SEMI-LOCAL FEATURES FOR THE CLASSIFICATION OF SEGMENTED OBJECTS

Robert Sorschag
Institute of Software Technology and Interactive Systems, Vienna University of Technology
Favoritenstrasse 9-11, A-1040 Vienna, Austria

Keywords: Visual Features, Segmentation, Object Detection.

Abstract: Image features are usually extracted globally from whole images or locally from regions-of-interest. We propose different approaches to extract semi-local features from segmented objects in the context of object detection. The focus lies on the transformation of arbitrarily shaped object segments to image regions that are suitable for the extraction of features like SIFT, Gabor wavelets, and MPEG-7 color features. In this region transformation step, decisions arise about the used region boundary size and about modifications of the object and its background. Amongst others, we compare uniformly colored, blurred and randomly sampled backgrounds versus simple bounding boxes without object-background modifications. An extensive evaluation on the Pascal VOC 2010 segmentation dataset indicates that semi-local features are suitable for this task and that a significant difference exists between different feature extraction methods.

1 INTRODUCTION

The main research question of this work is: How to extract state-of-the-art texture and color features best from segmented objects to classify them? This question is relevant because a set of object detection approaches have been proposed where segmentation is used as a pre-processing step (Pantofaru et al., 2008), (Li et al, 2007), (Rabinovich et al., 2007), (Russel et al., 2006). They outperform sliding window approaches although almost the same features and classification techniques are used. We believe that customized features that are less distracted by the object's background can further improve these results. The features proposed in this work exploit this benefit. Furthermore, they are simple and fast to compute which makes them suitable to assist segmentation-based object detection systems.

Generally, the detection of class-level objects in real-world images is a challenging task for automated systems that is far from solved. Objects can be situated everywhere and at every size in an image. They can be occluded and shown under all kinds of perspective distortions or under different lighting conditions. Moreover, intra class differences and inter class similarities can complicate this task. Even humans sometimes fail to distinguish between closely related classes like bicycles and motorbikes when only a single image with difficult examples is shown. However, the complexity of object detection can be reduced when a set of segmented object hypotheses are given in the first place (Li et al., 2007) because it is accurately known where to search for an object.

In this work, we extract well-established image features semi-locally from segmented objects. Thereby, color and texture features are generated from image regions that contain the entire object. We use the term semi-local features because these features are locally extracted from the image but globally extracted from the object. Furthermore, we show that the use of differently prepared image regions facilitates the power of these features. For instance, the object background is excluded and

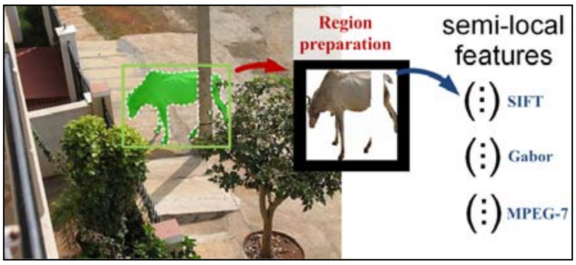

Figure 1: Semi-local features.

Sorschag R. (2012).
SEMI-LOCAL FEATURES FOR THE CLASSIFICATION OF SEGMENTED OBJECTS.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 170-175
DOI: 10.5220/0003712301700175

replaced by white pixels in Figure 1.

This work contributes to object detection research with an extensive study on the suitability of semi-local features for the classification of segmented objects and the influence of different region preparation techniques. The used set of image features and dissimilarity measures should ensure that the evaluation results are as universally valid as possible. We do not propose a complete object detection system with object segmentation and classification techniques. Instead we work on interactively generated segmentations that are provided by the Pascal VOC challenge (Everingham et al., 2010) and use a simple nearest neighbor classification. In addition to this perfect segmentation, we simulate inaccurate segmentations for comparison.

The remainder of the paper is organized as follows. Section 2 describes related work in the field of object detection and segmentation. Section 3 presents semi-local features. Section 4 explains the experiments and Section 5 draws conclusions.

2 RELATED WORK

Local features (Mikolajczyk and Schmid, 2005a) are a part of the best practice for object detection systems. First, these features are regularly sampled or extracted around interest-regions (Mikolajczyk et al., 2005b) before they are generalized to one or more bag-of-features (BoF) per image (Van de Sande et al., 2010), (Lazebnik et al., 2006). This BoF approach produces fixed-length vectors for classification. In order to locate objects within an image, many sub-regions are then investigated with a sliding window (Lampert et al., 2008). In addition to BoFs, global and semi-local features have been successfully used for related tasks, like scene classification (Oliva and Torralba, 2006), geometric context retrieval (Hoiem et al., 2005), and human body detection (Dalal and Triggs, 2005).

