

Praktikumsversuch
Aktives Sehen

C. Bahlmann, M. Schael, O. Ronneberger und H. Burkhardt

Inhaltsverzeichnis

1 Aktives Sehen	5
1.1 Einleitung	5
1.1.1 Was ist Aktives Sehen?	5
1.1.2 Bewegungsschätzung, Objektverfolgung und Aktives Sehen	6
1.1.3 Bewegungsschätzung im Praktikum	6
1.2 Methoden zur Bewegungsschätzung	8
1.2.1 Implizite Bewegungsschätzung	10
1.2.2 Explizites Verfahren zur Bewegungsschätzung	14
1.2.3 Vergleich der Verfahren zur Bewegungsschätzung	20
1.3 Versuchsaufbau	23
1.4 Praktikumsaufgaben	25
1.4.1 Aufgaben zum Impliziten Verfahren	25
1.4.2 Aufgaben zum Expliziten Verfahren	25
1.5 Gauß'sche Mischverteilungen	26
1.5.1 Das Modell	26
1.5.2 Der EM-Schätzalgorithmus für die Verteilungsparameter .	28
1.5.3 Bestimmung der Komponentenanzahl	33

1 Aktives Sehen

1.1 Einleitung

1.1.1 Was ist Aktives Sehen?

Der Begriff *Aktives Sehen* beschreibt eine Methodik in der Bildverarbeitung, in dem die Kamera in einem Bildverarbeitungssystem eine *aktive* Rolle einnimmt. Gemäß einer Definition von Swain und Stricker [8], bezeichnet *Aktives Sehen* Sehsysteme mit der Eigenschaft, entsprechend der behandelten Aufgabe aktiv Änderungen von Kameraparametern vorzunehmen. Beispielsweise kann man sich Änderungen von Position und Orientierung der Kamera oder auch von Fokus, Zoom, Apertur oder bei binokularen Systemen der Vergenz, d.h. der Orientierung der beiden Kameraachsen zueinander, vorstellen. Als typisches Merkmal gilt dabei stets, daß die Bildverarbeitungsoperationen und die hervorgerufenen Aktionen außerordentlich eng gekoppelt sind.

Die Möglichkeiten, die Sehsysteme mit aktiver Kamerasteuerung bieten, sind vielfältig. Der Artikel von Swain und Stricker [8] beschreibt eine *Aufmerksamkeitsselektion* in aktiven Bildverarbeitungssystemen. Hierbei wird die Bildszene in bestimmten Parametern (Swain und Stricker [8] geben dafür Raum, Auflösung und Zeit an) *selektiv* erfaßt. So kann zum Beispiel—ähnlich den Charakteristika des menschlichen Sehsystems—der zentrale Bereich einer Szene mit einer höheren Bildauflösung erfaßt werden.

Einen typischen Anwendungsbereich für Aktive Sehsysteme sieht Ballard [1] in der *physikalischen Suche*. Dieser Begriff beschreibt das Vorgehen, aktiv durch Änderung der Kameraparameter, also z.B. von Position, Blickwinkel oder Zoom eine natürliche Szene zu explorieren. Die Autoren argumentieren, daß auf diese Weise die Bildverarbeitung in vielen Fällen effektiver, effizienter und robuster gehandhabt werden kann als im Gegensatz dazu bei einer *algorithmischen Suche* auf der Aufnahme eines einzelnen Bildes.

1.1.2 Bewegungsschätzung, Objektverfolgung und Aktives Sehen

In diesem Praktikum *Aktives Sehen* werden die Begriffe *Bewegungsschätzung* bzw. *Objektverfolgung* den Schwerpunkt bilden. Bei der *Bewegungsschätzung* handelt es sich um die Problemstellung, einen Bewegungsparametervektor \mathbf{T} von zwei zeitversetzt aufgenommenen Bildern eines bewegten Objektes zu bestimmen. Typische Parameter des Vektors \mathbf{T} sind Translations-, Rotations- und eventuell Skalierungsparameter. Der geschätzte Vektor \mathbf{T} kann anschließend z.B. von einem aktiven Sehsystem verwendet werden, um Position und Orientierung der Kamera nachzustellen. Ein solcher Vorgang wird *Objektverfolgung* bezeichnet.

Wie lassen sich die oben eingeführten Begriffe in das Gebiet *Aktives Sehen* einordnen? Dazu stelle man sich eine Szene vor, in dem sich das interessierende Objekt, z. B. ein menschliches Gesicht, vor einem Hintergrund bewegt. Möchte man das Objekt weiterverarbeiten, z. B. die Blickrichtung des Gesichts bestimmen, so läßt sich die Aufgabe unter dem Aspekt des aktiven Sehens und der oben beschriebenen Aufmerksamkeitsselektion vereinfachen, wenn die Bewegungsparameter des Gesichtes bekannt sind. Anhand dieser kann die Kamera dem Gesicht nachgeführt werden und das Gesicht kann mit einer höheren Auflösung aufgenommen werden.

Weitere Anwendungsgebiete der hier behandelten Thematik liegen in der Robotik, der Meteorologie, der Bewegtbildcodierung, in einer berührungslosen Geschwindigkeitsmessung, oder auch bei Multimediaanwendungen wie einer Sprecherfixation bei einer Videokonferenz.

1.1.3 Bewegungsschätzung im Praktikum

Das Praktikum wird zwei verschiedene Strategien zur Bewegungsschätzung besprechen. Sie unterscheiden sich insbesondere durch die zugrundeliegenden Schätztechniken, die die Bewegungsparameter berechnen:

1. einen *impliziten* Algorithmus, in dem die Bewegungsparameter über eine Korrelation geschätzt werden, und
2. eine *explizite* Schätzformel.

Die vorliegende Praktikumsbeschreibung ist folgendermaßen gegliedert: Im folgenden Kapitel 1.2 werden die beiden Schätzverfahren vorgestellt. Abschnitt 1.2.1 beschäftigt sich mit dem impliziten, Abschnitt 1.2.2 mit dem expliziten Verfahren. Eine vergleichende Betrachtung wird anschließend in Abschnitt 1.2.3 vollzogen. Kapitel 1.3 erklärt dann die Umgebung, in der der praktische Teil des Versuchs bearbeitet wird, bevor in Kapitel 1.4 schließlich die Versuchsaufgaben formuliert

sind. Der Anhang 1.5 erläutert ein Modell von Wahrscheinlichkeitsverteilungen, welches bei der expliziten Schätzmethode benötigt wird, die Gauß'schen Mischverteilungen.

1.2 Methoden zur Bewegungsschätzung

Der Begriff Bewegungsschätzung beschreibt den Vorgang, die Transformationsparameter \mathbf{T} einer auf ein Objekt angewandten Transformation $\mathbf{h}(\mathbf{x}, \mathbf{T})$ aus einer Bildfolge zu schätzen. Die theoretischen Grundlagen der Bewegungsschätzung werden ausführlich in der Vorlesung [3] vermittelt; in diesem Abschnitt werden sie zusammengefaßt.

Die Transformation \mathbf{h} beschreibt hierbei die Bewegung (Rotation und/oder Translation) eines 2D oder 3D Objektes im 2D oder 3D Raum. Es ist evident, daß bei der Projektion eines Objektes im 3D Raum auf die 2D Kameraebene Information dezimiert wird, und die Transformationsparameter schwierig zu rekonstruieren sind. In [5] werden für die Zentral- und Parallelprojektion eines planaren Objektes im 3D Raum die Formel der Projektion angegeben. Für die Zentralprojektion beschreibt ein achtelementiger Parametervektor \mathbf{T} die Transformation. Der Parametervektor der Parallelprojektion umfaßt noch sechs Elemente.

In diesem Kapitel der Praktikumsanleitung werden wir uns auf die Schätzung von translatorischen Bewegungen planarer Objekte in der Ebene beschränken. Wobei die Translationsebene parallel zur Kameraebene ist. Abbildung 1.1 gibt einen Eindruck von der Translation eines Objektes. Von einem Objekt werden zwei aufeinanderfolgende Aufnahmen \mathbf{I} und \mathbf{I}_T gemacht. Die Bilder \mathbf{I} und \mathbf{I}_T unterscheiden sich durch eine auf das Objekt ausgeübte Transformation \mathbf{h} , welche über die Transformationsparameter \mathbf{T} spezifiziert wird. Der formale Zusammenhang zwischen den Bildern läßt sich folgendermaßen angeben:

$$\begin{aligned}\mathbf{I}(\mathbf{x}) &= \mathbf{I}_O(\mathbf{x}) + \mathbf{N}(\mathbf{x}), \\ \mathbf{I}_T(\mathbf{x}, \mathbf{T}) &= \mathbf{I}_{OT}(\mathbf{x}, \mathbf{T}) + \mathbf{N}_T(\mathbf{x}) = \mathbf{I}_O(\mathbf{h}(\mathbf{x}, \mathbf{T})) + \mathbf{N}_T(\mathbf{x}),\end{aligned}$$

wobei $\mathbf{x} = (x, y)^T$. Das Bild \mathbf{I}_O repräsentiert das durch die Projektion aus dem 3D Raum entstandene Bild von dem Objekt. Dementsprechend stellt $\mathbf{I}_{OT}(\mathbf{x}, \mathbf{T})$ resp. $\mathbf{I}_O(\mathbf{h}(\mathbf{x}, \mathbf{T}))$ das Bild von dem transformierten Objekt dar. Zwischen der Abbildung \mathbf{I}_O des Objektes und dem Gesamtbild \mathbf{I} wird unterschieden, da neben der Projektion vom Objekt auch ein Teil der Umgebung des Objektes auf die Kameraebene abgebildet wird. Vereinfachend wird angenommen, daß die Umgebung annähernd homogen ist. Diese Annahme kann auch als Segmentierung des Objektes resp. als Definition einer "Region-of-Interest" (ROI) verstanden werden.

