# Segmentation Based Interest Points and Evaluation of Unsupervised Image Segmentation Methods

Piotr Koniusz
P.Koniusz@surrey.ac.uk

Krystian Mikolajczyk
K.Mikolajczyk@surrey.ac.uk

CVSSP, University of Surrey, UK

## Abstract

This paper investigates segmentation based interest points for matching and recognition. We propose two simple methods for extracting features from the segmentation maps, which focus on the boundaries and centres of the gravity of the segments. In addition, this can be considered a novel approach for evaluating unsupervised image segmentation algorithms. Former evaluations aim at estimating segmentation quality by how well resulting segments adhere to the contours separating ground-truth foregrounds from backgrounds and therefore explicitly focus on particular objects of interest. In contrast, we propose to measure the robustness of segmentations by the repeatability of features extracted from segments on images related by various geometric and photometric transformations. Further, our evaluation provides a new insight into suitability of the segmentation methods for generating local features for image retrieval or recognition. Several segmentation methods are evaluated and compared to state-of-the art interest point detectors using the repeatability criteria as well as standard matching and recognition benchmarks.

## 1 Introduction

One of the crucial issues in image retrieval or recognition is the extraction of salient features. Segmentation methods seem to have great potential of delivering good features as their main goal is to separate foreground objects from backgrounds. For instance, in [18] multiple segmentations were used to find objects and their extent in collections of images. The assumption was that all similar objects across images give rise to segments alike, and those irrelevant appear dissimilar. Reminiscent approaches were taken in [11]. They concluded that even over-complete representations may be insufficient to achieve satisfactory repeatability of segmentation maps. Similar scenes affected by natural lighting conditions, angle of view and scale result in a range of different collections of segments. Thus, partial matching was taken into further investigation in [8]. We argue that stability of produced partitions is more important than unambiguous foreground/background separation for such applications. Techniques undertaken in [8, 11, 18] show that segmentation methods can be used in the recognition field. Since the state-of-the art recognition results [4] are still produced with interest point based methods, the advantages of segmentations for such purposes are yet to be proved with the appropriate approach.

This paper reports on a set of tests which aimed at identifying what kinds of features from general-purpose segmentation algorithms come as stable. This enables further exploitation of these stable parts to build reliable representations for image content retrieval or classification systems. We are unaware of any previous evaluation that targets stability of segmentations with the use of interest points and recognition, which makes this work novel in these areas and contributes to the segmentation evaluation problem. Furthermore, it also contributes towards bridging the gap between local features and unsupervised segmentations. Along with a simple testing protocol we deliver local feature detectors based on the segmentation maps that may be directly used in many applications utilising interest points. Extensive evaluation demonstrates the performance of local features that indirectly characterises the quality of different segmentations. In contrast to the existing evaluations like [1, 6, 12], we quantify the segmentation performance in terms of suitability for recognition with means of local descriptors.

**Related work.** Exhaustive surveys on state-of-the-art local invariant features and their evaluations can be found in [14, 15, 22]. The interest points are usually characterised in terms of repeatability and invariance to different geometric and photometric changes. The most known testing approach is based on repeatability of features for a number of undergoing changes between two images related by a homography matrix. Interest point detectors based on first or second order gradient [22] achieved the highest scores in such evaluations. Also, MSER detector [13] based on Watershed segmentation performed extremely well there.

According to a recent survey on quality of segmentation [6], three most robust methods are Mean Shift [2], Efficient Graph-Based Image Segmentation [5], and Normalised Cuts [20]. The results were obtained on a database of 1023 images by gauging how well resulting segments adhered to the contours separating ground-truth foregrounds from backgrounds. Combined performance of segmentations was also estimated and their complementary nature emphasised. Another important benchmark proposed in [12] was based on evaluation of segmentation quality using human-labelled multi segment ground-truth as opposed to single-mask foregrounds in [6]. Their database consists of 12000 hand-labelled segmentations of 1000 images taken from Corel dataset. The test measures the difference between union and intersection of two corresponding regions segmented out of similar images. If two compared segmentations are just refined/generalised versions of each other, this approach yields small errors. Precision-recall curves relying on a measure of matched pixels of boundaries between two segmentations were applied in [3]. Thy ranked segmentations as follows: SE Min-Cut, Canny Edge Detector, Mean Shift, Local Variation, and Normalised Cuts. Lastly, recent survey in [23] categorised different evaluation methods for segmentations and compared nine of them. It was pointed out that the evaluation methods are reliable for a given segmentation, though they struggle to remain objective when comparing a few different segmentations at a time. Our approach addresses some of these issues.