2.1 Segmentation-based Detection

Object detection approaches that operate on segmented objects (Pantofaru et al., 2008), (Li et al., 2007), (Rabinovich et al., 2007), (Russel et al., 2006) work similar to sliding window approaches but with a heavily reduced search-space. Thus, more powerful (and computationally more expensive) recognition approaches can be applied. However, this benefit is not extensively exploited so far: In Pantofaru et al. (2008) color histograms and RCF (regionSIFT) descriptors are extracted from the segmented objects. Li et al. (2007), Rabinovich et al. (2007), and Leibe et al. (2008) generate BoFs from SIFT (Lowe, 2004), colorSIFT (Van de Sande et al., 2010), local shape context (Leibe et al., 2008), and gray-value patches. In Li et al. (2007) independent BoFs are extracted from the segmented object and its background within a bounding box as well as semi-local HoG features (Dalal and Triggs, 2005). Rabinovich et al. (2007) sets all background pixels to black and extracts local features from interest-regions that overlap with the segmented object. We use a similar zero-masking step to generate features with a higher weighting of the object shape.

Only Toshev et al. (2010) propose segmentation specific features, called boundary object shape, where the geometric relations of object boundary edges are measured. We further explore this idea and propose customized features for the classification of segmented objects. To the knowledge of the authors, no work has been proposed so far that investigates such semi-local features for object detection.

2.2 Segmentation Approaches

Different object segmentation approaches including Normalized Cuts (Shi and Malik, 1997), MinCuts (Carreira and Sminchisescu, 2010), and Mean-Shift (Comaniciu and Meer, 2002) have been used for the object detection systems described above. A good overview of segmentation approaches can be found in Hoeim et al. (2011). In contrast to semantic segmentation (Csurka and Perronnin, 2010), these approaches work without knowledge about the segmented objects and they are used to generate a ‘soup’ of many overlapping segmentations. Such multi-segmentation approaches can achieve higher object detection rates when overlapping segments are individually classified and combined afterwards (Li et al., 2007). All of the described object detection systems work with unsupervised segmentation. However, it can be useful to test single stages of such detection systems on interactively generated object segments that are almost perfect (Pantofaru et al., 2008). We use this strategy to compare different semi-local features that are extracted from perfectly and inaccurately segmented objects.

3 SEMI-LOCAL FEATURES

We extract and classify semi-local features from segmented objects in following steps. First, a set of transformed image regions are prepared from every segmented object. Next, different color and texture features are extracted from these regions and stored in a database. The features of each object are then matched against the features of all other objects using a nearest neighbor strategy with several dissimilarity measures. At last, we evaluate the percentage of correctly matched features for each object class.

3.1 Region Preparation

In this work, semi-local features are extracted from regions around segmented objects using different object-background modifications, segment-ation accuracies and bounding boxes. In the following, these region preparation methods are explained and their effects on the resulting feature properties are discussed.

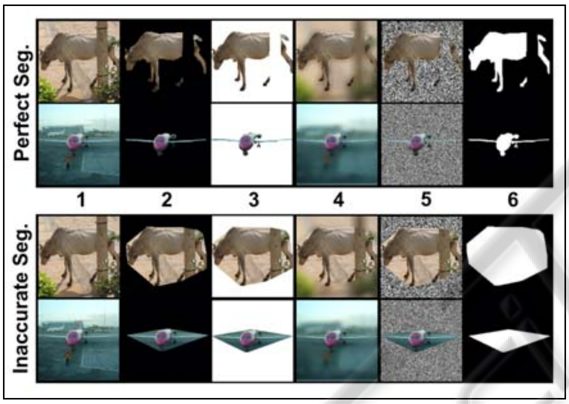

Figure 2: Region preparation techniques.

Object-background modifications: We use six different modification techniques, shown in the columns of Figure 2. Region 1 (leftmost column) is equivalent to bounding boxes without segmentation. No focus is set to specific properties of the object in these regions. In the opposite, shape is the only attribute left to describe in Region 6 (rightmost column). In Region 2 and Region 3 black and white backgrounds are used. These regions set the focus to the object shape and its content. Region 4 keeps the characteristics of the original background although the object is focused and the object boundaries are sharpened. We use Gaussian smoothing to blur the background of these regions heavily. The Gaussian noise of Region 5 also sets focus to the object but with fewer weighting of the object shape. In preliminary experiments, we have tested further object-background modifications (e.g. object bound expansion) but the six selected ones performed best.

