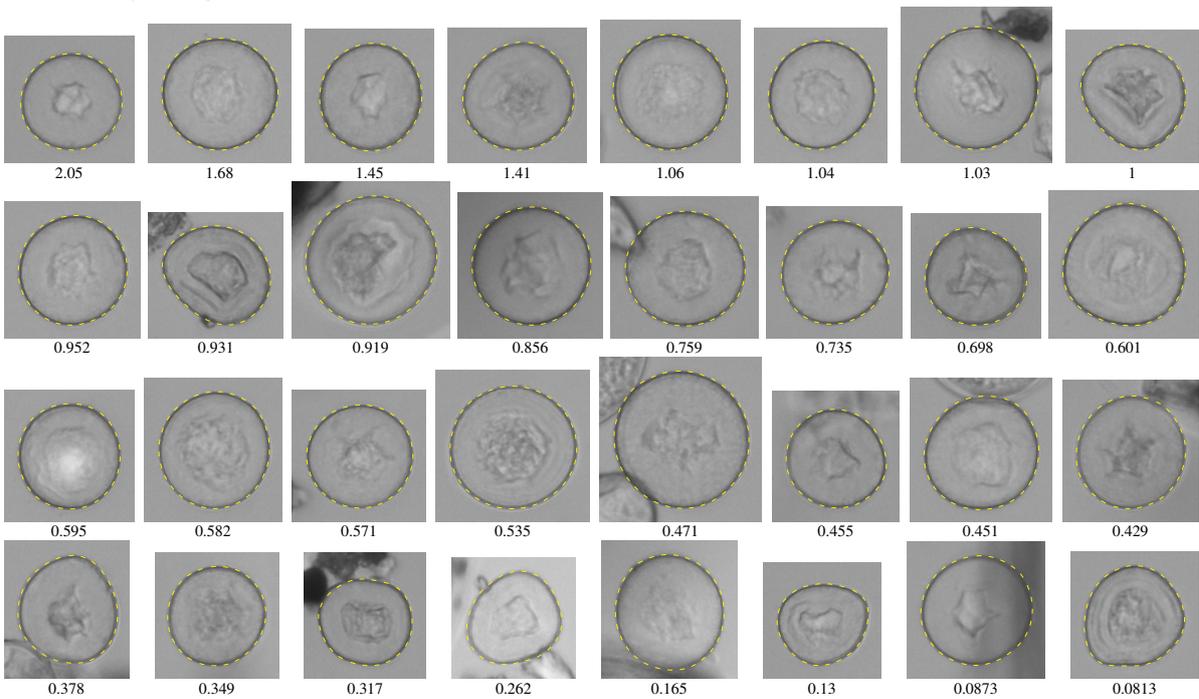
Confocal data set



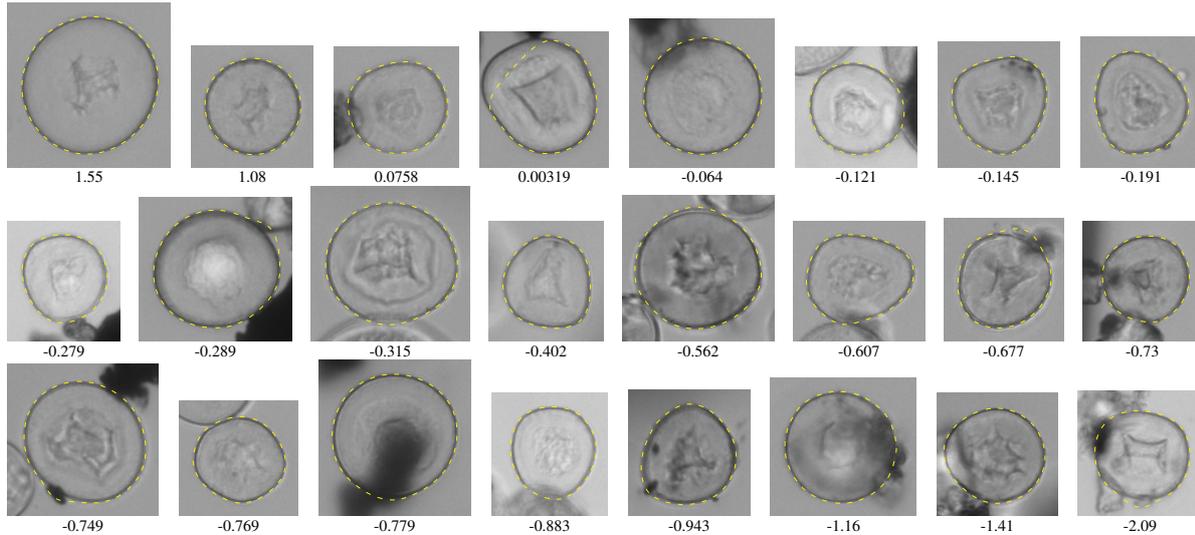
78 *Taxus* in pollenmonitor test data set (Freiburg + Zürich)

(27 in training data set)

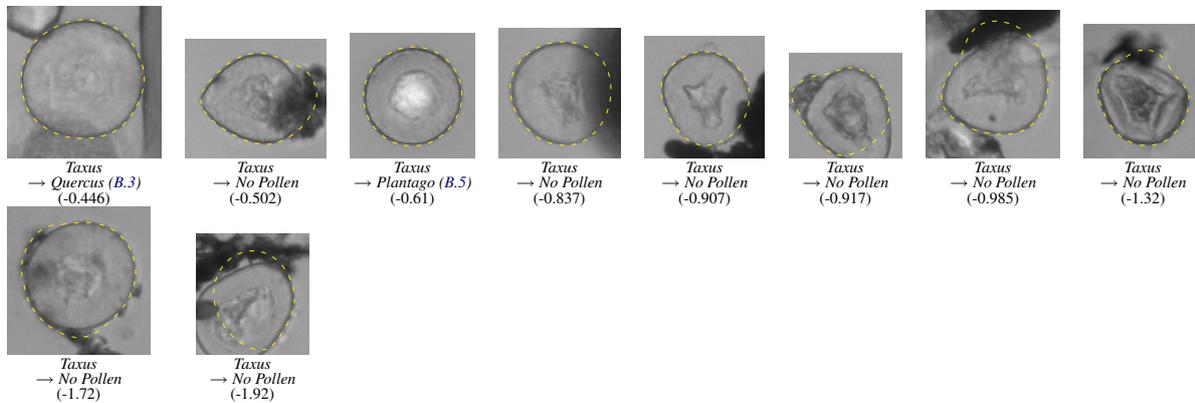
36 correctly recognized as *Taxus* (46.2%):



32 rejected (41%):



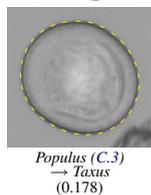
10 classified as other objects (12.8%):