# 2  Segmentation Based Interest Points

This section briefly discusses the investigated segmentation approaches and then presents the methods for extracting local features from their segmentation maps.

## 2.1  Unsupervised Segmentation Methods

This study follows the findings of [6] and focuses on measuring performance of Efficient Graph-Based Image Segmentation (EGO), Mean Shift (MS), Watershed (WA) and Normalised Cuts (NC) in terms of their stability.
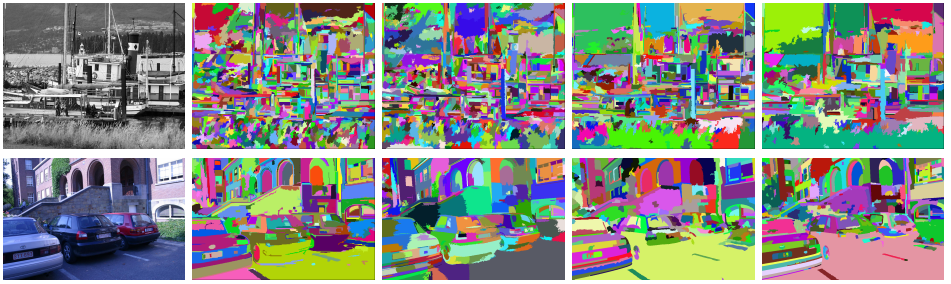
Figure 1: Examples of images along with corresponding segmentations. (Top) Results of EGO for over-, well-, and under-segmented sets (from left to right). (Bottom) Segmentation maps of EGO, MS, NC, and WA methods on well-segmented set (from left to right).

*EGO* [5] is a graph-based technique where all vertices represent pixel coordinates of an image and edges represent a similarity measure between neighbouring vertices by the difference of their RGB colours. This method relies on the bottom-up scheme and the merging strategy is based on a pair-wise region comparison.

*MS* [2] is a simple and robust technique based on a connectedness criterion. All pixels of an image are considered as vectors in 5D consisting of spatial and colour coordinates. Centroid based mode detection is employed and coordinates are ascribed modes. Recursive fusion of basins of attraction merges the modes located within a certain radius.

*Normalised Cuts* [20] is also based on vertices and edges representing pixel coordinates and pixel similarities in a graph storing $N \times (width \times height)^2$ bytes of data. N represents a single weight in bytes. Image partitioning is performed by a cut between two disjoint sets of vertices which optimise normalised cost function. We modified the segmentation process (further referred as NC) to overcome the complexity issues due to which images larger than 200x200 pixels cannot be easily handled. Larger images were split into a set of half-overlapping sub-windows. The resulting segments were merged by using only those non adjacent to the boundaries of sub-windows. Since performance of the original implementation seemed plunging immediately in presence of scale/affine changes, a merging method similar to the Watershed post-processing described further in the text was applied. This gave satisfactory segmentation results and significantly reduced the processing time.

*Watershed* [9] segmentation acts on an image luminance/RGB map and uses the gradient descent method to seek for local minima. Thus, the pixels are attracted to the minima within a given basin of attraction. This method benefits when combined with an anisotropic filtering (further referred as WA) introduced in [17]. We introduced additional post-processing step sorting all segments in ascending order (by size) and merging first N percent of adjacent small segments based on their similarity of average colour.

## 2.2 Detection of Interest Points from Segmentation Maps

Inspired by evaluation of affine region detectors [14, 15], we focused on two kinds of keypoints locating potentially salient parts of segments.

*Ellipses* inscribed in the segments are potentially repeatable features. Centre estimation and ellipse fitting can be performed on either contour coordinates or over the whole area. We found that area fitted ellipses are more repeatable as associated segments often suffer from partial spilling into noisy structures under both geometric or photometric changes.

