

Wavelet-based Salient Points with Scale Information for Classification

Alexandra Teynor and Hans Burkhardt

Department of Computer Science, Albert-Ludwigs-Universität Freiburg, Germany
{teynor, Hans.Burkhardt}@informatik.uni-freiburg.de

Abstract

The calculation of local features at points of interest is a vital part of many current image retrieval and object detection systems. The wavelet-based interest point detector by Louprias et al. was especially developed for image retrieval applications. We show how the detector can be extended by a Laplacian scale selection mechanism to provide scale information and compare it to other state of the art detectors. The extended detector is very well suited for visual object class recognition using feature cluster histograms. It discovers a variety of image structures distributed over the entire image, and the number of regions obtained can be adjusted easily. These properties lead to superior performance, which we confirmed by tests on a difficult animal categorization problem.

1. Introduction

Current content based image retrieval (CBIR) and object class recognition systems use local representations of objects. The advantage of this can easily be seen: local representations can deal with variations in object shape and partial occlusions. Since later steps in the recognition chain heavily depend on the quality of the region detection step, special care has to be taken. Some detectors for example deliver a very sparse set of points, so that large areas of an image are not covered. Many current approaches for CBIR or visual object class recognition use histograms of feature clusters, typically with some 1000s of dimensions. If only few local features can be computed from an image, these histograms are very sparse and recognition performance might be spoiled. Nowak et al. [10] have shown that a sufficient number of interest points is crucial for a bag-of-features image classification.

The wavelet based interest point detector by Louprias et al. [5] was developed in the context of CBIR. The objective was to obtain sufficient interest points that do

not only detect image structures of a particular type, e.g. corners, but any “visually meaningful” structure. Moreover, the points should not only be gathered in small areas, but cover the whole image in order to represent the entire content. However, the basic detector does not provide any scale information, so features had to be extracted from fixed size windows around the point.

2. Related Work

Interest point detection has a long tradition in computer vision for finding point correspondences in image pairs to reconstruct 3D scenes from 2D views. Lindeberg shows in [4] how scale information can be assigned to different image structures using a scale space representation. The principles described there have been adopted for many current covariant region detectors, e.g., the Difference of Gaussian (DoG) detector [6] or the Harris-/Hessian-Laplace [8] detector. Our extension can also be seen in this tradition. Alternative methods for scale selection evaluate some stability criterion inside a region, e.g., the Salient Region detector by Kadir & Bradey [3].

3. Wavelet Based Salient Points

The wavelet based salient point detector by Louprias et al. [5] uses a Discrete Wavelet Transform (DWT) as a tool for the multi resolution analysis of the image. The image f is studied at a number of scales 2^{-j} , $j \in \mathbb{Z}^+$. The wavelet coefficients $W_{2^{-j}} f$ are obtained as the convolution of the image with the wavelet function dilated at scale 2^{-j} . Wavelet coefficients with high absolute value get traced over the scales. This is possible, since wavelets with compact support are used and for each coefficient, a region of support in the preceding decomposition stage can be determined. From the set of ancestors, the one with the highest absolute value is chosen and traced further. In the final stage, the gradient magnitudes are used to disambiguate between the pixels. The sum of all traced wavelet coefficients is used as the

4.2. Loupias-Laplace Detector

The original Loupias detector does not provide scale information. Although the DWT uses a multi scale representation of the image, it is not equivalent to a linear scale space representation, since in the DWT, we are restricted to orthogonal wavelets with compact support. The Gaussian kernel as needed for a linear scale space does not fulfill this condition. To still obtain scale information, we apply a LoG scale selection mechanism to the points provided by the original detector. The algorithm works as follows:

1. Apply the wavelet based interest point detector to the image. Obtain a set of M candidate points $\mathbf{S}_c = \{s_1, \dots, s_M\}$, $s = (x, y)$, by choosing the M points with the largest saliency values.
2. For each point $s_i \in \mathbf{S}_c$, build a scale space representation by evaluating a LoG filter at different scales t_j .
3. For each point s_i , search for the maximum filter response, and assign this t_j as scale for the interest point. If no maximum can be found, the interest point is discarded or a default scale assigned.

It is also possible to allow multiple maxima, when they are isolated within a certain scale range.