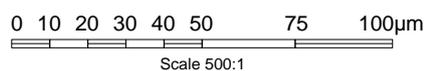
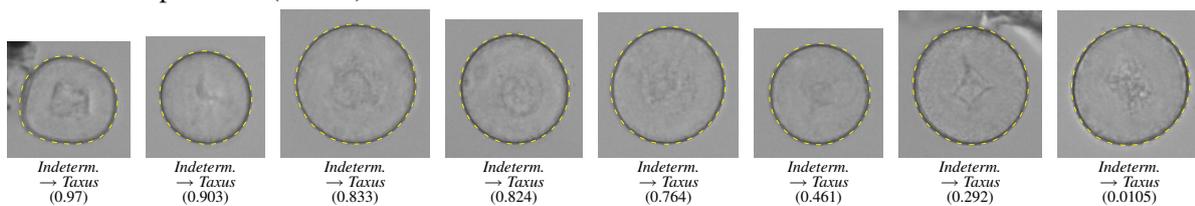
Other objects that were confused with *Taxus*

(relative to 47 predicted *Taxus*)

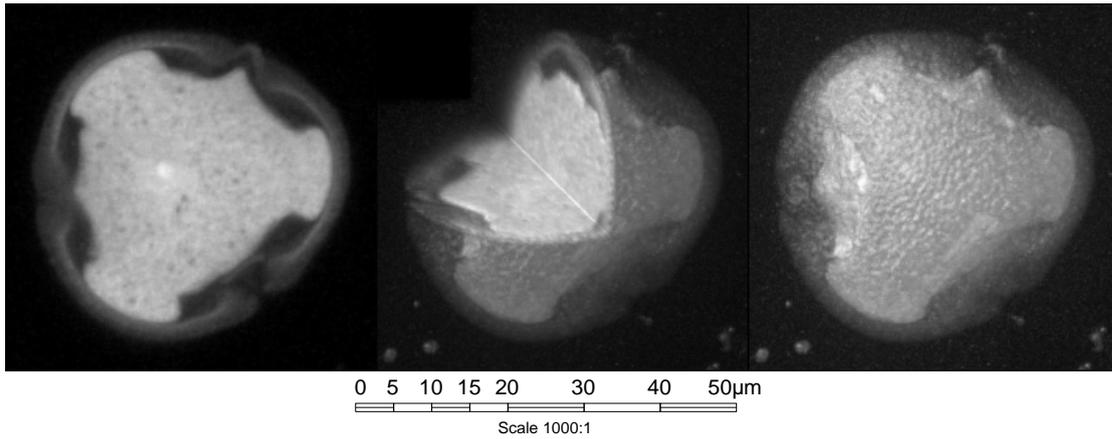
1 false positives (2.13%):



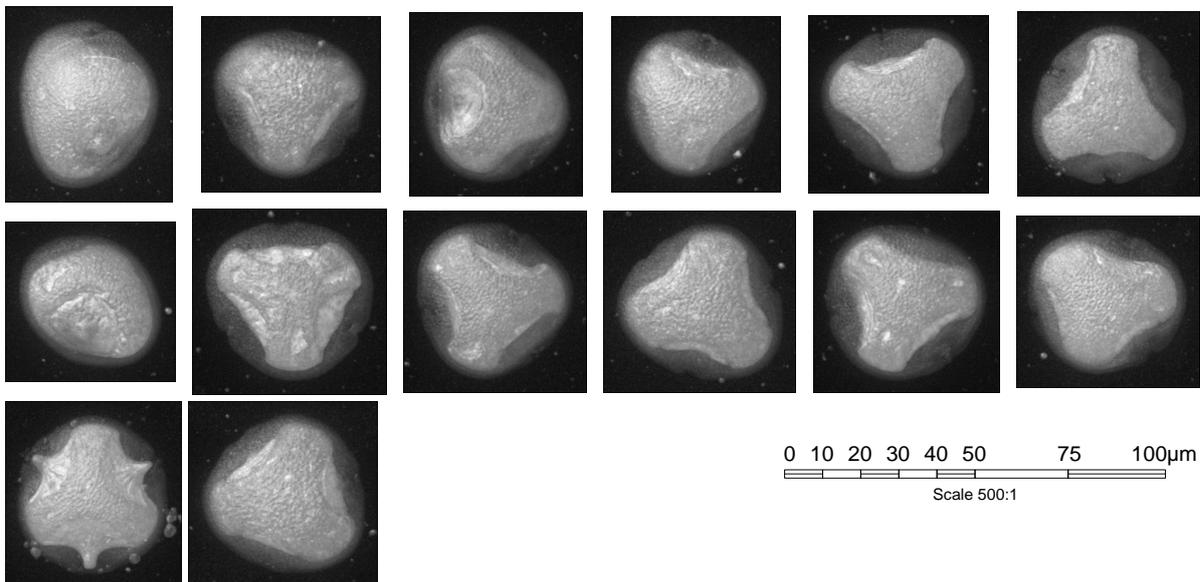
10 unknown positives (21.3%):



D.13 Tilia (Linde)



Confocal data set

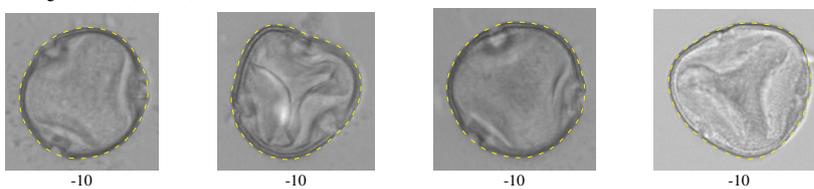


5 *Tilia* in pollenmonitor test data set (Freiburg + Zürich)

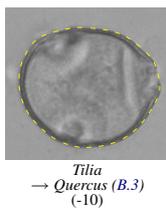
(0 in training data set)

0 correctly recognized as *Tilia* (0%):

4 rejected (80%):



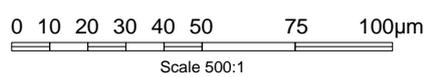
1 classified as other objects (20%):



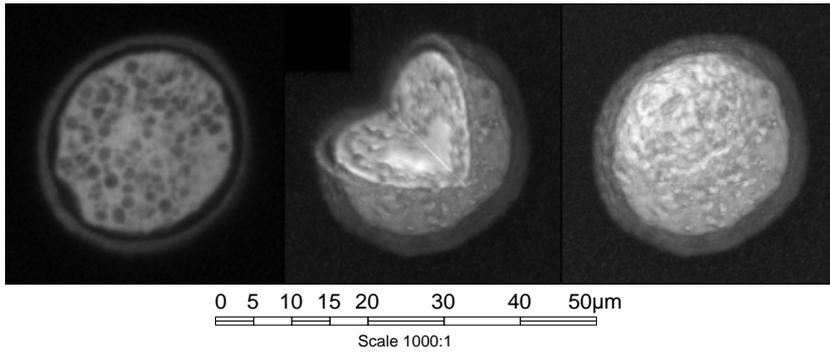
Other objects that were confused with *Tilia* (relative to 0 predicted *Tilia*)

0 false positives (nan%):