*Corners* located on region boundaries are salient features which may overcome the spilling problem. SUSAN detector [21] is very well tailored to detect corners and junctions on seg-
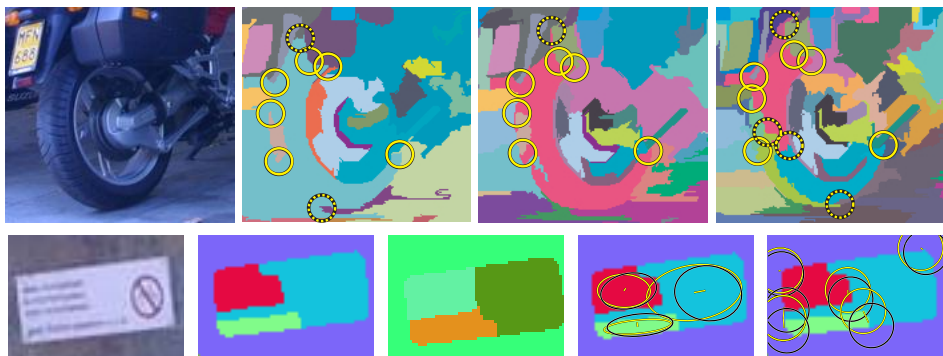
Figure 2: (Top) An example image with under-, well-, and over-segmented tire (from left to right) and detected corners. (Bottom) Segmentation results before and after applying a blur as well as matched ellipses and corners (explained in the text), from left to right respectively.

ment boundaries. It is also efficient and highly repeatable. Our implementation of SUSAN detector simply scans the boundaries of the segments with a 19x19 window and counts the number of pixels with the same label as the central point. These numbers are recorded in another image map and smoothed with a Gaussian filter. Finally, non-minima are suppressed, and corners and junctions detected. Figure 3 illustrates both types of features.

**Discussion on Evaluation of Boundary and Centre Based Features.** The investigated segmentations yield disjoint regions covering the whole area of images. Thus, the affine regions retrieved by fitting ellipses into segments form a complete representation of the image in contrast to other interest point detectors which often provide too many features in some areas and none in others. Moreover, the segment based features often capture contextually meaningful parts of the objects, *e.g.* cars, wheels, windows, limbs, etc. If the segmentation approaches produced repetitive results, the features based on these results would also be very repetitive. Figure 2(top) illustrates a tire of a bike along with the detected corners. Although the tire appears to be segmented out correctly only in the well-segmented result, there are correctly matched corners (yellow circles) amongst all three segmentations and only few corners remain unmatched (dotted circles). This highlights the repetitive capabilities of boundary based feature points as opposed to the centre based regions which suffer more from over- or under-segmentation. On the other hand, a blur applied to the image in figure 2(bottom) gave rise to two slightly different segmentations. This time all three segments between two segmentations were well matched (yellow and black ellipses connected by a line denote match between two regions). Although, there remained a few corners unmatched.

To quantify both effects, we employed two complementary measures based on the homography ground-truth. The region overlap from [15] was defined by a ratio of intersection to union of reference region $R_r$ and projected region $R_p$: $\varepsilon_o = 1 - \frac{R_r \cap R_p}{R_r \cup R_p}$. This measure was used to evaluate centre based regions by the percentage of correspondences for which $\varepsilon_o \leq 0.3$. In addition, for the boundary based points we used the distance in pixels between the interest point and its nearest projected correspondence. The correspondences were considered correct if $\varepsilon_n \leq 4$ pixels. We refer to this measure as nearest neighbour (NN). The goal of adopting the overlap based repeatability [15] was to examine to what extent segments from a given segmentation are roughly preserved over a wide range of transformations. The NN repeatability measure [19] was applied to quantify the accuracy of segment boundaries. Figure 3 visualises the overlap (left) and distance (right) based correspondences.
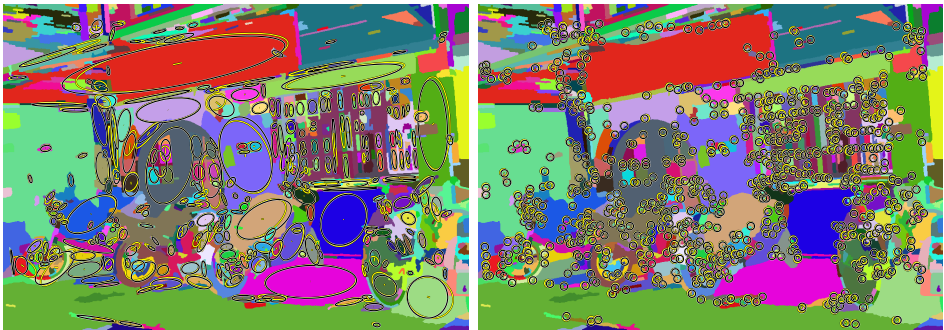
Figure 3: Illustration of segment (left) and boundary (right) based features (yellow) in the reference image together with their correspondences (black) projected from another image.