Die Translation des Objekts läßt sich durch den Zusammenhang

$$\mathbf{x}_T = \mathbf{h}(\mathbf{x}, \mathbf{T}) = (x - t_x, y - t_y)^T$$

ausdrücken. Der Parametervektor der Transformation ist demnach $\mathbf{T} = (t_x, t_y)^T$. Bei den Termen \mathbf{N} und \mathbf{N}_T handelt es sich um Störungen der Bilder in Form von additiven Rauschen.

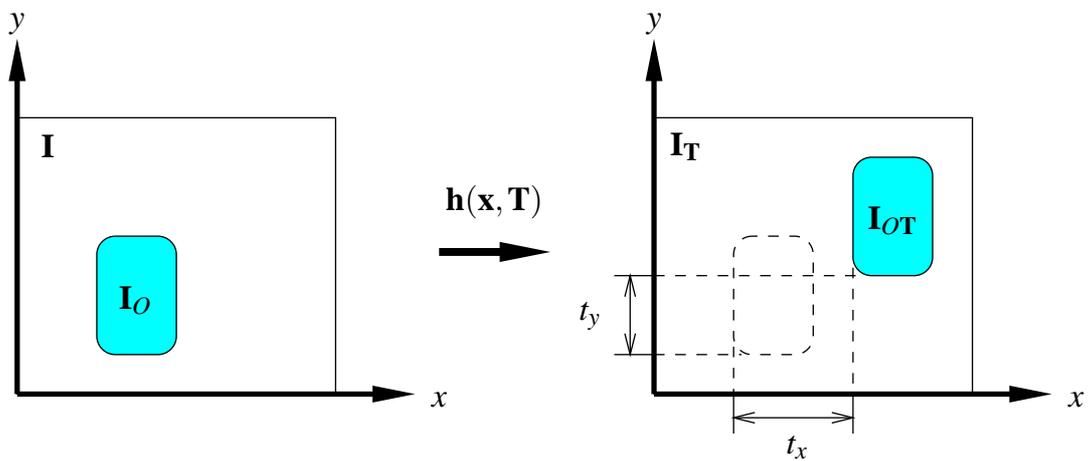


Abbildung 1.1: Translatorische Transformation eines planaren Objektes in der Ebene.

1.2.1 Implizite Bewegungsschätzung

In diesem Kapitel wird nun ein implizites Verfahren zur Bewegungsschätzung hergeleitet. Implizite Ansätze dienen zur numerischen Lösung von Problemstellungen, deren Lösung nicht (oder nur mit hohem Aufwand) explizit über eine Gleichung angegeben werden kann. Zu den Problemstellungen, die sich meist nur implizit lösen lassen, gehören zum Beispiel Differentialgleichungen die ein dynamisches System beschreiben. Implizite Verfahren sind somit auch Bestandteil der numerischen Mathematik [6]. Auch das Problem der Bewegungsschätzung läßt sich über einen impliziten Ansatz lösen.

Modelladaptive Parameterschätzung

Der hier vorgestellte Ansatz der modelladaptiven Parameterschätzung zur Bewegungsschätzung basiert auf der Dissertation von N. Diehl [5]. Für den Ansatz wird vorausgesetzt, daß die Kamerahauptachse orthogonal zur Objektebene liegt und keine Verdeckungseffekte auftreten. Desweiteren wird bei dem Ansatz ein starres Objekt gefordert. Die Transformation \mathbf{h} — also die mögliche Bewegung — gilt es als bekannt vorauszusetzen.

In Abbildung 1.2 ist das allgemeine Prinzip der Bewegungsschätzung mittels einer modelladaptiven Parameterschätzung dargestellt. Ein Objekt ist in dem Originalbild \mathbf{I} auf einem homogenen Hintergrund abgebildet. Bei dem Bild kann es sich um ein Grauwertbild oder ein Farbbild handeln, was allerdings in der Realisierung der Parameterschätzung zu berücksichtigen ist. Der Einfachheit halber werden wir hier nur Grauwertbilder betrachten. Auf das Bild \mathbf{I} wirkt eine Transformation \mathbf{h} im Sinne einer Bewegung. $\mathbf{I}_{\mathbf{T}}$ bezeichnet das transformierte Bild. Ziel ist es, den Parametervektor \mathbf{T} der Transformation \mathbf{h} aus den aufeinanderfolgenden Bildern \mathbf{I} und $\mathbf{I}_{\mathbf{T}}$ zu schätzen. Um die Herleitung der Theorie für die implizite Bewegungsschätzung überschaubar zu halten, wird der Idealfall angenommen: es treten keine Störungen in Form von additiven Rauschen auf. D.h. $\mathbf{N} \equiv \mathbf{0}$ und $\mathbf{N}_{\mathbf{T}} \equiv \mathbf{0}$.

Über einen geschätzten Transformationsparametersatz $\hat{\mathbf{T}}$ wird ein Modellbild $\mathbf{I}_{\hat{\mathbf{T}}}$ basierend auf dem Originalbild \mathbf{I} berechnet. Für das Modellbild gilt:

$$\mathbf{I}_{\hat{\mathbf{T}}}(\mathbf{x}, \hat{\mathbf{T}}) = \mathbf{I}_O(\mathbf{h}(\mathbf{x}, \hat{\mathbf{T}})) + \mathbf{N}(\mathbf{h}(\mathbf{x}, \hat{\mathbf{T}}))$$

Wie schon erwähnt, gehen wir davon aus, daß keine Störungen auftreten. D.h. das Modellbild $\mathbf{I}_{\hat{\mathbf{T}}}$ ergibt sich direkt aus dem transformierten Objektbild \mathbf{I}_O . Da \mathbf{T} nicht bekannt ist, wird die Schätzung mehr oder weniger fehlerhaft sein. Die Abweichung des Modellbildes $\mathbf{I}_{\hat{\mathbf{T}}}$ von dem Transformationsbild $\mathbf{I}_{\mathbf{T}}$ wird über die Differenz

$$\mathbf{e}(\hat{\mathbf{T}}) = \mathbf{I}_{\hat{\mathbf{T}}}(\mathbf{x}, \hat{\mathbf{T}}) - \mathbf{I}_{\mathbf{T}}(\mathbf{x}, \mathbf{T})$$

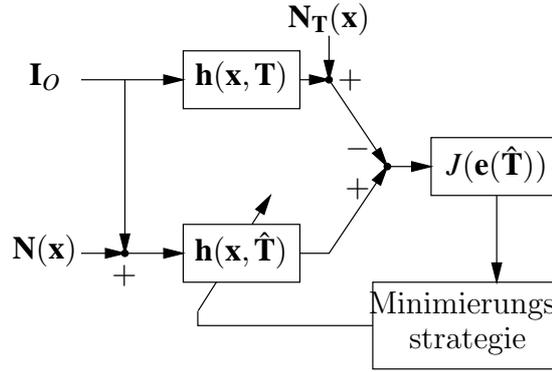


Abbildung 1.2: Schema eines modelladaptiven Schätzverfahrens zur Bewegungsschätzung.

berechnet. Für einen gegebenen Modellparameter $\hat{\mathbf{T}}$ läßt sich also ein Fehler für jede Bildposition \mathbf{x} bestimmen. Ausgehend von dem Modellfehler \mathbf{e} wird jetzt das Fehlermaß J über den Erwartungswert des quadratischen Modellfehlers definiert:

$$\begin{aligned} J(\hat{\mathbf{T}}) &= \frac{1}{2} E \{ \mathbf{e}(\hat{\mathbf{T}})^2 \} \\ &= \frac{1}{2} E \{ (\mathbf{I}_{\hat{\mathbf{T}}}(\mathbf{x}, \hat{\mathbf{T}}) - \mathbf{I}_{\mathbf{T}}(\mathbf{x}, \mathbf{T}))^2 \} \end{aligned} \quad (1.1)$$

Die Erwartungswertbildung ist hier im Sinne einer Ortsmittelung – also eine Mittelung über alle Bildpositionen \mathbf{x} – zu verstehen. Die Fehlerfunktion J ist also durch eine skalare Funktion repräsentiert.

Ziel der Bewegungsschätzung ist es, den Transformationsparameter so zu schätzen, daß $\hat{\mathbf{T}} \approx \mathbf{T}$ ist, und der Fehler J minimal wird:

$$J(\hat{\mathbf{T}}) = \min_{\hat{\mathbf{T}}}! \quad (1.2)$$

Das Minimierungsproblem der Varianz der Modellabweichung läßt sich alternativ über einen Korrelationsansatz lösen. Hierzu wird der Erwartungswert des Modellfehlers als Innenprodukt formuliert. Es gilt:

$$2J(\hat{\mathbf{T}}) = \|\mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}}\|^2 = \langle \mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}}, \mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}} \rangle .$$

Das Innenprodukt $\langle \cdot, \cdot \rangle$ läßt sich zu folgendem Ausdruck vereinfachen:

$$\langle \mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}}, \mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}} \rangle = \|\mathbf{I}_{\hat{\mathbf{T}}}\|^2 + \|\mathbf{I}_{\mathbf{T}}\|^2 - 2 \langle \mathbf{I}_{\hat{\mathbf{T}}}, \mathbf{I}_{\mathbf{T}} \rangle .$$

Für den vereinfachten Fall (Objekt vor schwarzem Hintergrund) sind die Ausdrücke $\|\mathbf{I}_{\hat{\mathbf{T}}}\|^2$ und $\|\mathbf{I}_{\mathbf{T}}\|^2$ nicht von dem Transformationsparameter \mathbf{T} abhängig. Damit der gesamte Ausdruck $\|\mathbf{I}_{\hat{\mathbf{T}}} - \mathbf{I}_{\mathbf{T}}\|^2$ minimal wird, muß dann nur der Term $\langle \mathbf{I}_{\hat{\mathbf{T}}}, \mathbf{I}_{\mathbf{T}} \rangle$ maximal werden. Dies entspricht einer Maximierung der Kreuzkorrelation der Bilder, die über die Formel

$$\langle \mathbf{I}_{\hat{\mathbf{T}}}, \mathbf{I}_{\mathbf{T}} \rangle = \sum_x \sum_y \mathbf{I}_{\hat{\mathbf{T}}}(x, y) \mathbf{I}_{\mathbf{T}}(x, y)$$

berechnet werden kann. Allerdings ist der zur direkten Berechnung der Korrelation notwendige Aufwand erheblich. Für jeden auszuwertenden Parametersatz $\hat{\mathbf{T}}$ sind $W \cdot H$ Multiplikationen und Additionen auszuführen. Durch Herunterskalieren der Bilder lassen sich diese Berechnungen für jeden möglichen Parametersatz $\hat{\mathbf{T}}$ jedoch noch mit vertretbarem Zeitaufwand ausführen.