Segmentation accuracy: As shown in Figure 2, we use two different segmentation accuracies. On the one hand, perfect segmentations are given from the Pascal VOC dataset (Everingham et al., 2010). The object pixels are thereby used as foreground and all others are used as background. On the other hand, we simulate an inaccurate segmentation using the convex hull of all pixels that belong to a perfectly segmented object. No holes are given in this approach but the actual object shape is heavily changed.

Bounding boxes: Most image features are extracted from square image regions. However, segmented objects are given as arbitrarily shaped polygons or image masks, and thus we operate on bounding boxes around such object segments. As shown in Figure 3, we select two different bounding boxes for each object. First, we use tight, rectangular bounding boxes that touch the segment bounds on all four sides. These regions are resized to squares in a pre-processing step. Secondly, we use squared bounding boxes that touch the object bounds only in the larger dimension. These regions contain larger parts of the object's background but no additional resize step changes the aspect ratio of these regions.

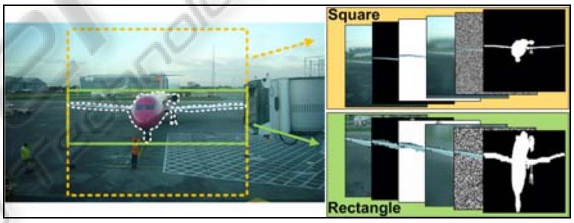

Figure 3: Bounding boxes.

3.2 Image Features

In this work, four state-of-the-art texture and color features are used: SIFT, Gabor wavelets, MPEG-7 ColorLayout and ScalableColor. We omit to add specific shape features because the used texture features extracted from Region 6 (white object on black background) already present effective shape features. All features are computed on 64 x 64 pixel regions.

SIFT features (Lowe, 2004) consist of 8-dimensional orientation histograms that are computed from the image gradients in 16 slightly overlapping sub-regions on a 4x4 grid. The feature is normalized to increase the robustness against color and illumination changes. In the proposed semi-local feature approach, we extract only one SIFT feature from the entire object region. No interest point detection is used in this process.

Gabor wavelets (Frigo and Johnson, 2005) are computed with a bank of orientation and scale sensitive Gabor filters. The mean and standard deviation of each filter output is thereby used as features values.

MPEG-7 ColorLayout features (Manjunath et al., 2001) present the spatial distribution of colors in a very compact form. They cluster an image or image region into sub-regions of 8x8 pixels and compute the average pixel value for each of them. Finally, the first low frequency coefficients of a discrete cosine transform are selected.

MPEG-7 ScalableColor features (Manjunath et al., 2001) are computed from a quantized HSV color histogram. A scalable binary representation is then generated by indexing the probability values of each histogram bin before a discrete Haar transformation is applied. The resulting feature is scale invariant.

3.3 Classification

We compute the nearest neighbor for segmented objects using all described region preparation techniques and feature types independently. Thereby, each segmented query object is matched against all segmented objects in the dataset that do not stem from the same image as the query object. The object class of the nearest neighbor is then used to classify a query object. We perform this nearest neighbor classification with following dissimilarity measures to get as general findings as possible.

- Minkowski family distances: L1, Euclidian, and Fractional distance
- Cosine function based dissimilarity
- Canberra metric
- Jeffrey divergence
- Chi-Square statistics

These measures have been chosen according to their high performance for image retrieval with global features in Liu et al. (2008) where further information about these measures can be found. We believe that more sophisticated classification approaches can be used to achieve better results, but it is out of the scope of this work to identify the best classification strategies. Instead, we try to perform a fair comparison between the proposed feature extraction techniques.

4 EVALUATION

In the experiments of this work, we used two different evaluation strategies. On the one hand, the recall of correctly classified objects is computed for each object class and for all classes combined. On the other hand, we perform an additional k-nearest neighbor classification to evaluate the percentage of query objects with at least one correct match in the top k entries (k = 1-10).

4.1 Dataset

We used the open Pascal VOC 2010 segmentation dataset (Everingham et al., 2010) for experiments. In this dataset, 20 different object classes (see x-axis of Figure 4) are perfectly segmented in 1928 Flickr images. The ground-truth contains a total number of 4203 objects whereby several object classes occur more often than other ones. For instance, 928 persons and 108 dinning tables are given. All images are provided with jpg encoding and a longer dimension side of 500 pixels.

4.2 Results

The results are organized according to following aspects: the suitability of semi-local features to classify segmented objects and the role of region preparation, segmentation accuracy, used image feature types, and dissimilarity measures. Figure 4 and Table 1 are used to illuminate these points. Both show the achieved recall of nearest neighbor classification for Jeffrey divergence on squared bounding boxes.