# 3  Evaluation Results

This section outlines the experimental results of segmentation based feature evaluations. We first discuss our experimental setup and then present the results for repeatability test, matching of descriptors, complementarity of feature points and recognition performance.

## 3.1  Experimental Setup

We exploited a set of well-known test images from [15]. Each image sequence consisted of 6 images with gradual distortions: bike/blur, boat/scale-rotation, car/illumination, graffiti/affine, house/JPEG compression, bark/zoom-rotation, tree/blur, wall/affine. These changes seem the most ubiquitous and reflect phenomena taking place during image acquisition. The homography based ground-truth made this data very suitable for a quantitative evaluation. The resolution of images varied from 800x640 to 1000x700 pixels.

According to [6], the optimal performance of general-purpose segmentations was achieved for range of 10 up to 80 segments per a 200x200 pixels image. Though, it is unclear how segmentations can be compared provided wide range of their tweaking parameters. Enforcing arbitrary number of segments does not guarantee appropriate scale of observation. To address this issue we adopted a simple ad hoc solution which uses EGO to generate three different control sets of segment maps at different scales of observation, namely: over-, well-, and under-segmented. The remaining segmentations were tweaked to fit to the control sets to their best abilities. In order to avoid damaging effect of exact fitting, we built histograms of sizes of segments for all tested methods and all images from the control sets. The segmentation parameters which produced the most similar histograms to the control set according to $Chi^2$ distance were selected. Finally, we used three sets of parameters for each method. Figure 1(top) illustrates the results of EGO with the under-, well-, and over-segmented images whilst (bottom) shows all four methods on the well-segmented set. We report only a subset of the results but the observations are valid for most of them unless stated otherwise.

We followed the protocol from [15] to evaluate the segment features using the repeatability measures. Results of state-of-the art MSER and Hessian detectors [14] operating at fine scale were added as a reference. We also report the percentage of correct matches obtained with SIFT [10] to evaluate the proposed features for matching applications. We additionally investigated the intra detector complementarity. Correspondence sets (repeatable points) of methods under scrutiny were computed between testing images 1-2, 1-3, ..., 1-6. Further, correspondence sets of the reference MSER/Hessian detectors were extracted in the same manner. Subsequently, all correspondences form testing sets having significant overlap/NN
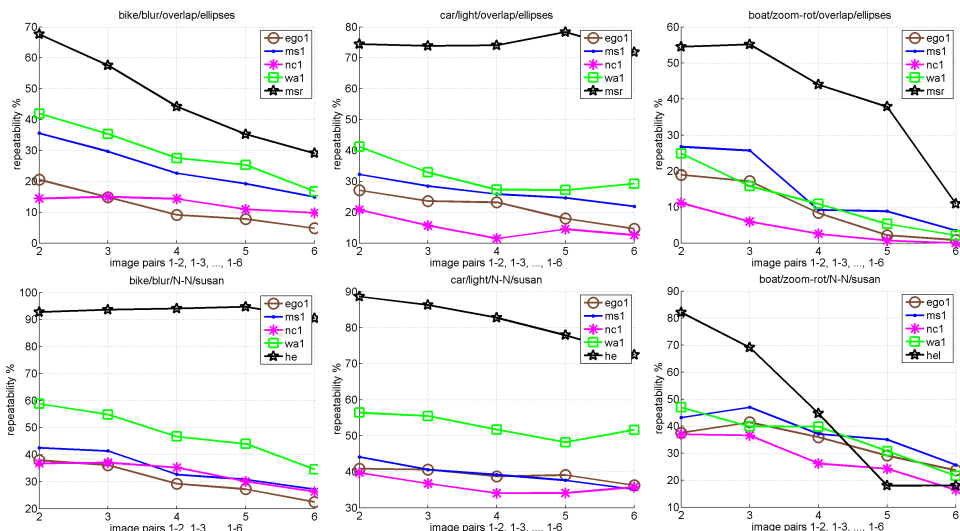
Figure 4: Repeatability results for bike (left), car (middle), and boat (right) with the ellipse based regions (top) and SUSAN corners (bottom) on the over-segmented set.

proximity with correspondences in referencing sets were removed. The ratio of remaining correspondences to their original amount is called the exact complementarity. If any tested detector yields *e.g.* 90% in such test, this indicates that 90% of all repeatable points are novel while remaining 10% are also repeatable and present at the same time in a reference method. Another measure called the relaxed complementarity differs in a way that points directly detected by a reference detector on images 2, 3, ..., 6 are used instead of correspondences from reference sets 1-2, 1-3, ..., 1-6 while subtracting them from correspondences of testing sets.