Bei noch komplexeren Transformationen (z.B. Translation und Rotation) ist das nicht mehr möglich. Hier ist man auf iterative Verfahren angewiesen (siehe Vorlesung Bildverarbeitung II)

Objektverfolgung vor ungleichmäßigem Hintergrund



Abbildung 1.3: Bewegtes Objekt vor ungleichmäßigem Hintergrund

Die Vereinfachungen aus dem letzten Abschnitt (gleichmäßiger schwarzer Hintergrund) sind in der Realität nur selten anzutreffen. Für die vorliegende Aufgabenstellung, Verfolgung eines Gesichts mit der Kamera, muss ein beliebiger Hintergrund erlaubt sein (siehe Abb. 1.3)

Dazu muss vom Originalbild nur der Teil verwendet werden, der das Gesicht enthält. Der Einfachheit halber wird dieser Ausschnitt rechteckig gewählt. Für die Berechnung des Fehlermaß J heißt das dann nur, dass das Integrationsintervall entsprechend angepasst werden muss, d.h. es werden nicht mehr alle Pixel des Modellbildes mit denen des Transformationsbildes verglichen, sondern nur noch diejenigen, die zum Gesicht gehören (Abb. 1.4)

Auch hier lässt sich das Fehlermaß wieder als Korrelation ausdrücken, allerdings gilt hier nicht mehr die Unabhängigkeit des Ausdrucks $\|\mathbf{I}_{\mathbf{T}}\|^2$ vom Transformationsparameter \mathbf{T} , so dass wir hier den normierten Korrelationskoeffizient verwenden wollen:

$$r(\hat{\mathbf{T}}) = \frac{\langle \mathbf{I}_{\hat{\mathbf{T}}} - \hat{m}, \mathbf{I}_{\mathbf{T}} - m \rangle}{\|\mathbf{I}_{\hat{\mathbf{T}}} - \hat{m}\| \cdot \|\mathbf{I}_{\mathbf{T}} - m\|} \quad (1.3)$$



Abbildung 1.4: Vom Originalbild wird nur der Teil benutzt, der das Gesicht enthält

wobei \hat{m} den Mittelwert des Modellbildes $\mathbf{I}_{\hat{\mathbf{T}}}$ und m den Mittelwert des Transformationsbildes $\mathbf{I}_{\mathbf{T}}$ bezeichnet. Der normierte Korrelationskoeffizient hat einen Wertebereich von -1 bis +1 und außerdem die schöne Eigenschaft sowohl gegenüber additiven als auch multiplikativen Veränderungen der Grauwerte invariant zu sein.

Aufgabe 1: Schreiben sie die Formel 1.3 unter Verwendung von Summen Σ auf. Achten sie insbesondere auf die korrekten Summationsgrenzen. Nutzen Sie x, y als "Laufvariablen" und $\Delta x, \Delta y$ als Parameterisierung für die Translation $\hat{\mathbf{T}}$. Die Größe des Gesichts-Bildes sei w, h . Der Index des linken oberen pixels soll (Matlab-konform) mit 1,1 bezeichnet werden. Das Bild mit dem Gesicht soll mit \mathbf{T} (template) bezeichnet werden und das aktuelle Kamerabild mit \mathbf{I}

1.2.2 Explizites Verfahren zur Bewegungsschätzung

Einleitung

Für eine vergleichende Untersuchung mit dem in Kapitel 1.2.1 besprochenen, *impliziten* Lösungsansatz zur Parameterschätzung wird in diesem Abschnitt ein Schätzverfahren beschrieben, welches die Bewegungsparameter *explizit* bestimmt. Dies bedeutet, die Bewegungsparameter \mathbf{T} werden direkt in einer geschlossenen Berechnungsvorschrift, im Gegensatz zu einer iterativen Vorgehensweise bestimmt.

Eine Vergleichsbetrachtung von impliziter und expliziter Methodik und eine Diskussion der Vor- und Nachteile solcher Vorgehensweisen wird in diesem Abschnitt noch nicht angestellt; dies wird im Anschluß in Kapitel 1.2.3 geschehen.

Ein Überblick über die hier vorgestellte Idee gibt die Abbildung 1.5 wieder: Unter Verwendung eines Modells \mathbf{M}_O des zu verfolgenden Objektes werden die interessierenden Bewegungsparameter \mathbf{T} des transformierten Bildes $\mathbf{I}_T(\mathbf{x}, \mathbf{T})$ in einer geschlossenen Formel berechnet. Das Modell \mathbf{M}_O wird durch Merkmale einer „Region-of-Interest“ (ROI) \mathbf{I}_O des Objektes geformt. Man beachte, daß \mathbf{M}_O über den gesamten Verlauf der Objektverfolgung nicht verändert wird. Mit dem gleichen Modell werden auch die Bewegungsparameter \mathbf{T}' eines weiteren transformierten Bildes $\mathbf{I}_{T'}(\mathbf{x}, \mathbf{T}')$ berechnet.

Literatur zu diesem Schätzverfahren befindet sich in einem Aufsatz von Raja et al. [7]. Eine auf dieses Praktikum abgestimmte Beschreibung erfolgt in den folgenden Abschnitten. Es werden die drei wesentlichen Aspekte „Region-of-Interest Extraktion“ (Abschnitt 1.2.2), „Merkmalextraktion“ (Abschnitt 1.2.2) und „Schätzformel der Bewegungsparameter“ (Abschnitt 1.2.2) erklärt.

Festlegung der „Region of Interest“

Ausgehend von einer Bildaufnahme, welche das zu verfolgende Objekt beinhaltet, wird eine—nicht notwendigerweise rechteckige—Region \mathbf{I}_O markiert, welche für die weitere Vorgehensweise als Referenz für das zu verfolgende Objekt dienen soll. Dieser Bereich wird in der Literatur meistens „Region of Interest“ (ROI) genannt. Abbildung 1.6 zeigt an einem Beispielbild die Extraktion einer ROI.

Merkmalextraktion

Wie in der Einleitung angesprochen wurde, werden die Bewegungsparameter des transformierten Bildes \mathbf{I}_T nicht basierend auf dem unverarbeiteten Modellobjekt

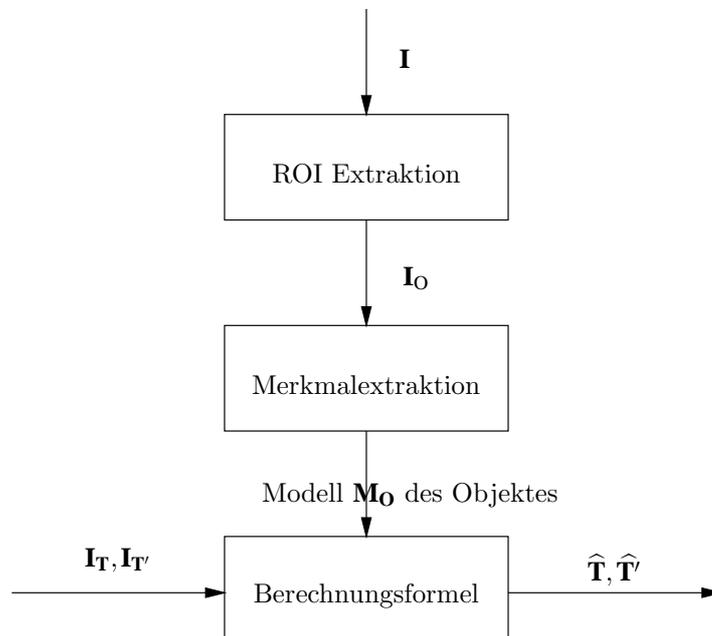


Abbildung 1.5: Blockdiagramm, welches das hier beschriebene, explizite Verfahren beschreibt: Die Transformationsparameter T werden im Gegensatz zu dem impliziten Verfahren ohne einen Iterationsalgorithmus direkt aus dem transformierten Bild I_T mit Hilfe eines Objektmodelles M_O berechnet. Das Modell M_O besteht dabei aus einem Merkmalsatz, der aus der „Region-of-Interest“ des betrachteten Objektes gewonnen wird. Es bleibt im Gegensatz zu dem impliziten Verfahren über den gesamten Prozeß der Objektverfolgung konstant. Daher werden auch die Bewegungsparameter T' eines weiteren transformierten Bildes $I_{T'}$ unter Berücksichtigung von M_O berechnet.

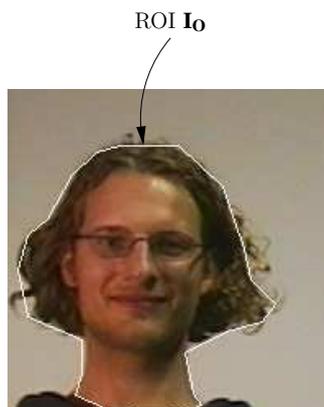


Abbildung 1.6: Beispiel für die Extraktion einer „Region of Interest“ (ROI).