5 Evaluation

We compare the new detector to current state-of-the-art detectors, in particular the DoG detector [6], the Harris-Laplace detector [8], the Hessian-Laplace detector [8] and the Kadir & Bradey Salient Region detector [3]. For these detectors, executable programs provided by the authors were used, and the standard parameter settings used. Moreover, we compared the results to random detections in the image, in order to verify the suitability of interest point detectors. We tested two variants of the Loupias-Laplace detector. In the first version, whenever no distinct scale could be determined for a detection, the point was discarded (“Loupias-Laplace stable”), in the other version a default radius was used (“Loupias-Laplace all”). We also tested features calculated from the original Loupias detector with a fixed radius of 15 pixels (“Loupias fixed scale”). For all interest point detectors, regions smaller than 5 pixels were discarded, since we deem them to be too unstable for feature extraction.

The database used for the following tests is a challenging set of 1300 animal pictures belonging to 14

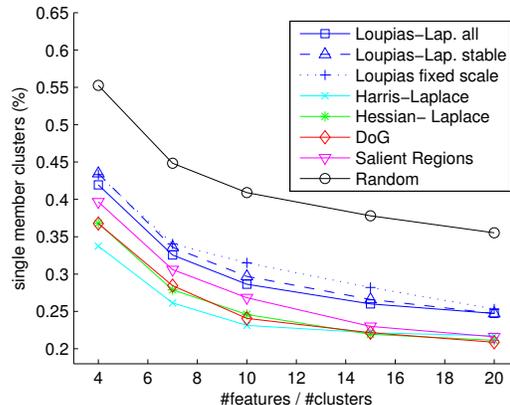


Figure 3. Percentage of single member clusters for different cut values, they show the diversity of the structures detected.

animal classes. The animals in the pictures are photographed in diverse poses, sizes and in front of different backgrounds. The database was proposed by the MUSCLE initiative [2], and is a subset of the Corel database.

For each type of interest point detector, three types of features capturing different properties of a local structure were computed. For color, we used an HSV color histogram with quantization factors $8 \times 4 \times 3$. The local shape of the patch was captured by a SIFT-like gradient location and direction histogram [6], here we used a 3×3 grid for the locations and 6 orientations. To capture texture information, we calculated the mean of the wavelet coefficients corresponding to a detected region at the three largest scales and the three orientations. A large value of such a feature means a rough texture, a small value describes a smooth region. All features were concatenated, so the altogether feature dimension for each local area was 159.

Structure Variety A first test is concerned with the variety of structures extracted from the detected regions. It follows closely one proposed in [7]. Traditional interest point detectors cover only a specific type of structure in an image, i.e., corners or blobs. In order to test the variety of structures extracted, we test how the data clusters in feature space. We use agglomerative clustering with pairwise average linking of features extracted from 30000 random regions detected by a respective detector type. In the process of clustering, a tree is build by joining the clusters with the shortest distance. The tree can then be “cut” at various levels in order to ob-

tain a different number of clusters. Since the detectors deliver a different number of regions, and in order to be comparable to the results given in [7], the trees were cut at levels resulting in specific ratios of detections vs. the number of clusters (i.e. 4, 7, 10, 15, 20). The test measures the percentage of clusters with just one member. These represent distinct structures. The single cluster ratio for the different detectors and cut values are listed in figure 3.

We can see that the features extracted from Loupias-Laplace regions are more diverse than features from the Harris-Laplace or the DoG regions, since more single member clusters exist. The Loupias-Laplace detector does not only extract a specific image structure (e.g. blobs), but a larger variety. As expected, the Loupias detector without scale selection shows a higher single cluster ratio than the detectors with scale selection. Since the scale information of the local structures is not considered, the regions obtained are more diverse. Regions extracted from random areas are very distinct. For these structures, it is difficult to find a compact representation.

Animal Image Categorization Our actual goal is to use the detector in an object classification task in a bag-of-features [1] classification approach. For histogram creation, codebooks with 4000 visual words were obtained via agglomerative clustering. The histograms are classified with a standard multi class SVM with an histogram intersection (HI) kernel in a one-vs-one mode. The results were obtained using five-fold cross validation. Other kernels were tested as well but were found less suitable compared to the HI kernel.

A nice property of the Loupias-Laplace detector is the fact that the number of points retrieved can be controlled easily by considering the M most salient points. The quality of the histograms obtained from the local features does not only depend on the type of the detected structures, but also on the number of them. Few detections result in very sparse histograms that can be unstable. Some of the standard detectors deliver very few regions, because they only focus on a certain type of structure. Changing thresholds there changes to what extent structures are considered to be of the desired type, but only within a certain range.