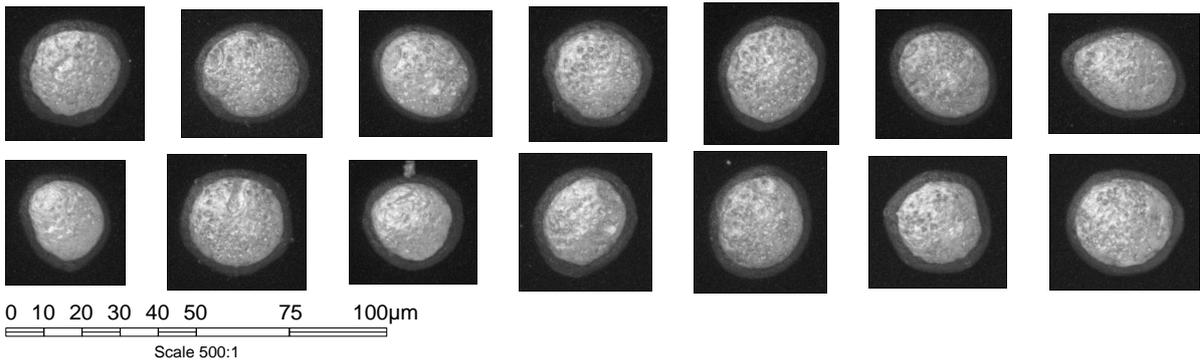
0 unknown positives (nan%):



D.14 Ulmus (Ulme)



Confocal data set

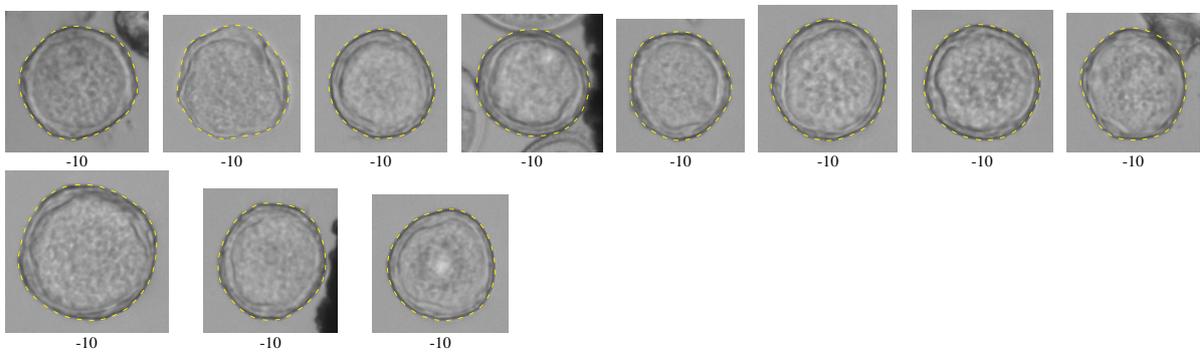


12 *Ulmus* in pollenmonitor test data set (Freiburg + Zürich)

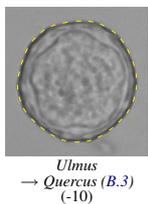
(0 in training data set)

0 correctly recognized as *Ulmus* (0%):

11 rejected (91.7%):



1 classified as other objects (8.33%):

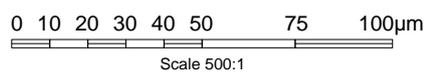


Other objects that were confused with *Ulmus*

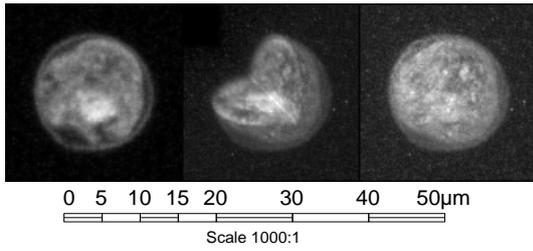
(relative to 0 predicted *Ulmus*)

0 false positives (nan%):

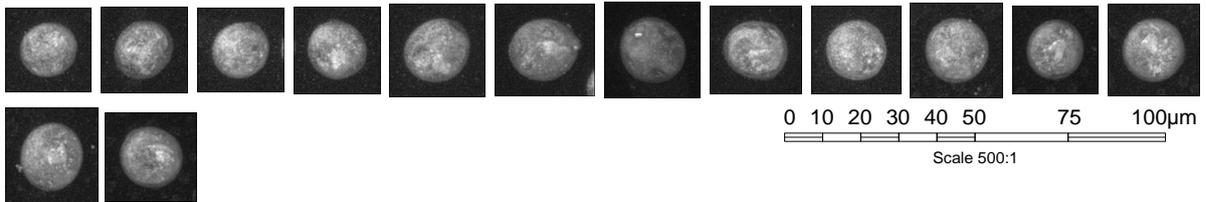
0 unknown positives (nan%):



D.15 *Urtica* (Brennessel)



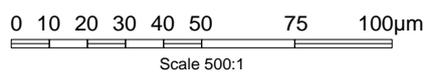
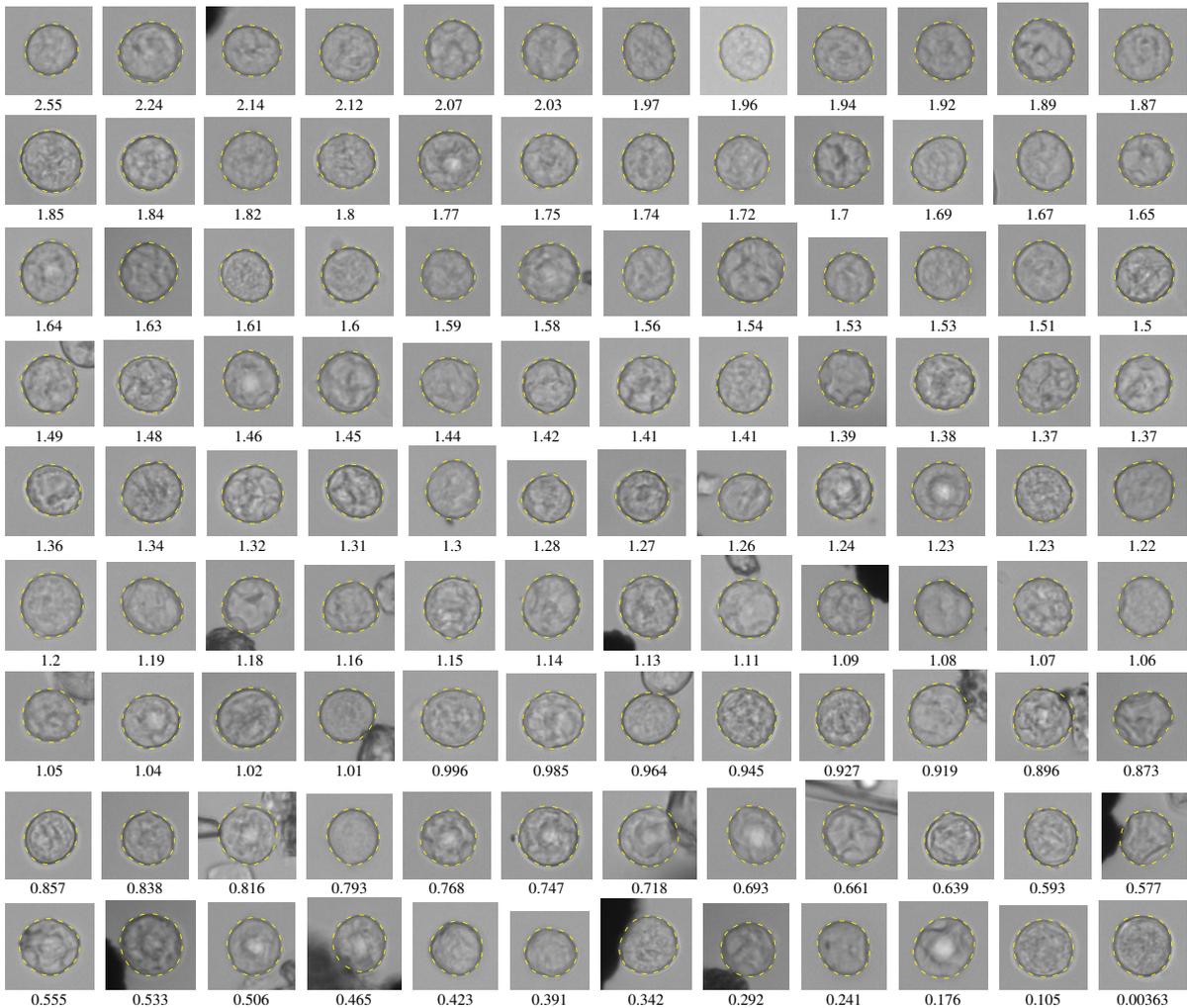
Confocal data set