\mathbf{I}_O , sondern auf *Merkmalen* \mathbf{M}_O von \mathbf{I}_O berechnet.

Die Bildung von Merkmalen ist eine in der Muster- bzw. Bilderkennung fundamentale Methode, welche beabsichtigt, den wesentlichen Informationsgehalt aus den Bildern zu extrahieren und den unwesentlichen Informationsgehalt zu verwerfen (siehe Vorlesung [4]). Dessen Zweck liegt in der Dimensionsreduktion des meistens sehr hochdimensionalen Bild-Vektorraumes auf einen niedriger dimensionalen Merkmalraum. Bei einer der Problemstellung entsprechenden Merkmalextraktion kann die Mustererkennung in der Regel effizienter und robuster durchgeführt werden.

Das in diesem Schätzverfahren verwendete Mustermerkmal wird die *Farbverteilung* des zu verfolgenden Objektes bilden. Die Autoren des Aufsatzes [7] argumentieren, daß in Bezug auf Verdeckungen, Rotation in der Tiefe und Skalierungs- bzw. Auflösungsänderungen Merkmale mit Farbverteilungsinformation robuster gegenüber geometrischen Ansätzen sind.

Dieses Vorgehen impliziert jedoch, daß die vorausgesetzten Objekte ausschließlich anhand ihrer Farbverteilung erkannt werden können. Sämtliche Information z.B. über die geometrische Struktur der Objekte bleiben bei dem vorgenommenen Ansatz unberücksichtigt.

Zur Modellierung der Farbverteilung eines Bildbereichs $\mathbf{X} = \{\mathbf{x}_{ij}\}_{i=1\dots M, j=1\dots N}$ ($\mathbf{x}_{ij} = (i, j)$) betrachte man zunächst ein Farbattribut ξ . Das Attribut ξ kann zum Beispiel durch einen Vektor der drei RGB-Farbwerte, der HSV-Farbwerte oder auch durch einen Vektor in einem RGB- bzw. HSV-Unterraum gebildet werden. In diesem Fall wird ξ durch den H- (gibt den *Farbton* an, vergl. Praktikumsversuch „Farbe“) und den S-Farbwert (gibt die *Sättigung* an) des HSV-Farbraumes definiert, $\xi = (H, S)$. Durch das Ignorieren der Größe V (gibt die *Helligkeit* an) erhält man bereits eine gewisse Invarianz gegenüber Helligkeitsschwankungen, die z. B. durch eine variierende Beleuchtung hervorgerufen werden können.

Das indizierte Attribut $\xi_{\mathbf{x}}$ gibt entsprechend den Farbwert des Bildes an einem Pixel $\mathbf{x} \in \mathbf{X}$ an.¹

Farbmerkmalbildung mit Histogrammen Die Farbverteilung der ROI \mathbf{I}_O soll zunächst durch ein *Farbhistogramm* modelliert werden. Ein Farbhistogramm $H_F[\mathbf{O}]$ ist eine Transformation des Bildausschnittes \mathbf{I}_O , welche die absoluten Auftretshäufigkeiten der Farbwerte ξ von \mathbf{I}_O innerhalb definierter Intervalle ξ^i des Farbattributes zählt, $H_F[\mathbf{O}] : \{\xi^i\} \mapsto N$. Ein Intervall wird bei Histogrammen oft auch als Behälter oder englisch *Bin* bezeichnet. Zur Anschauung des Histogrammbegriffs ist in Abbildung 1.7 das Farbhistogramm $H_F[\mathbf{O}]$ des bekannten

¹Man beachte: Der Vektor \mathbf{x} steht für den Ortsvektor zu einem Bildpixel, nicht für dessen Farbwert.

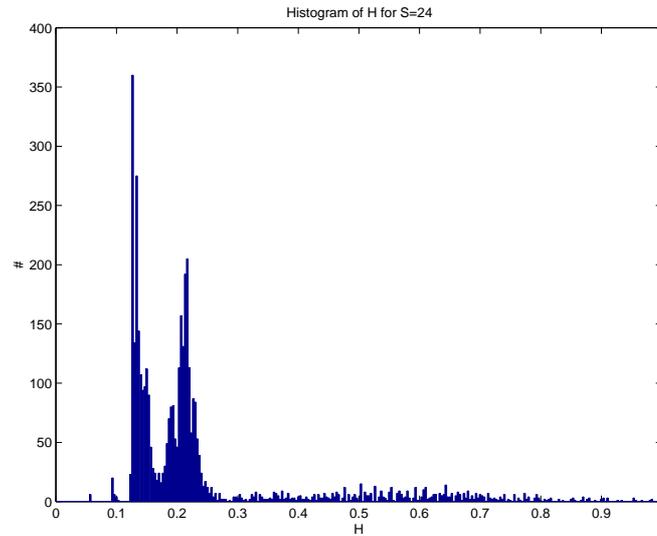


Abbildung 1.7: Farbhistogramm des Farbwerts H für ein festes $S_0 = 0.1$ des Beispielfarbbildes der Abbildung 1.6. In diesem Beispiel sind 300×300 Bins gewählt worden.

Beispielfarbbildes illustriert.

Ein Problem bei der Histogrammbildung läßt sich an Abbildung 1.7 gleich erkennen: beim Übergang zwischen benachbarten Binsäulen können große Unstetigkeiten auftreten. Diese Unstetigkeiten spiegeln in der Regel nicht die reale Situation in der Bildszene wieder, sondern sind eine Folge von einem Mißverhältnis von der Breite der Bins und der Anzahl der zur Verfügung stehenden Daten. Eine Möglichkeit zur Lösung dieses Problems würde z. B. ein Verfahren geben, welches die Binbreite adaptiv einstellt. In diesem Praktikum soll jedoch, dem Aufsatz [7] folgend, ein Ansatz beschrieben werden, der eine andere Idee aufgreift, nämlich die Farbverteilungsmodellierung mit *parametrischen Verteilungsdichtefunktionen* anstelle von Histogrammen.

Farbmerkmalbildung mit Verteilungsdichtefunktionen Im Gegensatz zur Verteilungsmodellierung mit Histogrammen wird bei der Modellierung durch Verteilungsdichtefunktionen die Farbverteilung nicht durch Angabe von diskreten Häufigkeitswerten an bestimmten Stützstellen ξ^i des Farbattributes, sondern als parametrische kontinuierliche Verteilungsdichtefunktion geschätzt.

Die hier verwendete Klasse von Verteilungsdichtefunktionen wird durch die *Gauß'schen Mischverteilungen* gebildet. Eine Erklärung dieses Modelles wird im Anhang 1.5 gegeben. An dieser Stelle soll dem Anhang vorgegriffen und angenommen werden, daß die Farbverteilung des Objektes \mathbf{I}_O durch eine parametrische Verteilungsdichtefunktion $p_F[\mathbf{O}](\xi)$ anstelle eines Histogramms $H_F[\mathbf{O}](\xi)$ modelliert ist.

Die Schätzformel für die Bewegungsparameter

In den vorigen Abschnitten sind die Extraktion der ROI und die Merkmalbildung beschrieben worden. Ein Blick auf Abbildung 1.5 zeigt, daß nun für eine vollständige Beschreibung des Verfahrens eine—explizite—Schätzvorschrift für die Bewegungsparameter \mathbf{T} angegeben werden muß.

Die für dieses Praktikum interessierenden Bewegungsparameter ist genau wie in Kapitel 1.2.1, der Translationsvektor $(t_x, t_y)^T$ des zu verfolgenden Objektes.

Es soll angenommen werden, daß die Lage des Objektes zum Zeitpunkt t durch den Ortsvektor \mathbf{m}_t beschrieben wird (was \mathbf{m}_t genau bedeutet, siehe unten). Anhand dieser Größe schlägt der Artikel [7] als Berechnungsformel

$$\mathbf{T} = (t_x, t_y)^T = \mathbf{m}_t - \mathbf{m}_{t-1} = \frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) (\mathbf{x} - \mathbf{m}_{t-1})}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})}, \quad (1.4)$$

vor, wobei \mathbf{x} die Koordinaten der Bildpixel und $p(\xi_{\mathbf{x}})$ die Mischverteilungsdichte der ROI $\mathbf{I}_{\mathbf{O}}$ für die Koordinate \mathbf{x} bezeichnet, welche durch die im vorigen Abschnitt erläuterte Methode geschätzt wurde.

Eine physikalische Anschauung für \mathbf{m}_t wird offensichtlich, wenn Gleichung (1.4) umgeformt wird:

$$\begin{aligned} \mathbf{m}_t &= \frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) (\mathbf{x} - \mathbf{m}_{t-1})}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})} + \mathbf{m}_{t-1} \\ &= \frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) \mathbf{x}}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})} - \frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) \mathbf{m}_{t-1}}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})} + \mathbf{m}_{t-1} \\ &= \frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) \mathbf{x}}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})} \end{aligned} \quad (1.5)$$

Ein Vergleich mit der physikalischen Formel für die Schwerpunktberechnung

$$\mathbf{s} = \frac{\sum_{\mathbf{x}} m_{\mathbf{x}} \mathbf{x}}{\sum_{\mathbf{x}} m_{\mathbf{x}}}$$

zeigt, daß sich die Objektkoordinaten \mathbf{m}_t als *Massenschwerpunkt* des vorliegenden Bildbereichs interpretieren lassen, wobei die Bedeutung der Punktmasse $m_{\mathbf{x}}$ durch die Wahrscheinlichkeitsdichte $p(\xi_{\mathbf{x}})$ des Farbwertes an der Koordinate \mathbf{x} übernommen wird.²

Aufgabe 2: Zum Verständnis von Gleichung (1.5) betrachte man Abbildung 1.8: Teilabbildung a) soll für die folgenden Überlegungen als (ein zugegebenermaßen konstruiertes) Objektbild $\mathbf{I}_{\mathbf{O}}$ und Teilabbildung b) als „Suchbild“ $\mathbf{I}_{\mathbf{T}}$ betrachtet werden. Im Unterschied zu den vorangegangenen Überlegungen wird als Merkmal im Folgenden der Grauwert anstatt das H-S-Farbraummerkmal zugrundegelegt.