Semi-local features: Figure 4 shows that the classification rates of the best matching object classes are significantly above 50% for texture features. Furthermore, the results of all objects are clearly above random classification (5%) independent of the used feature type. The fact that all 4-legged animals (sheep, horse, cow, cat, dog) are below the average, indicates that inter class similarities decrease their classification. As shown in Table 1, the highest overall classification rate of 46,5% was achieved with SIFT features from perfectly segmented Region 6. Moreover, 80% of all objects have at least one correct match within the first 10 retrieved objects for the same configuration. These results clearly indicate that semi-local features are able to facilitate the classification of accurately segmented objects.

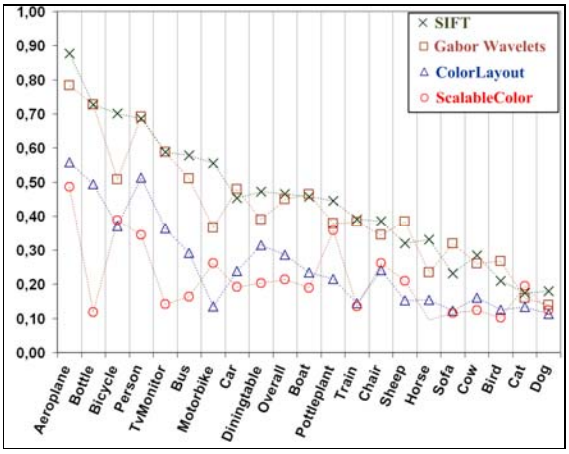

Figure 4: Classification results per object class for perfect segmentations and square bounding boxes. For each feature type we give the recall of the best performing region.

Region preparation: Table 1 shows that texture features achieved the best results on Region 6 (white foreground on black background) where only shape information is given. This is also true for most object classes. MPEG-7 color descriptors generally perform best with original objects on uniformly colored background (Regions 2 and 3). These regions are also the best choice for texture features when no accurate segmentation is given. At the first glance, white background outperforms black background on the given dataset but the results of k-nearest neighbour matching did not verify this assumption. Moreover, square bounding boxes always achieved better results than rectangle bounding boxes for SIFT and MPEG-7 features by an average increase of 2%. This indicates that the effect of changing the object's aspect ratio is worse than using a larger amount of background. However, for Gabor wavelets no significant changes have been measured between square bounding boxes and rectangle ones.

Segmentation accuracy: In order to simulate inaccurate segmentations from the given test set, we used the convex hull around perfectly segmented objects. Table 1 shows the classification results of perfectly and inaccurately segmented objects. These results indicate that accurate segmentation can improve the classification significantly (up to +24,5%) when the region is prepared appropriately. In contrast, only smaller improvements of about 2% are achieved between unmodified regions (Region 1) and modified ones for inaccurate segmentation. Only the results of Gabor wavelets improved from 20,5% to 25.3% and 24.8% for uniformly colored backgrounds. Region 6 performs worse than all other regions for inaccurate segmentation because these regions only contain very rough object contours, as shown in Figure 2.

Table 1: Overall results (recall) for square bounding boxes.

		R.1	*R.2*	*R.3*	*R.4*	*R.5*	*R.6*
Perfect Seg	**SIFT**	25,0	38,3	40,4	32,0	29,8	**46,5**
	GW	20,5	37,2	39,9	21,0	31,5	45,0
	CL	15,4	22,4	23,6	19,6	15,0	28,7
	SC	16,4	21,8	21,4	21,6	16,5	-
Inacc. Seg.	**SIFT**	25,0	27,2	27,5	22,5	27,2	12,1
	GW	20,5	25,3	24,8	19,1	25,1	10,8
	CL	15,4	16,8	18,6	17,9	15,2	15,1
	SC	16,4	16,5	16,8	16,5	15,8	-

Feature types: The performance of SIFT and Gabor wavelets is similar for both segmentation accuracies and all regions except Region 1 and Region 4 where the background is left unmodified and blurred, respectively. Gabor wavelets perform slightly better on rectangular bounding boxes while SIFT achieves better results on square regions. MPEG-7 ColorLayout and ScalableColor features perform worse than texture features for the given task. Although Figure 4 indicates that ColorLayout outperforms ScalableColor this is only true because the best performing region preparation approach (Region 6) is not applicable for pure color features, like ScalableColor, where no spatial information is used.