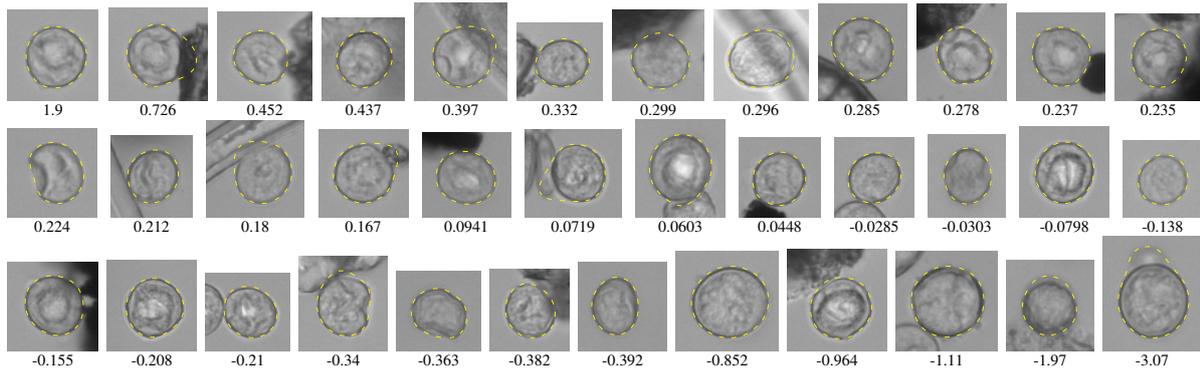
1094 *Urtica* in pollenmonitor test data set (Freiburg + Zürich)

(165 in training data set)

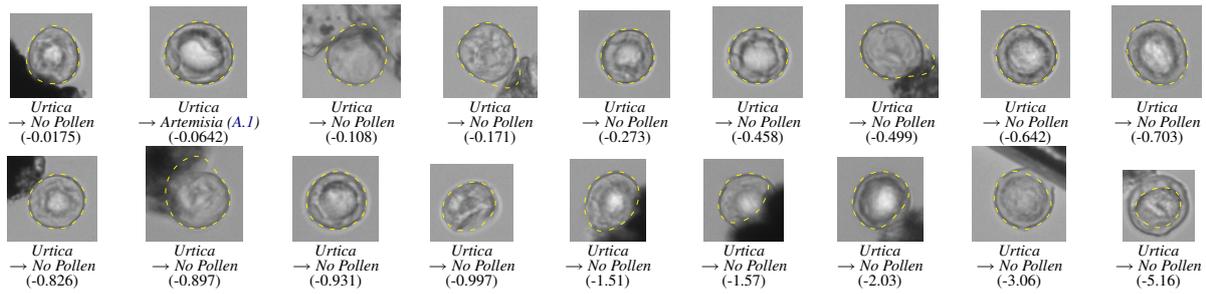
991 correctly recognized as *Urtica* (90.6%):



46 rejected (4.2%):



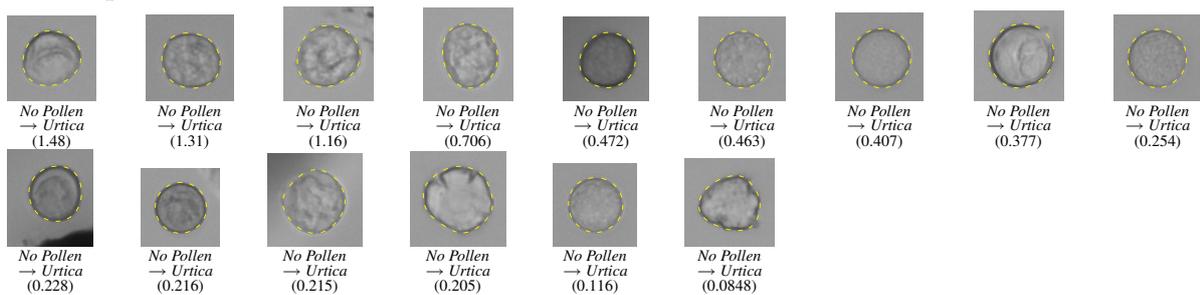
57 classified as other objects (5.21%):



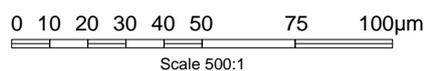
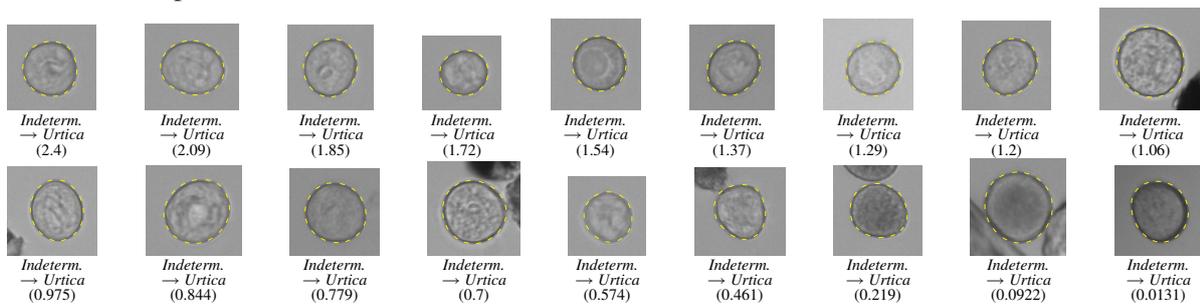
Other objects that were confused with *Urtica*