²Die Masse $m_{\mathbf{x}}$ ist nicht zu verwechseln mit den Objektkoordinaten \mathbf{m}_t .

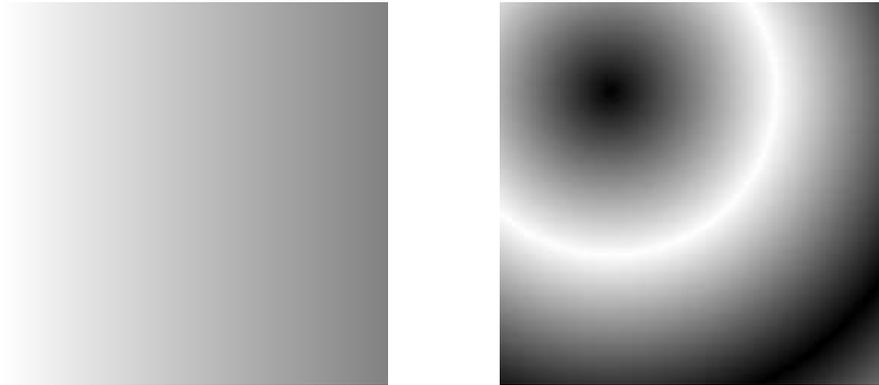


Abbildung 1.8: Zu Aufgabe 5: a) Objektbild \mathbf{I}_O b) transformiertes Bild \mathbf{I}_T

1. Man skizziere die Wahrscheinlichkeitsdichte $p_{\text{Grauwert}}[\mathbf{O}](\xi)$ für die Grauwerte $[0, \dots, 255]$ bezüglich \mathbf{I}_O .
2. Man skizziere die „Massenbelegung“ $p(\xi_{\mathbf{x}})$ für \mathbf{I}_T (z.B. durch Konturlinien) und zeichne ungefähr den „Massenschwerpunkt“ \mathbf{m} ein.

Anmerkungen

Im folgenden sollen noch einige Anmerkungen gegeben werden, durch welche Maßnahmen das beschriebene Schätzverfahren universeller und robuster gestaltet werden kann.

Anpassung des Referenzmodelles In den zuvor beschriebenen Erläuterungen wird das Referenzmodell \mathbf{M} (siehe Abbildung 1.5) nur ein einziges Mal am Anfang der Objektverfolgung gebildet. Dieses Vorgehen führt zu Problemen, wenn über die Zeitdauer der Verfolgung Veränderungen in der Umgebung, z. B. in der Beleuchtung auftreten. Daher geben die Autoren in [7] eine Methode zur zeitlichen Adaption des Modells an.

Schätzformel für Größenänderungen Neben der Schätzformel für den Translationsvektor $\Delta \mathbf{m}$ (1.4) kann eine weitere Schätzformel für die Objektgröße σ durch Berechnung der Standardabweichung aufgestellt werden,

$$\sigma = \sqrt{\frac{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}}) \{(\mathbf{x} - \mathbf{m}_{t-1}) - \mathbf{m}_t\}^2}{\sum_{\mathbf{x}} p(\xi_{\mathbf{x}})}}.$$

Anhand dieser Information kann z. B. Rückschluß auf Tiefenbewegungen gezogen werden, worauf man zusätzlich zur Kameraorientierung die Zoomparameter so einstellen kann, daß das Bild des Objekts stets eine gleiche Größe besitzt.

1.2.3 Vergleich der Verfahren zur Bewegungsschätzung

Merkmalextraktion

Ein weiterer wesentlicher Unterschied zwischen den beiden Ansätzen liegt in der Merkmalextraktion: das implizite Verfahren identifiziert das zu verfolgende Objekt ohne eine solche, wohingegen das explizite Verfahren durch die vorgeschaltete Merkmalextraktion ausschließlich Informationen der Objektfarbverteilung ausnutzt. Es wurde in Abschnitt 1.2.2 bereits diskutiert, daß dabei Information über die Objektstruktur verloren geht. Die Rolle der Merkmalextraktion soll nun unter zwei Aspekten näher beleuchtet werden: unter dem Begriff der *Äquivalenzklasse* und der Robustheit gegenüber *Rotationen*.

Zunächst zu den Äquivalenzklassen: Eine Folge der Merkmalextraktion bei dem expliziten Verfahren ist, daß neben dem zu verfolgenden Objekt \mathbf{M} ein anderes Objekt \mathbf{M}' die Bestimmung der Bewegungsparameter stören kann, sofern es die gleichen Merkmale besitzt. Ein Beispiel hierfür soll in Aufgabe 5 überlegt werden. Für den Fall gleicher Merkmale von \mathbf{M} und \mathbf{M}' spricht man davon, daß \mathbf{M} und \mathbf{M}' in eine Äquivalenzklasse fallen.

Aufschlußreich ist auch die Untersuchung der Merkmalextraktion bezüglich des zweiten Aspektes, der Rotation. Die Merkmalextraktion nimmt bei der expliziten Schätzmethode eine entscheidende Rolle im Umgang mit—in diesem Praktikum bisher nicht berücksichtigten—Rotationen ein. Durch die Rotationsinvarianz der Farbverteilungsmodellierung gewinnt das Verfahren die praktisch nützliche Invarianzeigenschaft gegenüber Rotationen. Ein Nachteil dieser Invarianz ist jedoch, daß keine explizite Bestimmung der Rotationsparameter, die z. B. für die Nachstellung eines Orientierungsparameters der Kamera benötigt würde, erfolgen kann.

Das hier vorgestellte implizite Verfahren berücksichtigt keine Rotation. Eine Erweiterung hierauf würde, wie in Abschnitt 1.2.1 angesprochen, eine Anpassung des Bewegungsmodells $\mathbf{h}(\mathbf{x}, \mathbf{T})$ erfordern. Ähnlich zu der Berechnung der Translationsparameter würden die Rotationsparameter anschließend durch die Minimierung der Fehlerfunktion bestimmt werden. Dieses Vorgehen hat im Vergleich zu dem expliziten, merkmalbasierten Ansatz den Vorteil einer höheren Flexibilität in der Auswahl der zugelassenen Bewegungstransformationen.

Aufgabe 3: Welche Art von Objekten \mathbf{O}' fallen durch die Merkmalbildung mit Farbverteilungen fälschlich in die Äquivalenzklasse eines zu verfolgenden Objektes \mathbf{O} ?

Aufgabe 4: Wie reagieren die beiden Verfahren bei der Bestimmung der Bewegungsparameter wohl darauf, wenn ein weitere Objekte \mathbf{O}'' , \mathbf{O}''' , \mathbf{O}'''' in das Bild gebracht wird. Diese Überlegung soll für folgende Nebenbedingungen angestellt

werden:

1. \mathbf{O}'' unterscheidet sich von \mathbf{O} nur im Farbhistogramm, jedoch nicht in den Flächenverteilungen.
2. \mathbf{O}''' unterscheidet sich von \mathbf{O} nur Anordnung der Farbflächenverteilungen, nicht aber im Farbhistogramm.
3. \mathbf{O}'''' unterscheidet sich von \mathbf{O} in der Farbverteilung als auch in der Struktur.

Berechnungsaufwand

Bei einem Vergleich der beiden Verfahren spielt natürlich auch der Berechnungsaufwand eine Rolle. Im folgenden soll die Berechnungskomplexität für die vorgestellten Ansätze untersucht werden. Es wird stets davon ausgegangen, daß die Parameterschätzung auf einem Bild der Größe $M \cdot N$ durchgeführt wird.

Ergänzend zur Berechnungskomplexität findet eine Untersuchung der absoluten Berechnungszeit in den Versuchsaufgaben (Kapitel 1.4) statt.

Berechnungskomplexität des expliziten Verfahrens Bei dem expliziten Parameterschätzverfahren muß zwischen

1. dem einmalig vollzogenen Vorverarbeitungsschritt und
2. der wiederholt berechneten Schätzvorschrift Gleichung (1.4)

differenziert werden.

Der Vorverarbeitungsschritt besteht aus der ROI-Extraktion und der Aufstellung des Farbverteilungsmodells mit Gauß'schen Mischverteilungen. Da die ROI-Extraktion von Hand ausgeführt wird, soll sie bei der Aufwandsabschätzung nicht berücksichtigt werden. Der Rechenaufwand für die Bestimmung der Gauß'schen Mischverteilungsparameter berechnet sich pro EM-Iterationsschritt (siehe Abschnitt 1.5) gemäß Gleichungen (1.12)–(1.15) zu $O(d \cdot R)$ mit $d = \dim(\text{Merkmalvektor}) = 2$ und R als Parameter für die Anzahl der ROI-Bildpunkte. Abhängig von der Iterationsanzahl K des EM-Algorithmus ergibt sich als Gesamtberechnungsaufwand für den Vorverarbeitungsschritt

$$O(K \cdot d \cdot R)$$

Eine Abschätzung der Laufzeitkomplexität für den Parameterschätzschritt soll in Aufgabe 8 vorgenommen werden.