Finally, we performed a recognition experiment on standard data from PASCAL VOC Challenge [4] to show the features performance in a different application scenario.

## 3.2 Repeatability of Segmentation Methods

The repeatability of segment based features between the original and subsequently distorted images is presented in figure 4(top) for affine regions and in figure 4(bottom) for corners. The repeatability of MSER detector was greater than any of the other segmentation methods. However, unlike MSER, the other approaches did not apply any selection of the most stable regions. Similar observation is valid for Hessian points compared to segment boundary features. Otherwise, WA segmentation performed consistently better than other methods.

For the over-segmented set, WA was the winner with the repeatability of 42% for graffiti, bike, car, and house. Second best was MS with 30% for bark, boat, tree, and wall. Clearly, WA behaved bettter on structured scenes whilst MS was second best scoring on average 33% repeatability. Further, MS was the clear winner for the natural scenes where WA scored rather low. EGO was the third best reachig rougly 23% for structured and 20% for the natural views. NC scored 16% on average across all sequences.

For the well-segmented set, WA was comparable to MS on the structured images with about 40% repeatability. MS again outperformed other methods for natural scenes with average of 32%. EGO yielded roughly 20% and NC 16% across all image categories. In terms of the number of correspondences on structured scenes, WA produced approximately 150 correspondences between the original and first distorted images, and MS gave 190. These num-
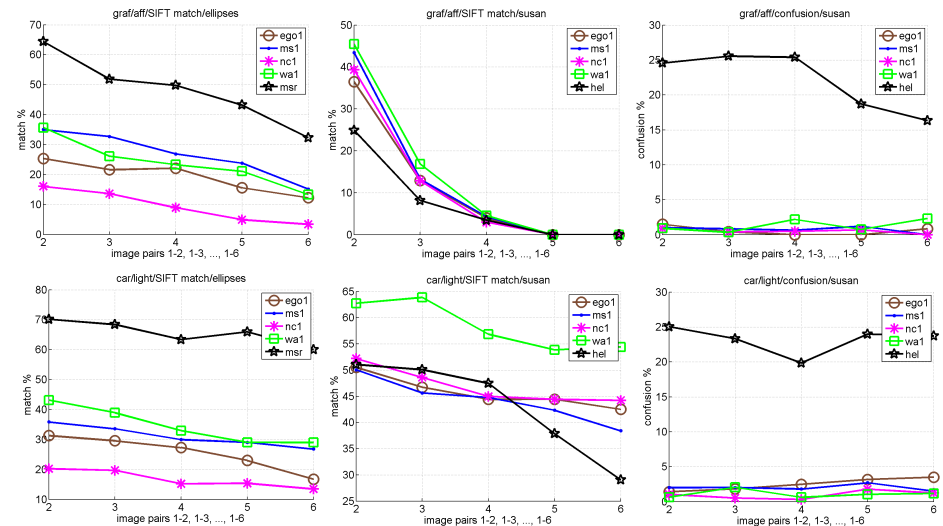
Figure 5: Matching results from the over-segmented set for ellipses (left), SUSAN corners (middle), confusion (right) for graffiti images (top) and car (bottom).

bers reached 200 for MS and 50 for WA on natural scenes. Regarding the under-segmented set, MS outperformed the other segmentations. In case of natural scenes, all segmentations but EGO produced very few correspondences ($\ll 50$) and only small persistent structures remained matched. Satisfactory amount of correspondences ($\geq 50$) was produced for the structured scenes. WA gave best results in most of the sequences followed by MS and EGO.

SUSAN corners proved repetitive in figure 4(bottom), although they performed significantly lower than Hessian (HE). WA won again on the over-segmented set (structured scenes) with maximum 58% repeatability. MS led in natural scenes with the average repeatability of 41% where NC performed second best. WA kept up the same trend for well-segmented structured scenes with average repeatability of 54% and MS consistently won throughout natural scenes reaching 41%. In case of the over-segmented image set, roughly the same results were obtained for NC and EGO. In under-segmented set, structured scenes processed by WA gave again best average repeatability of 52%. The biggest shift took place on natural under-segmented images where both EGO and NC were winners with similar performance of 35% repeatability. They delivered around 500 and 100 correspondences respectively.