(relative to 1125 predicted *Urtica*)

15 false positives (1.33%):



119 unknown positives (10.6%):



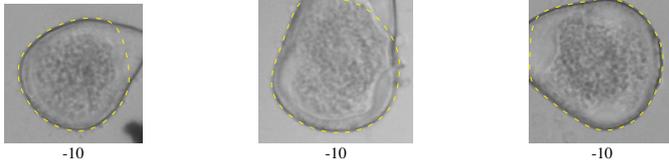
D.16 Cyperaceae (Sauergräser)

5 *Cyperaceae* in pollenmonitor test data set (Freiburg + Zürich)

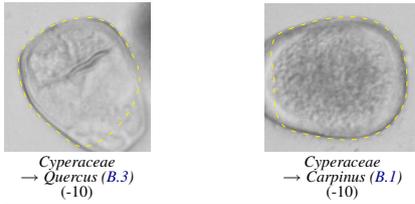
(0 in training data set)

0 correctly recognized as *Cyperaceae* (0%):

3 rejected (60%):



2 classified as other objects (40%):

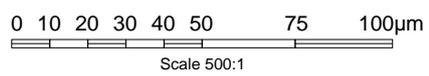


Other objects that were confused with *Cyperaceae*

(relative to 0 predicted *Cyperaceae*)

0 false positives (nan%):

0 unknown positives (nan%):

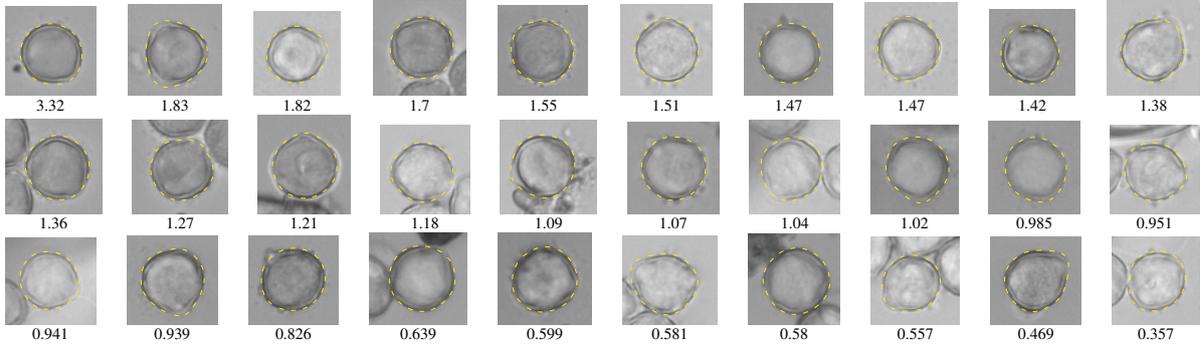


D.17 Sambucus (Holunder)

39 Sambucus in pollenmonitor test data set (Freiburg + Zürich)

(63 in training data set)

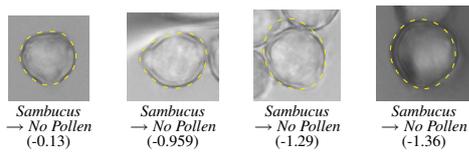
30 correctly recognized as *Sambucus* (76.9%):



5 rejected (12.8%):



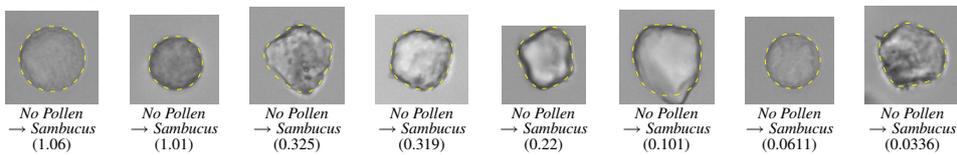
4 classified as other objects (10.3%):



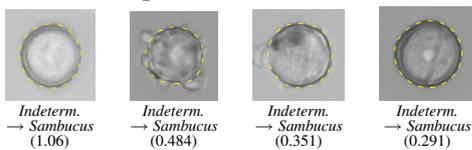
Other objects that were confused with *Sambucus*

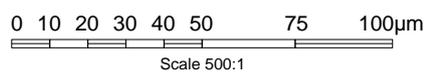
(relative to 42 predicted *Sambucus*)

8 false positives (19%):



4 unknown positives (9.52%):





E Categorization of Pollen Taxa into Different Allergenic Levels

In Germany the allergenic impact of the several pollen taxa is described by two categories “allergenic” and “non-allergenic”. The allergenic category contains the six taxa that are reported in the daily pollen forecasts: *Corylus*^(A.5), *Alnus*^(A.2), *Betula*^(A.4), *Poaceae*^(A.6), *Secale*^(A.7), *Artemisia*^(A.1).

In Switzerland four allergy levels (high, moderate, low, none) are used. The highly-allergenic category contains the same 6 taxa as in Germany and additionally the *Fraxinus*^(B.4).

Within the last years a new highly allergenic pollen taxa became important. Due to the global warming the *Ambrosia*^(A.8) plant expands to the north and is now present in parts Switzerland and the southern regions of Germany.

In this thesis a mixture of these two categorizations is used to obtain the best possible compatibility: We use the swiss categories, but move the *Fraxinus*^(B.4) to the moderately allergenic pollen taxa (see table E.1

pollen taxon	categorization in this thesis	categorization in Switzerland	categorization in Germany
<i>Artemisia</i> ^(A.1)	high	high	allergenic
<i>Alnus</i> ^(A.2)	high	high	allergenic
<i>Alnus viridis</i> ^(A.3)	high	high	allergenic
<i>Betula</i> ^(A.4)	high	high	allergenic
<i>Corylus</i> ^(A.5)	high	high	allergenic
<i>Poaceae</i> ^(A.6)	high	high	allergenic
<i>Secale</i> ^(A.7)	high	high	allergenic
<i>Ambrosia</i> ^(A.8)	high	high	allergenic
<i>Fraxinus</i> ^(B.4)	moderate	high	non-allergenic
<i>Carpinus</i> ^(B.1)	moderate	moderate	non-allergenic
<i>Ostrya Carpinifolia</i> ^(B.2)	moderate	moderate	non-allergenic
<i>Quercus</i> ^(B.3)	moderate	moderate	non-allergenic
<i>Plantago</i> ^(B.5)	moderate	moderate	non-allergenic
<i>Rumex</i> ^(B.6)	moderate	moderate	non-allergenic
<i>Fagus</i> ^(C.2)	low	low–moderate	non-allergenic
<i>Castanea</i> ^(C.1)	low	low	non-allergenic
<i>Populus</i> ^(C.3)	low	low	non-allergenic
<i>Salix</i> ^(C.4)	low	low	non-allergenic
<i>Acer</i> ^(D.1)	none	none	non-allergenic
<i>Humulus</i> ^(D.2)	none	none	non-allergenic
<i>Chenopodium</i> ^(D.3)	none	none	non-allergenic
<i>Compositae</i> ^(D.4)	none	none	non-allergenic
<i>Cruciferae</i> ^(D.5)	none	none	non-allergenic
<i>Aesculus</i> ^(D.6)	none	none	non-allergenic
<i>Juglans</i> ^(D.7)	none	none	non-allergenic
<i>Larix</i> ^(D.8)	none	none	non-allergenic
<i>Picea</i> ^(D.9)	none	none	non-allergenic
<i>Pinus</i> ^(D.10)	none	none	non-allergenic
<i>Platanus</i> ^(D.11)	none	none	non-allergenic
<i>Taxus</i> ^(D.12)	none	none	non-allergenic
<i>Tilia</i> ^(D.13)	none	none	non-allergenic
<i>Ulmus</i> ^(D.14)	none	none	non-allergenic
<i>Urtica</i> ^(D.15)	none	none	non-allergenic
<i>Cyperaceae</i> ^(D.16)	none	none	non-allergenic
<i>Sambucus</i> ^(D.17)	none	none	non-allergenic