Berechnungskomplexität des impliziten Verfahrens Für das implizite Verfahren ist keine Vorverarbeitung notwendig. Es braucht also nur der Schätzalgorithmus bei der Betrachtung des Berechnungsaufwandes berücksichtigt werden. Dies soll ebenfalls in Aufgabe 8 bearbeitet werden.

Aufgabe 5: Man gebe die Berechnungskomplexität für

1. die Schätzvorschrift des expliziten Verfahrens und
2. den Schätzalgorithmus des impliziten Verfahrens an.



Abbildung 1.9: Die verwendete Tracking-Farbkamera *Sony EVI-D31*. Viele Kameraparameter lassen sich mit einer Infrarot Fernbedienung als auch über die serielle Schnittstelle eines Computers einstellen.

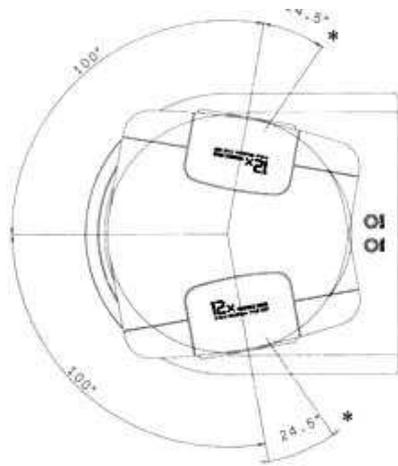
1.3 Versuchsaufbau

Die Kamera: Das „Herzstück“ des Praktikums ist die Farbkamera *Sony EVI-D31*. Abbildung 1.9 zeigt ein Foto. Die EVI-D31 ist eine in zwei Freiheitsgraden schwenkbare Trackingkamera mit eingebauter Objektverfolgungsoption.³ Zu einer Integration in ein aktives Sehsystem lassen sich verschiedene Kameraparameter einstellen bzw. auch abfragen; dies sind u.a.:

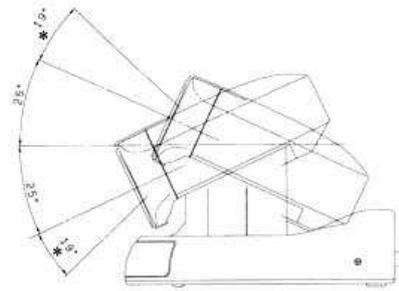
- der seitliche Schwenkwinkel (*Pan-Winkel*) mit einem einstellbaren Bereich von -100° bis $+100^\circ$ (siehe Abbildung 1.10 (a))
- der Hoch-Tief-Schwenkwinkel (*Tilt-Winkel*) mit einem einstellbaren Bereich von -25° bis $+25^\circ$ (siehe Abbildung 1.10 (b))
- der 12-fach Zoom
- der Fokus

Eine Einstellung dieser Größen kann sowohl mit Hilfe einer Infrarot-Fernsteuerung als auch über die serielle RS232C-Schnittstelle eines Computers durch das VISCA (Video System Control Architecture) Protokoll erfolgen. Für das Praktikum stehen *Matlab*-Funktionen zur Verfügung, die von letzterer Methode Gebrauch machen.

³Diese wird bei der Versuchsdurchführung ausgeschaltet und soll durch die in Abschnitt 1.2 behandelten Verfolgungsmethoden ersetzt werden.



(a) Pan Schwenkbereich



(b) Tilt Schwenkbereich

Abbildung 1.10: Die Bewegungsfreiheitsgrade der EVI-D31.

1.4 Praktikumsaufgaben

1.4.1 Aufgaben zum Impliziten Verfahren

In dieser Aufgabe soll das in der Anleitung vorgestellte implizite Verfahren zur Bewegungsschätzung näher kennengelernt werden.

Aufgabe 1: Schneiden Sie ein Gesicht aus und korrelieren Sie es mit dem aktuellen Kamera-Bild. Stellen Sie die Korrelationsfunktion live als surface-plot dar.

TIP: Damit die Geschwindigkeit der Korrelation erträglich wird, skalieren Sie die Bilder etwa um den Faktor 8 herunter.

Aufgabe 2: Zeichnen Sie in das aktuelle Kamera-Bild an die Stelle der Maximalen Korrelation ein Kreuz. d.h. das Kreuz sollte dann immer etwa auf der Nasenspitze sitzen.

TIP: Das Kreuz zeichnet man am einfachsten, indem man die gesamte Spalte und die gesamte Zeile zur gefundenen Position auf 255 setzt.

Aufgabe 3: Führen Sie die Kamera dem Gesicht hinterher.

1.4.2 Aufgaben zum Expliziten Verfahren

Aufgabe 4: Transformieren Sie das aktuelle Bild in den HSV Farbraum. Berechnen Sie ein 2D Histogramm (dafür gibt es keine Matlab-Funktion) aus den "hue" und "saturation" Werten und stellen es live als surface-plot dar. Beobachten Sie, wie sich das Histogramm ändert, wenn verschiedene Objekte in die Kamera gehalten werden. Auch hier empfiehlt es sich, das Bild herunterzuskalieren.

TIP: Beachten Sie beim Berechnen des Histogramms, daß bei Matlab das erste Element eines Arrays den Index 1 (und nicht 0) hat

Aufgabe 5: Berechnen Sie das Histogramm eines hautfarbenen Teilbildes. Nutzen Sie dieses Histogramm um alle hautfarbenen Bereiche des aktuellen Kamerabildes entsprechend hervorzuheben.

Aufgabe 6: Nutzen Sie statt dem Histogramm die Gauß'sche Mischverteilung um die hautfarbenen Bereiche zu lokalisieren.

1.5 Gauß'sche Mischverteilungen

Eine geeignete Klasse von Funktionen für die Approximation von Verteilungsdichtefunktionen (und auch allgemeiner Funktionen) stellen die *Gauß'schen Mischverteilungen* dar, welche eine lineare Superposition von Gauß'schen Normalverteilungen darstellt. Eine wichtige Eigenschaft von Gauß'schen Mischverteilungen ist, daß durch sie eine beliebige Verteilungsdichte beliebig genau approximiert werden kann, vorausgesetzt das Modell hat eine genügend große Anzahl von Gauß'schen Normalverteilungskomponenten und die Modellparameter der jeweiligen Komponenten sind korrekt gewählt.

Die nun erfolgende Beschreibung von Gauß'schen Mischverteilungen orientiert sich an den Ausführungen des Buches [2].

1.5.1 Das Modell

Gauß'sche Mischverteilungen, welche im weiteren, der Notation aus [2] folgend, mit $p(\mathbf{x})$ bezeichnet werden, sind eine lineare Kombination von sogenannten *Verteilungskomponenten* $p(\mathbf{x}|j)$, wobei $p(\mathbf{x}|j)$ durch sphärisch invariante Gaußverteilungen, d.h. durch „runde Gaußglocken“ ohne Vorzugsrichtung, modelliert werden:⁴

$$p(\mathbf{x}) = \sum_{j=1}^M P(j) p(\mathbf{x}|j) \quad (1.6)$$

$$p(\mathbf{x}|j) = \frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left\{-\frac{\|\mathbf{x} - \mu_j\|^2}{2\sigma_j^2}\right\} \quad (1.7)$$

Der Parameter d steht in diesem Fall für $\dim(\mathbf{x})$. Die Koeffizienten $P(j)$ beschreiben die *Mischparameter*. Für sie gelten die Nebenbedingungen

$$\sum_{j=1}^M P(j) = 1, \quad 0 \leq P(j) \leq 1.$$

Man kann sich Gauß'sche Mischverteilungen anschaulich als einen zweistufigen Zufallsprozeß vorstellen. Um einen Datenpunkt gemäß (1.6) zu erzeugen, wird

⁴Die Forderung nach sphärischer Invarianz wird an dieser Stelle nur gestellt, damit sich die nachfolgenden Rechnungen einfacher gestalten. Sie ist bei den Gauß'schen Mischverteilungen nicht zwingend.

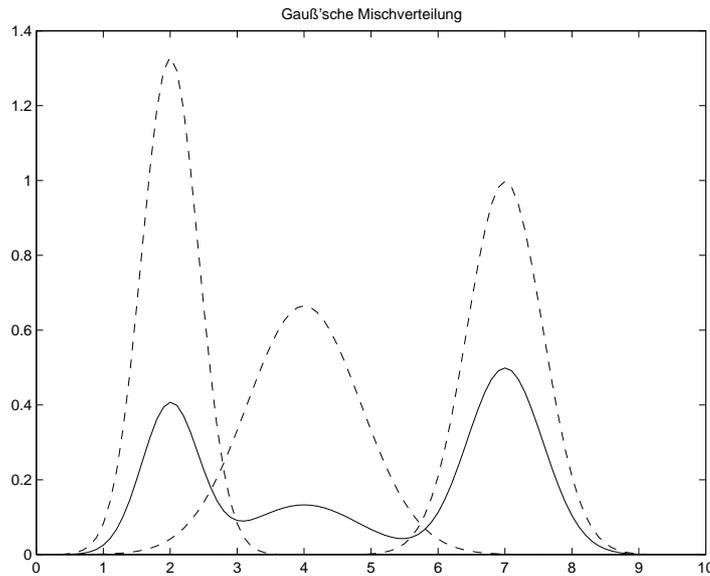


Abbildung 1.11: Beispiel einer eindimensionalen, dreikomponentigen Gauß'schen Mischverteilung. Sie wird mit der durchgezogenen Linie dargestellt. Die gewählten Parameter lauten $(\mu_1, \mu_2, \mu_3) = (2, 4, 7)$, $(\sigma_1, \sigma_2, \sigma_3) = (0.3, 0.6, 0.4)$ und $(P(1), P(2), P(3)) = (0.3, 0.2, 0.5)$. Die einzelnen Gaußkomponenten sind mit gestrichelten Linien eingezeichnet.

zunächst entsprechend $P(j)$ eine erzeugende Verteilungskomponente j und anschließend entsprechend der j -ten Verteilungsdichte $p(\mathbf{x}|j)$ ein Datenpunkt erzeugt. Den Mischparametern $P(j)$ geben in diesem Zusammenhang die *a-priori*-Wahrscheinlichkeiten der Verteilungskomponenten j an.