Dissimilarity measures: The difference between the best and the worst dissimilarity measure for all features is about 3-5%. For instance, the results of SIFT features for Region 6 on perfect segmentations lie between 46,5% for the best (Jeffrey divergence) and 42,4% for the worst measure (Canberra metric). The highest variations are caused by MPEG-7 ScalableColor features. It seems that the ranking of dissimilarity measures does not depend on the used region preparation technique because the results of all measures are similarly ordered for all techniques. The best dissimilarity measure for all features was Jeffrey divergence followed by Chi-Squared statistics. The worst measure was Fractional distance for all features followed by Canberra metric for texture features. L1 metric performed best of the Minkowski family measures, especially for texture features where the difference to Euclidian distance was above 2,5%.

5 CONCLUSIONS

We have proposed semi-local features for the classification of segmented but unknown objects. In this approach, state-of-the-art texture and color features are extracted from regions that cover the entire object with and without background-modifications. Results of an extensive evaluation indicate that the proposed approach offers the opportunity to improve the task of object class detection in combination with efficient segmentation approaches. The experiments of this work investigated perfect segmentations as well as inaccurate ones. The classification was done with a nearest neighbor matching strategy and different dissimilarity measures to keep the evaluation as simple and universally valid as possible.

In the evaluation, we have first shown that it does matter how the regions of segmented objects are prepared for semi-local feature extraction. Regions where the object and its background are modified can improve the overall classification rate significantly compared to unmodified regions, especially for accurate segmentations. Secondly, square bounding boxes achieves better results than tight, rectangular bounding boxes. Thirdly, texture features perform better than color features and improvements of a few percent can be achieved when the right dissimilarity measures are chosen. The Jeffrey divergence and Chi-Square correlation performed best for all feature types and region preparation techniques. We conclude that semi-local features are good candidates to improve object detection systems due to their simplicity and the promising results in this work. Furthermore, we plan to investigate semi-local features in an integrated object detection system to verify this assumption.

ACKNOWLEDGEMENTS

The author would like to thank Horst Eidenberger for his feedback and support. The research leading to this publication has received funding from the Austrian FIT-IT project ‘IV-ART – Intelligent Video Annotation and Retrieval Techniques’.

REFERENCES

Carreira, J., Sminchisescu, C., 2010. Constrained parametric min cuts for automatic object segmentation. *CVPR*.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *PAMI*.

Csurka, G., Perronnin, F., 2010. An efficient approach to semantic segmentation. *IJCV*.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. *CVPR*.

Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A., 2010. The PASCAL Visual Object Classes (VOC) challenge. *IJCV*.

Frigo, M., Johnson, S., 2005. The design and implementtation of FFTW3. *Proc. Program Generation, Optimization, and Platform Adaptation*

Hoiem, D., Efros, A., Hebert, M., 2005. Geometric context from a single image. *ICCV*.

Hoiem, D., Stein, A., Efros, A., Hebert, M., 2011. Recovering occlusion boundaries. *IJCV*.

Lampert, C., Blaschko, M., Hofmann, T., 2008. Beyond sliding windows: Object localization by efficient subwindow search. *CVPR*.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *CVPR*.

Leibe, B., Leonardis, A., Schiele, B., 2008. Robust object detection with interleaved categorization and segmentation. *IJCV*.

Li, F., Carreira, J., Sminchisescu, C., 2007. Object recognition as ranking holistic figure-ground hypotheses. *CVPR*.

Liu, H., Song, D., Rüger, S., Hu, S., Uren, V., 2008. Comparing dissimilarity measures for content-based image retrieval. *AIRS*.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *IJCV*.

Manjunath, B., Ohm, J.-R., Vasudevan, V., Yamada, A., 2001. Color and texture descriptors. *Trans. on Circuits and Systems for Video Technology*.

Mikolajczyk K., Schmid, C., 2005a. A performance evaluation of local descriptors. *Trans. PAMI*.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005b. A comparison of affine region detectors. *IJCV*.

Oliva, A., Torralba, A., 2006. Building the GIST of a Scene: The Role of Global Image Features in Recognition, *Visual Perception, Progress in Brain Research*.

Pantofaru, C., Schmid C., Hebert, M., 2008. Object recognition by integrating multiple image segmentations. *ECCV*.

Rabinovich A., Vedaldi, A., Belongie, S., 2007. Does image segmentation improve object categorization? *Tech. Rep. CS2007-090*.

Russell, B., Freeman, W., Efros, A., Sivic, J., Zisserman, A., 2006. Using multiple segmentations to discover objects and their extent in image collections. *CVPR*.

Shi, J., Malik, J., 1997. Normalized cuts and image segmentation. *CVPR*.

Toshev, A., Taskar, B., Daniilidis, K, 2010. Object detection via boundary structure segmentation. *CVPR*.

Van de Sande, K., Gevers, T., Snoek, C., 2010. Evaluating color descriptors for object and scene recognition. *PAMI*.