Table E.1: The allergenic levels for the different pollen taxa

Bibliography

- A. d. Albuquerque Araujo, L. Claudino, and R. Augusto Rabelo Oliveira. Pollen grain information retrieval by neural representation. In *Computer Graphics and Image Processing, 2001 Proceedings of XIV Brazilian Symposium on*, pages 398–, 2001a.
- A. d. Albuquerque Araujo, L. Perroton, R. Oliveira, L. Claudino, S. Guimarães, and E. Bastos. Non-linear features extraction applied to pollen grain images. In *Technical Summary Digest of the IS&T/SPIE Electronic Imaging Conference 4304: Non Linear Image Processing and Pattern Analysis XII*, pages 175–176., San Jose, CA, USA, 2001b. doi:10.1117/12.424990.
- G. P. Allen, R. M. Hodgson, S. R. Marsland, G. Arnold, R. C. Flemmer, J. Flenley, and D. W. Fountain. Automatic recognition of light-microscope pollen images. In *Image Vision and Computing New Zealand 2006, IVCNZ 06 proceedings*, pages 355–360, november 2006.
- R. Baumgartner. Erkennung von Pollen in Honig-Proben. Diplomarbeit, Albert-Ludwigs-Universität Freiburg, August 2005.
- P. Bonton, A. Boucher, M. Thonnat, et al. Colour image in 2d and 3d microscopy for the automation of pollen rate measurement. *Image Analysis & Stereology*, 21:25–30, march 2002. URL <http://www.wise-t.com/ias/article.php?id=79&year=2002&issue=3>.
- A. Boucher and M. Thonnat. Object recognition from 3d blurred images. In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 1, pages 800–803, 2002. doi:10.1109/ICPR.2002.1044880.
- A. Boucher, P. Hidalgo, M. Thonnat, J. Belmonte, C. Galan, P. Bonton, and R. Tomczak. Development of a semi-automatic system for pollen recognition. *Aerobiologia*, 18:195–201, 2002. doi:10.1023/A:1021322813565.
- Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004. ISSN 0162-8828. doi:10.1109/TPAMI.2004.60.
- R. N. Bracewell. *The Fourier Transform and Its Applications, 3rd ed*, chapter Convolution Theorem, pages 108–112. McGraw-Hill, 1999. doi:10.1036/0073039381.
- H. Burkhardt and S. Siggelkow. Invariant features in pattern recognition - fundamentals and applications. In C. Kotropoulos and I. Pitas, editors, *Nonlinear Model-Based Image/Video Processing and Analysis*, pages 269–307. John Wiley & Sons, 2001.
- M. B. Bush and C. Y. Weng. Introducing a new (freeware) tool for palynology. *Journal Of Biogeography*, 34(3):377–380, Mar. 2007. doi:10.1111/j.1365-2699.2006.01645.x.
- C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- C. Chen, E. A. Hendriks, R. P. W. Duin, J. H. C. Reiber, P. S. Hiemstra, L. A. de Weger, and B. C. Stoel. Feasibility study on automated recognition of allergenic pollen: grass, birch and mugwort. *Aerobiologia*, 22(4):275–284, 2006. doi:10.1007/s10453-006-9040-0.

- P. Comtois, P. Alcazar, and D. Néron. Pollen counts statistics and its relevance to precision. *Aerobiologia*, 15(1):19–28, march 1999. doi:10.1023/A:1007501017470.
- G. Dahme, E. Ribeiro, and M. Bush. Spatial statistics of textons. In *International Conference of Computer Vision Theory and Applications - VISAPP*, Setubal, Portugal, 2006.
- J. R. Driscoll and D. M. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15(2):202–250, June 1994. doi:10.1006/aama.1994.1008.
- A. Duller, G. Duller, I. France, and H. Lamb. A pollen image database for evaluation of automated identification systems. *Quaternary Newsletter*, 89:4–9, 1999a.
- A. Duller, G. Duller, I. France, and H. Lamb. Bangor/aberystwyth pollen image database. http://www.informatics.bangor.ac.uk/~ian/pdbase/pollen_dbase.html, 1999b.
- J. Fehr. Lokalisierung und Erkennung von Zellkernen in Gewebeschnitten. Diplomarbeit, Albert-Ludwigs-Universität Freiburg, Novmeber 2004.
- J. Fehr, O. Ronneberger, H. Kurz, and H. Burkhardt. Self-learning segmentation and classification of cell-nuclei in 3d volumetric data using voxel-wise gray scale invariants. In W. Kropatsch and R. Sablating, editors, *Pattern Recognition - Proc. of the 27th DAGM Symposium, Vienna, Austria*. Springer, Berlin, Aug. 2005. doi:10.1007/11550518_47.
- P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. Technical Report TR2004-1963, Cornell Computing and Information Science, September 2004.
- J. Flenley. The problem of pollen recognition. In M. B. Clowes and J. Penny, editors, *Problems in Picture Interpretation*, pages 141–145. CSIRO, Canberra, 1968.
- I. France, A. Duller, H. Lamb, and G. Duller. A comparative study of model based and neural network based approaches to automatic pollen identification. In *British Machine Vision Conference*, pages 340–349, 1997. URL <http://www.bmva.ac.uk/bmvc/1997/papers/036/036.html>.
- I. France, A. W. G. Duller, G. A. T. Guller, and H. F. Lamb. A new approach to automated pollen analysis. *Quaternary Science Reviews*, 19(6):537–546, February 2000. doi:10.1016/S0277-3791(99)00021-9.
- M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. special issue on "Program Generation, Optimization, and Platform Adaptation".
- R. Gilad-Bachrach, A. Navot, and N. Tishby. Margin based feature selection - theory and algorithms. In *Proc. 21'st International Conference on Machine Learning (ICML)*, 2004. doi:10.1145/1015330.1015352.
- B. Haasdonk. *Transformation Knowledge in Pattern Analysis with Kernel Methods*. PhD thesis, Albert-Ludwigs-Universität Freiburg, 2005.
- B. Haasdonk, A. Halawani, and H. Burkhardt. Adjustable invariant features by partial Haar-integration. In *Proc. of the 17th ICPR, Cambridge, United Kingdom*, volume 2, pages 769–774. IEEE, Aug. 2004. doi:10.1109/ICPR.2004.1334372.
- A. Halawani. *Local Integral Invariants for Image Retrieval*. PhD thesis, University of Freiburg, July 2006.
- R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems Man and Cybernetics*, 3(6):610–621, November 1973.