Ein Beispiel für eine eindimensionale, dreikomponentige Gauß'sche Mischverteilungsfunktion ist in Abbildung 1.11 illustriert.

Für unsere Anwendung kann man sich das bildlich folgendermaßen vorstellen: Die einzelnen Gaußkomponenten entsprechen dabei den Farbverteilungen von verschiedenen Regionen im Bild. In Abb. 1.12 könnte die Farbverteilung in den Haaren mit einer ersten Gaußverteilung und die Farbverteilung im Gesicht mit einer zweiten Gaußverteilung modelliert werden. Die Gesamtfarbverteilung des Bildes ist dann einfach die Summe dieser beiden Verteilungen.

Die prinzipielle Idee von Gauß'schen Mischverteilungen ist soweit erläutert. Ein wichtiger noch zu klärender Aspekt bleibt der, wie zu einer Modellierung einer Verteilungsfunktion geeignete Parameter geschätzt werden. Eine ausführliche Behandlung dieses Themas würde den Rahmen dieses Praktikums jedoch sprengen. Im folgenden soll dennoch eine kompakte und anschauliche Motivation gegeben werden.

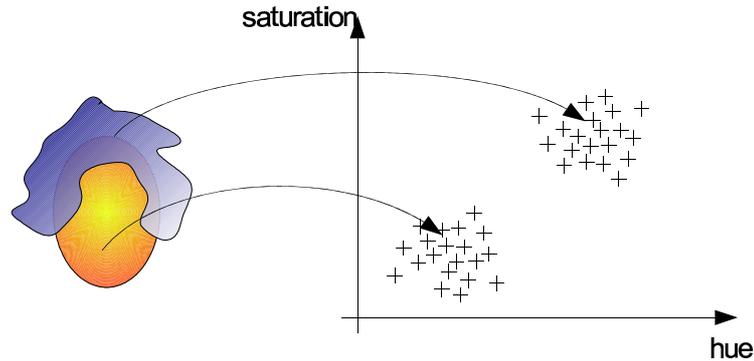


Abbildung 1.12: Modell für die Farbverteilung in einem Bild. Sowohl die Verteilung der Farben im Haar als auch die Verteilung der Farben im Gesicht könnte durch jeweils eine Gaußverteilung modelliert werden.

1.5.2 Der EM-Schätzalgorithmus für die Verteilungsparameter

Ein Blick auf Gleichungen (1.6) und (1.7) zeigt, daß es sich bei den zu schätzenden Parametern um μ_j , σ_j und $P(j)$, $j = 1, \dots, M$ handelt.

Diese können in einen *Maximum-Likelihood-Ansatz* (*MLE*) durch Maximierung der Verbundwahrscheinlichkeit

$$\mathcal{L}(\mu_j, \sigma_j, P(j)) := \prod_{n=1}^N p(\mathbf{x}^n | \mu_j, \sigma_j, P(j))$$

der Datenbasis \mathbf{x}^n , $n = 1, \dots, N$ unter Variation von μ_j , σ_j und $P(j)$ geschätzt werden. Das Ergebnis der Maximum-Likelihood-Schätzung kann zu

$$\hat{\mu}_j = \frac{\sum_{n=1}^N P(j|\mathbf{x}^n) \mathbf{x}^n}{\sum_{n=1}^N P(j|\mathbf{x}^n)} \quad (1.8)$$

$$\hat{\sigma}_j^2 = \frac{1}{d} \frac{\sum_{n=1}^N P(j|\mathbf{x}^n) (\mathbf{x}^n - \hat{\mu}_j)^2}{\sum_{n=1}^N P(j|\mathbf{x}^n)} \quad (1.9)$$

$$\hat{P}(j) = \frac{1}{N} \sum_{n=1}^N P(j|\mathbf{x}^n) \quad (1.10)$$

errechnet werden. Die Größe $P(j|\mathbf{x}^n)$ steht dabei für die *a-posteriori* Wahrscheinlichkeit, daß ein Datenpunkt \mathbf{x}^n , $n = 1, \dots, N$ von der Verteilungskomponente j erzeugt wurde.

Eine mathematische Herleitung der Gleichungen (1.8)–(1.10) wird in [2] gegeben. Die Schätzformeln lassen sich jedoch auch intuitiv nachvollziehen. Gleichung (1.8) entspricht der Anschauung, daß der Mittelwert der Komponente j durch den

Mittelwert der Datenpunkte gegeben ist, gewichtet jeweils mit der a-posteriori Wahrscheinlichkeit, daß der Datenpunkt durch die Komponente j erzeugt wurde. Entsprechende Überlegungen lassen sich für die anderen beiden Gleichungen anstellen.

Für den Spezialfall einer einkomponentigen Gauß'schen Verteilung reduzieren sich (1.8)–(1.10) auf die bekannten Schätzformeln

$$\begin{aligned}\hat{\mu}_1 &= \frac{1}{N} \sum_{n=1}^N \mathbf{x}^n \\ \hat{\sigma}_1 &= \frac{1}{dN} \sum_{n=1}^N (\mathbf{x}^n - \hat{\mu}_1)^2 \\ \hat{P}_1 &= 1\end{aligned}$$

Insoweit bieten die Gleichungen (1.8)–(1.10) zwar eine anschaulich einsichtige Lösung des Maximum-Likelihood-Ansatzes, eine explizite Berechnungsvorschrift für die Verteilungsparameter liefern sie aber nicht. Sie bilden nämlich ein System hochgradig nichtlinear gekoppelter Gleichungen, da die zu berechnenden Größen μ_j , σ_j und $P(j)$ implizit auf den rechten Seiten der Gleichungen vorkommen. Die benötigten a-posteriori Wahrscheinlichkeiten $P(j|\mathbf{x}^n)$ berechnen sich nach Bayes' Theorem nämlich zu

$$P(j|\mathbf{x}^n) = \frac{p(\mathbf{x}^n|j)P(j)}{p(\mathbf{x}^n)} = \frac{p(\mathbf{x}^n|j)P(j)}{\sum_{j=1}^M p(\mathbf{x}^n|j)P(j)}. \quad (1.11)$$

Diese Situation führt auf das bekannte Problem des Lösen einer impliziten Gleichung: Die Schätzung der Größen μ_j , σ_j und $P(j)$ hängt von den a-posteriori Wahrscheinlichkeiten $P(j|\mathbf{x}^n)$ ab, welche wiederum nur unter Kenntnis von μ_j , σ_j und $P(j)$ berechnet werden können. Abbildung 1.13 zeigt diesen Zusammenhang bildlich.

Gleichungen (1.8)–(1.11) legen jedoch ein iteratives Schema nahe: Angenommen, eine Anfangsbelegung für die Verteilungsparameter ist festgelegt worden, bezeichnet mit μ_j^{old} , σ_j^{old} und $P(j)^{\text{old}}$. Mit diesen Werten können dann Schätzungen für eine „neue“ Belegung der Parameter, μ_j^{new} , σ_j^{new} und $P(j)^{\text{new}}$, unter Berücksichtigung der Gleichungen (1.8)–(1.11), vorgenommen werden. Die neu errechneten Parameter übernehmen anschließend die Funktion der „alten“ Parameter, und der Iterationsschritt wird wiederholt.

In [2] wird gezeigt, daß in jedem Schritt die Größe $\mathcal{L}(\mu_j^{\text{new}}, \sigma_j^{\text{new}}, P(j)^{\text{new}})$, also das Maß für die Wahrscheinlichkeit der aktuellen Parameterbelegung ansteigt. Treten in einem Iterationsschritt keine bedeutenden Parameteränderungen mehr auf, so

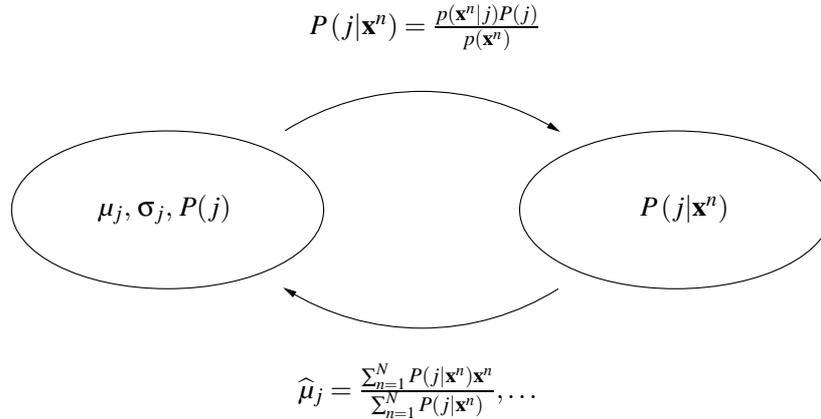


Abbildung 1.13: Das Problem der impliziten Gleichungen bei der Maximum-Likelihood-Schätzung der Mischverteilungsparameter: Es tritt eine gegenseitige Abhängigkeit der benötigten Größen auf. $\mu_j, \sigma_j, P(j)$ einerseits und $P(j|\mathbf{x}^n) = \frac{p(\mathbf{x}^n|j)P(j)}{p(\mathbf{x}^n)}$ andererseits auf. Der EM-Algorithmus löst dieses Problem auf, indem nach einer geeigneten Initialisierung in einem iterativen Schema die beiden Größen abwechselnd berechnet werden, bis ein lokales Maximum der Maximum-Likelihood Kriteriums $\mathcal{L}(\mu_j, \sigma_j, P(j))$ gefunden wurde.

wird die Iteration beendet und man betrachtet die aktuellen Parameterbelegungen als die gesuchte Lösung. Diese entsprechen dann einem lokalen Maximum für \mathcal{L} .