We evaluated the classification performance for the different detectors in relation to the average number of regions detected in the images. For the Harris-Laplace, the Hessian-Laplace and the Salient Region detector, we used the standard parameter setting as well as the lowest threshold we could tune, resulting in an increased number of detections. For the other detectors, the number of regions considered per image was

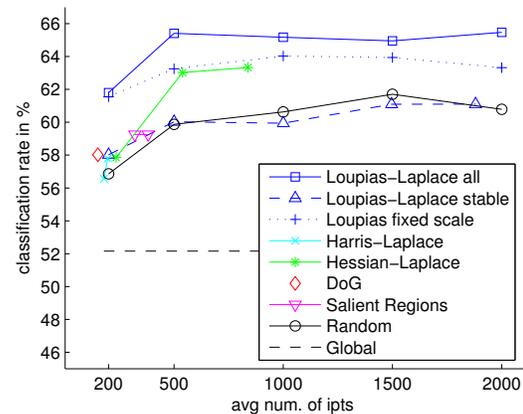


Figure 4. Classification results for the different detectors, depending on the number of interest points obtained.

adjusted to 200, 500, 1000, 1500 and 2000. As a baseline, we also computed global features as above from the entire image and used the concatenated feature vector directly for classification.

Classification results can be seen from figure 4. The best results could be obtained using the Loupias-Laplace all detector. Generally speaking, an increased number of detections results in more stable histograms and thus better classification rates. At about 1000-1500 detections, a saturation in the performance occurs. The DoG detector mainly delivered very small regions. Even after scaling them by a factor of 2, only few regions could pass the the required minimum radius of 5 pixels to be considered. The Harris-Laplace detector provides comparatively few regions for this type of images, still after adjusting the threshold. An interesting observation is that a large number of random regions could even outperform some of the sophisticated detectors. This observation is consistent with the results of [10]. There, they report superior classification results for a sufficient number of random detections compared to the LoG and the Harris-Laplace detector.

Incorporating context information in the from of background detections improves the classification rate. The reason is that the background and the animal class are of course somewhat related. This explains the rise in performance between the Loupias-Laplace stable, and the Loupias-Laplace all detector. Regions where the scale selection mechanism fails and thus are neglected by the first detector are mainly uniform regions of the background. It also explains the comparatively well performance of the original Loupias detector with a fix scale. There, no areas are discarded. The usefulness

of scale selection is demonstrated by the fact that for all numbers of detections, the results for the Loupias-Laplace all detector is better than the Loupias detector with a fixed radius.

6. Conclusions

In this work, we extended the wavelet based interest point detector by Loupias et al. by a Laplacian scale selection mechanism and compared it thoroughly to other state-of-the art detectors.

For visual object classification, the Loupias-Laplace detector is very well suited. It detects a variety of image structures beneficial for classification. The detector has the advantage that the number of retrieved points can be controlled in a straightforward way by taking the M most salient points, and these M points are usually spread out over the whole image. Especially for bag-of-feature type recognition algorithms, a sufficient number of interest points is necessary, and they should cover various “interesting” structures in an image. An alternative would only be to run multiple detector types, however, at increased cost. As demonstrated, our detector on its own is sufficient to retrieve a variety of suitable structures.

7. Acknowledgment

We like to thank Etienne Loupias for the provision of the source code of the original wavelet based salient point detector. Moreover, we would like to thank the authors of the other detectors for providing binaries of their programs for comparison. This work was partially

funded by the European Union MUSCLE NoE (FP6-507752).

References

- [1] G. Csurka, L. Dance, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Proceedings of the ECCV*, pages 59–74, 2004.
- [2] A. Hanbury and A. T. Targhi. A dataset of annotated animals. In *Proceedings of the 2nd MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation*, Alicante, Spain, 2006.
- [3] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *Proceedings of the ECCV*, pages 345–457, Prague, Czech Republic, 2004.
- [4] T. Lindeberg. *Principles for Automatic Scale Selection*, volume 2, pages 239–274. Academic Press, 1999. Handbook on Computer Vision and Applications.
- [5] E. Loupias, N. Sebe, S. Bres, and J.-M. Jolion. Wavelet-based salient points for image retrieval. In *Proceedings of the ICIP*, 2000.
- [6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.
- [7] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Proceedings of the ICCV*, volume 2, pages 1792–1799, 2005.
- [8] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proceedings of the ECCV*, volume 1, pages 128–142, 2002.
- [9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE TPAMI*, 27(10):1615–1630, 2005.
- [10] E. Nowak, F. Jurie, and W. Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings of the ECCV*, pages 490–503, Graz, Austria, 2006.