Bibliography

- D. M. Healy, D. N. Rockmore, P. J. Kostelec, and S. Moore. Ffts for the 2-sphere-improvements and variations. *Journal of Fourier Analysis and Applications*, 9(4):341–385, July 2003. doi:10.1007/s00041-003-0018-9.
- R. Hodgson, C. Holdaway, Y. Zhang, D. Fountain, and J. Flenley. Progress towards a system for the automatic recognition of pollen using light microscope images. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pages 76–81, 2005.
- N. P. Ivleva, R. Niessner, and U. Panne. Characterization and discrimination of pollen by raman microscopy. *Analytical and Bioanalytical Chemistry*, 381(1):261–267, 2005. doi:10.1007/s00216-004-2942-1.
- A. S. Jones. Image analysis applied for aerobiology. In *2nd European Symposium on Aerobiology*, page 2, 2000. URL <http://web.archive.org/web/20010203153900/betula.hno.akh-wien.ac.at/s2000/abstracts.html>.
- A. S. Jones, M. Hjelmroos-Koski, and E. R. Tovey. Image analysis can be used to identify airborne allergenic pollen. *J ALLERGY CLIN IMMUN*, 103(1):S1, poster 721, January, part 2 1999. URL <http://www2.us.elsevierhealth.com/inst/serve?action=searchDB&searchDBfor=art&artType=misc&id=ai1031ab704>.
- M. Langford, G. Taylor, and J. Flenley. The application of texture analysis for automated pollen identification. In *Proceedings of a Conference on Identification and Pattern Recognition*, pages 729–739, Toulouse, 1986. Univ. Paul Sabatini.
- M. Langford, G. Taylor, and J. Flenley. Computerized identification of pollen grains by texture analysis. *Review of Palaeobotany and Palynology*, 64:197–203, 1990.
- P. Li and J. Flenley. Pollen texture identification using neural network. *Grana*, 38:59–64, 1999.
- P. Li, W. J. Treloar, J. R. Flenley, and L. Empson. Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. *Journal of quaternary science*, 19(8):755–762, 2004. doi:10.1002/jqs.874.
- D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. doi:10.1023/B:VISI.0000029664.99615.94.
- M. Mazière. étude de faisabilité pour la reconnaissance automatique de grains de pollen. Rapport de dea, Université de Nice-Sophia Antipolis, 1997. URL <http://www-sop.inria.fr/orion/Publications/index.html>.
- K. Miyamoto and T. Hoshimiya. Measurement of the amount and number of pollen particles of cryptomeria japonica (taxodiaceae) by imaging with a photoacoustic microscope. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 53(3):586–591, March 2006. doi:10.1109/TUFFC.2006.1610567.
- M. A. A. Neil, R. Juskaitis, and T. Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope. *OPTICS LETTERS*, 22(24):1905–1907, 1997.
- M. Ranzato, P. E. Taylor, J. M. House, R. C. Flagan, Y. LeCun, and P. Perona. Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28(1):31–39, January 2007. doi:10.1016/j.patrec.2006.06.010.
- V. Razmovski, T. J. O’Meara, D. J. M. Taylor, and E. R. Tovey. A new method for simultaneous immunodetection and morphologic identification of individual sources of pollen allergens. *Journal of Allergy and Clinical Immunology*, 105(4):725–731, Apr. 2000. doi:10.1067/mai.2000.105222.

- M. Reisert. Equivariant holomorphic filters - theory and applications. Technical Report Internal Report 1/07, IIF-LMB, Computer Science Department, University of Freiburg, 2007. URL http://lmb.informatik.uni-freiburg.de/papers/download/re_report07_01.pdf.
- M. Reisert and H. Burkhardt. Learning equivariant functions with matrix valued kernels. *J. Mach. Learn. Res.*, 8:385–408, 2007. URL <http://portal.acm.org/citation.cfm?id=1248659.1248674&coll=GUIDE&dl=%23url.coll>.
- M. Reisert and H. Burkhardt. Invariant features for 3d-data based on group integration using directional information and spherical harmonic expansion. In *Proceedings of International Conference on Pattern Recognition (ICPR 2006)*, Hong Kong. LNCS, Springer, September 2006. doi:10.1109/ICPR.2006.721.
- M. Rodriguez-Damian, E. Cernadas, A. Formella, and R. Sa-Otero. Pollen classification using brightness-based and shape-based descriptors. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, volume 2, pages 212–215 Vol.2, 2004. doi:10.1109/ICPR.2004.1334098.
- M. Rodriguez-Damian, E. Cernadas, A. Formella, M. Fernandez-Delgado, and P. D. Sa-Otero. Automatic detection and classification of grains of pollen based on shape and texture. *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, 36(4):531–542, 2006. ISSN 1094-6977. doi:10.1109/TSMCC.2005.855426.
- O. Ronneberger, U. Heimann, V. Dietze, E. Schultz, H. Burkhardt, and R. Gehrig. Automated pollen recognition using gray scale invariants on 3d volume image data. In *2nd European Symposium on Aerobiology*, Vienna - Austria, 2000. URL <http://web.archive.org/web/20010203153900/betula.hno.akh-wien.ac.at/s2000/abstracts.html>.
- O. Ronneberger, H. Burkhardt, and E. Schultz. General-purpose object recognition in 3d volume data sets using gray-scale invariants – classification of airborne pollen-grains recorded with a confocal laser scanning microscope. In *Proceedings of the International Conference on Pattern Recognition*, Quebec, Canada, Sept. 2002a. doi:10.1109/ICPR.2002.1048297.
- O. Ronneberger, E. Schultz, and H. Burkhardt. Automated pollen recognition using 3d volume images from fluorescence microscopy. *Aerobiologia*, 18:107–115, 2002b. doi:10.1023/A:1020623724584.
- O. Ronneberger, J. Fehr, and H. Burkhardt. Voxel-wise gray scale invariants for simultaneous segmentation and classification. In *Pattern Recognition - Proc. of the 27th DAGM Symposium, Vienna, Austria*. Springer, Berlin, Aug. 2005a. doi:10.1007/11550518_11. DAGM-Award.
- O. Ronneberger, J. Fehr, and H. Burkhardt. Voxel-wise gray scale invariants for simultaneous segmentation and classification – theory and application to cell-nuclei in 3d volumetric data. Internal report 2/05, IIF-LMB, University Freiburg, april 2005b.
- O. Ronneberger, Q. Wang, and H. Burkhardt. 3d invariants with high robustness to local deformations for automated pollen recognition. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition – Proceedings of the 29th DAGM Symposium*, volume 4713/2007 of LNCS, pages 425–435, Heidelberg, Germany, September 12–14 2007. Springer Berlin / Heidelberg. doi:10.1007/978-3-540-74936-3_43.
- M. P. d. Sa-Otero, A. Gonzalez, M. Rodriguez-Damian, and E. Cernadas. Computer-aided identification of allergenic species of urticaceae pollen. *Grana*, 43(4):224–230, 2004. doi:10.1080/00173130410000749.