Zusammengefaßt lautet die Berechnungsvorschrift für die „neuen“ Schätzwerte $\mu_j^{\text{new}}, \sigma_j^{\text{new}}$ und $P(j)^{\text{new}}$ unter Berücksichtigung von (1.8)–(1.11) also:

$$\mu_j^{\text{new}} = \frac{\sum_{n=1}^N P^{\text{old}}(j|\mathbf{x}^n) \mathbf{x}^n}{\sum_{n=1}^N P^{\text{old}}(j|\mathbf{x}^n)} \quad (1.12)$$

$$(\sigma_j^{\text{new}})^2 = \frac{1}{d} \frac{\sum_{n=1}^N P^{\text{old}}(j|\mathbf{x}^n) (\mathbf{x}^n - \mu_j^{\text{new}})^2}{\sum_{n=1}^N P^{\text{old}}(j|\mathbf{x}^n)} \quad (1.13)$$

$$P(j)^{\text{new}} = \frac{1}{N} \sum_{n=1}^N P^{\text{old}}(j|\mathbf{x}^n) \quad (1.14)$$

wobei

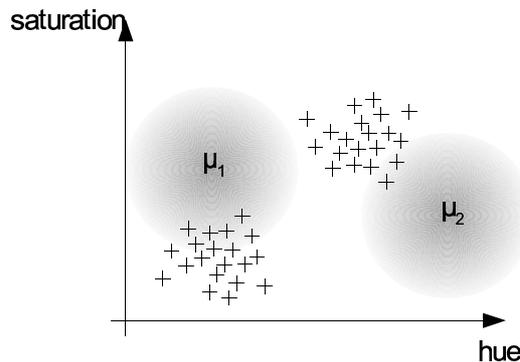
$$P^{\text{old}}(j|\mathbf{x}^n) = \frac{p^{\text{old}}(\mathbf{x}^n|j)P(j)^{\text{old}}}{p^{\text{old}}(\mathbf{x}^n)} \quad (1.15)$$

Der hier beschriebene Algorithmus zur Bestimmung der Mischverteilungsparameter ist eine Anwendung eines allgemeinen Verfahrens, welches in der Literatur unter dem Namen „EM-Algorithmus“ bekannt ist. Die beiden Buchstaben EM bezeichnen die zwei wesentlichen iterativen Schritte des Verfahrens, namentlich Expectate und Maximize. Der *Expectate*-Schritt bezieht sich dabei auf die Bildung des Erwartungswertes (engl. expectation value) für das Wahrscheinlichkeits-

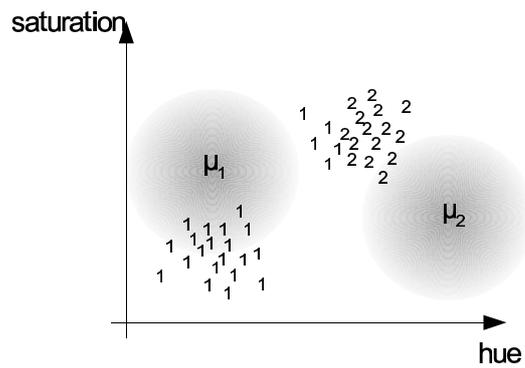
maß $\mathcal{L}(\mu_j, \sigma_j, P(j))$ der Parameterbelegung $\mu_j^{\text{old}}, \sigma_j^{\text{old}}$ und $P(j)^{\text{old}}$; der *Maximize*-Schritt auf die Maximum-Likelihood Schätzung zur Berechnung neuer Schätzwerte $\mu_j^{\text{new}}, \sigma_j^{\text{new}}$ und $P(j)^{\text{new}}$, welche sich in Gleichungen (1.12)–(1.15) ausdrückt. Der EM-Algorithmus begegnet dem interessierten Mustererkenner auch in anderen Aufgabenfeldern. Beispiele hierfür sind Vektor-Quantisierung oder die Parameterschätzung von Hidden-Markov-Modellen.

Für das obige Beispiel mit dem Bild des Kopfs sei hier der Ablauf des EM-Verfahrens noch einmal beispielhaft dargestellt: Gegeben ist die gemessene Verteilung der Farben (also alle Kreuzchen im hue/saturation Diagramm). Die Aufgabe ist nun zwei Gaußverteilungen an diese Daten anzupassen, so dass der Fehler minimal wird. Das Problem ist, dass wir anfangs weder die Parameter der beiden Gaussverteilungen μ_1, σ_1 und μ_2, σ_2 kennen, noch wissen, welcher Datenpunkt zur ersten, und welcher Datenpunkt zur zweiten Gaußverteilung "gehört". Diese "Zugehörigkeit" wird durch die a-posteriori-Wahrscheinlichkeit $P(j|\mathbf{x}^n)$ für jeden Datenpunkt n zur Verteilung j definiert. Die Anpassung läuft dann folgendermaßen ab:

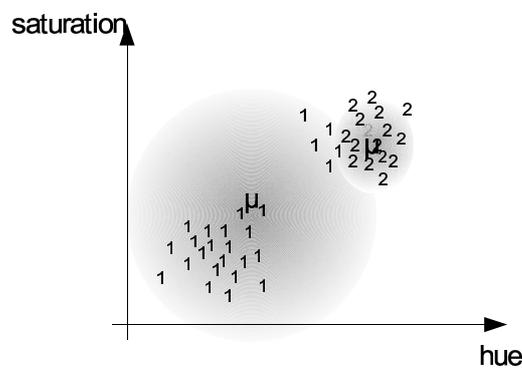
1. Für die beiden Gaußverteilungen werden zufällige Parameter μ_1, σ_1 und μ_2, σ_2 gewählt



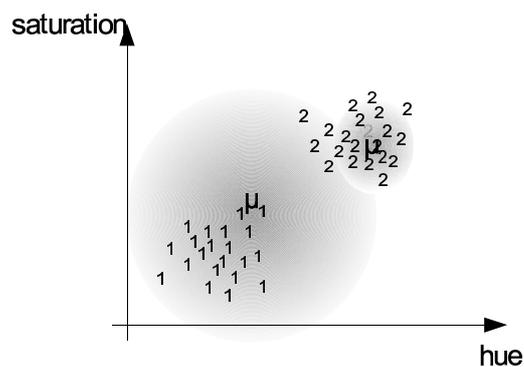
2. Für jeden Datenpunkt wird die Wahrscheinlichkeit $P(j|\mathbf{x}^n)$ geschätzt, mit der er zur ersten oder zur zweiten Gaußverteilung gehört (zur einfacheren Darstellung wird hier nur die wahrscheinlichere als Label an den Punkt geschrieben)



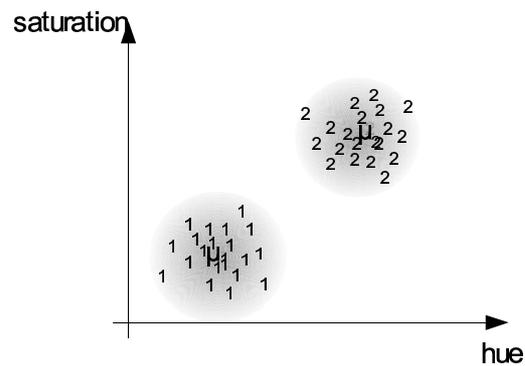
3. Anhand der geschätzten Zugehörigkeit für jeden Datenpunkt zur jeweiligen Gaussverteilung können jetzt die Parameter μ_1, σ_1 und μ_2, σ_2 neu berechnet werden



4. Mit den neuen Gaußverteilungen werden wieder die Wahrscheinlichkeiten $P(j|\mathbf{x}^n)$ bestimmt.



5. Und wieder neue μ_1, σ_1 und μ_2, σ_2 berechnet, bis schließlich die optimal passende Verteilung gefunden ist



1.5.3 Bestimmung der Komponentenanzahl

In den bisherigen Ausführungen ist davon ausgegangen worden, daß die Anzahl der Mischkomponenten M , eine vom Benutzer festzulegende Größe ist. Die Autoren von [7] geben ein adaptives Verfahren an, mit dem dies automatisch erfolgt. In der Grundfassung inkrementiert es die Größe M von der Belegung $M = 1$ startend so lange, bis kein Anstieg der Verbundwahrscheinlichkeit \mathcal{L} mehr zu erkennen ist.

Literaturverzeichnis

- [1] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48(1):57–86, February 1991.
- [2] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [3] H. Burkhardt. Digitale Bildverarbeitung II. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [4] H. Burkhardt. Mustererkennung. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [5] N. Diehl. *Methoden zur allgemeinen Bewegungsschätzung in Bildfolgen*. PhD thesis, Technische Universität Hamburg-Harburg, März 1988. Ersch. als Fortschrittbericht (Reihe 10, Nr. 92) der VDI-Zeitschriften, VDI-Verlag.
- [6] G. H. Golub and J. M. Ortega. *Wissenschaftliches Rechnen und Differentialgleichungen*, volume 6 of *Berliner Studienreihe zur Mathematik*. Heldermann Verlag, Berlin, 1995.
- [7] Yogesh Raja, Stephen J. McKenna, and Shaogang Gong. Colour model selection and adaptation in dynamic scenes. In Hans Burkhardt and Bernd Neumann, editors, *Computer Vision—ECCV'98*, volume 1, pages 460–473. Springer, 1998.
- [8] Michael J. Swain and Markus A. Stricker. Promising directions in active vision. *International Journal of Computer Vision—Special Issue on Active Vision I*, 2, 1993.