Bibliography

- L. H. Schaefer, D. Schuster, and J. Schaffer. Structured illumination microscopy: artefact analysis and reduction utilizing a parameter optimization approach. *Journal of Microscopy*, 216(2):165–174, November 2004. doi:10.1111/j.0022-2720.2004.01411.x.
- M. Schael. Invariantenbasierte Objekterkennung aus dreidimensionalen Sensordaten. Diplomarbeit, Technische Universität Hamburg-Harburg, 1996.
- M. Schael. Invariant greyscale features for 3D sensordata. Technical Report 9/98, Albert-Ludwigs-Universität, Freiburg, Institut für Informatik, 1998.
- M. Schael. *Methoden zur Konstruktion invarianter Merkmale für die Texturanalyse*. PhD thesis, Albert-Ludwigs-Universität Freiburg, 2004.
- M. Schael and S. Siggelkow. Invariant grey-scale features for 3d sensor-data. In *Proceedings of the International Conference on Pattern Recognition (ICPR2000)*, pages 531 – 535, Barcelona, Spain, September 2000. doi:10.1109/ICPR.2000.906129.
- S. Scharring, A. Brandenburg, G. Breitfuss, H. Burkhardt, W. Dunkhorst, M. v. Ehr, M. Fratz, D. Giel, U. Heimann, W. Koch, H. Lödding, W. Müller, O. Ronneberger, E. Schultz, G. Sulz, and Q. Wang. Online monitoring of airborne allergenic particles (omnibuss). In J. Popp and M. Strehle, editors, *Biophotonics: Visions for Better Health Care*, pages 31–88. Wiley-VCH, june 2006. doi:10.1002/3527608842.ch2.
- T. Schmidt. Entwicklung von Vorverarbeitungsverfahren für fluoreszenzmikroskopische 3D Volumendaten zur Verbesserung der nachfolgenden Erkennung mittels grauwertbasierter Invarianten. Studienarbeit, Albert-Ludwigs-Universität Freiburg, Dezember 2003.
- T. Schmidt. Automatische Selektion von optimalen Kernfunktionen für grauwertbasierte Invarianten. Diplomarbeit, Albert-Ludwigs-Universität Freiburg, April 2004.
- J. Schulz. Lokalisierung von Zellkernen und Bestimmung ihrer Mitosephase in konfokalen 3d-Aufnahmen von *Arabidopsis thaliana*. Diplomarbeit, Albert-Ludwigs-Universität Freiburg, Dezember 2005.
- J. Schulz, T. Schmidt, O. Ronneberger, H. Burkhardt, T. Pasternak, A. Dovzhenko, and K. Palme. Fast scalar and vectorial grayscale based invariant features for 3d cell nuclei localization and classification. In *Proceedings of the 28th Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM 2006)*, Berlin, Germany. LNCS, Springer, 2006. doi:10.1007/11861898_19. DAGM-Award.
- H. Schulz-Mirbach. Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17. DAGM-Symposium "Mustererkennung"*, Informatik aktuell, pages 1–14. Springer, 1995a.
- H. Schulz-Mirbach. *Anwendung von Invarianzprinzipien zur Merkmalgewinnung in der Mustererkennung*. PhD thesis, Technische Universität Hamburg-Harburg, feb 1995b. Reihe 10, Nr. 372, VDI-Verlag.
- C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1): 10–21, 1949.
- S. Siggelkow. *Feature Histograms for Content-Based Image Retrieval*. PhD thesis, Albert-Ludwigs-Universität Freiburg, Dec. 2002.

- S. Siggelkow and M. Schael. Fast estimation of invariant features. In W. Förstner, J. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung, DAGM 1999*, Bonn, sep 1999. Springer.
- E. Stillman and J. R. Flenley. The needs and prospects for automation in palynology. *Quaternary Science Reviews*, 15:1–5(5), 1996. doi:10.1016/0277-3791(95)00076-3.
- Y. Takahashi, T. Nagoya, and H. Nitta. Autofluorescent image analysis in pollen counting. In *2nd European Symposium on Aerobiology*, 2000. URL <http://web.archive.org/web/20010203153900/betula.hno.akh-wien.ac.at/s2000/abstracts.html>.
- W. J. Treloar, G. E. Taylor, and J. R. Flenley. Towards automation of palynology 1: analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images. *Journal of quaternary science*, 19(8):745–754, 2004. doi:10.1002/jqs.871.
- Y. Ujiie. Brightness of pollen as an indicator of thermal alteration by means of a computer-driven image processor: statistical thermal alteration index (sttai). *Organic Geochemistry*, 32(1):127–141, Jan. 2001. doi:10.1016/S0146-6380(00)00146-7.
- V. N. Vapnik. *The nature of statistical learning theory*. Springer, 1995.
- A. Vossen. Optische Schnitte mit Hilfe strukturierter Beleuchtung. Studienarbeit, Albert-Ludwigs-Universität Freiburg, Januar 2004.
- Q. Wang. Expression of gray scale invariants with monomial kernels in fourier space. unpublished internal report, November 2004.
- S. Wilhelm, B. Gröbler, M. Gluch, and H. Heinz. *Confocal Laser Scanning Microscopy – Optical Image Formation, Electronic Signal Processing*. Carl Zeiss Jena, 2006. 45-0029 e/09.03.
- H. Winkler, R. Ostrowski, and M. Wilhelm. *Pollenbestimmungsbuch der Stiftung Deutscher Polleninformationsdienst*. Stiftung Deutscher Polleninformationsdienst, D-33175 Bad Lippspringe, 2001. ISBN 3-931732-09-6.
- C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *Image Processing, IEEE Transactions on*, 7(3):359–369, 1998. ISSN 1057-7149. doi:10.1109/83.661186.
- Y. Zhang, D. W. Fountain, R. M. Hodgson, J. R. Flenley, and S. Gunetileke. Towards automation of palynology 3: pollen pattern recognition using gabor transforms and digital moments. *Journal of quaternary science*, 19(8):763–768, 2004. doi:10.1002/jqs.875.
- Y. Zhang, R. Wang, and P. Hunter. Airborne pollen texture discrimination using wavelet transforms in combination with cooccurrence matrices. *Int. J. Intelligent Systems Technologies and Applications*, 1:143–156, 2005. doi:10.1504/IJISTA.2005.007312.