Concluding, the consistently best performer for structured categories was WA followed by MS. However, MS gave more stable results in the natural scenes. EGO performed on average as third best for either scene type. The whole segments were less repetitive than the boundary points. This is due to frequent spills of regions at low gradient boundary deeming these regions not repeatable. WA and MS upheld their stability both for the area based and boundary interest points. Lastly, NC segments seemed to produce less stable features.

## 3.3 Matching with SIFT Descriptor

This section provides details on matching capabilities of the segmentation based features. Matching with SIFT descriptor was carried out for both region and corner based features. The results are displayed in figure 5. In general, the region based interest points gave results consistent with the repeatability test in section 3.2. Despite large performance gap (compared to MSER), these regions provide useful features which are unique (WA for structured and MS
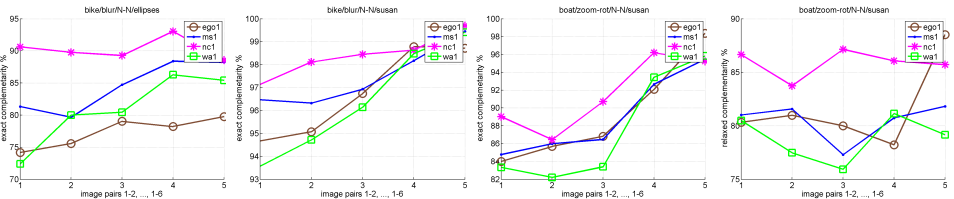
Figure 6: Inter detector complementarity as ratio of novel repeatable to all repeatable points delivered by region based and SUSAN detectors with respect to MSER/Hessian.

for natural scenes). The radii of the fitted ellipses were increased by a factor of 3 to include region boundaries into descriptors and made their sizes comparable to MSER features.

Matching of SUSAN corners brought prime results presented in figure 5(middle column). WA outperformed Hessian (HE) by 15%, 20%, 22%, and 7%, for the car, graffiti, boat, and bark sequences respectively. This is in contrast to the repeatability results in section 3.2 which showed HE as more repeatable than any joint venture of SUSAN and the tested segmentations. Further, for descriptor based matching, SUSAN combined with either of the segmentations outperformed HE for the graffiti, bark, tree, and the wall. Similar trends retained through other scales of observation. Though, the advantage of MS over WA became clear on natural scenes. We performed additional experiments to clarify the inconsistency between HE repeatability results and the matching scores. Figure 5(right) gives us insight into how many points from a given image were matched with more than one point in the corresponding transformed image. HE produced many multi-matches for the same local structures in contrast to the segmentation based points. This indicates much higher redundancy of HE. On the other hand, we attribute good performance of SUSAN to the segmentations which always strive to capture whole distinct regions. We argue these results are also due to the fact that the segmentation based corners are very salient feature points as they occur on the perimeter of two or more areas considered dissimilar by a given segmentation.

## 3.4   Inter Detector Repeatability

The following experiment aimed at capturing the levels of complementarity amongst the examined region/corner based features from segmentation maps and the MSER/HE reference detectors respectively. Highly complementary detectors can be used together to improve performance of matching or recognition. Figure 6 presents the complementarity test. The higher the exact complementarity (EC) the more novel repeatable interest points are detected with respect to the reference methods. For the region based points extracted from well-segmented images EC amounted to 79%, 78%, 93%, and 83% for EGO, MS, NC, and WA respectively. These are the average values concerning testing image pairs 1-2. Consecutive testing pairs 1-3, ..., 1-6 yielded rather monotonically increasing scores. The stronger image distortions the more novelty, although at a cost of fewer correspondences. Corner based feature points yielded the following scores: 91%, 92%, 93%, and 90% respectively. The relaxed complementarity (RC) resulted in similar trends, though lower by 2-4% on average. RC (also described in section 3.1) suffers from a bias towards non-repeatable noise detected by a reference detector. Figure 6(left) shows EC for bike (segment based regions and SUSAN corners). Two rightmost plots show the exact and relaxed repeatability scores for boat. RC usually increases with EC, though noise can affect it as in figure 6(right).

The complementarity score is expected to remain below 100% as the most distinctive features within an image should be extracted by any kind of a good detector. In general, it

seems that the segmentation based methods do introduce a significant number of additional, thus complementary features to state-of-the art interest points.

## 3.5 Object Recognition

In this experiment, Pascal 2008 data [4] was used to compare segment based corner features to MSER/Hessian points all combined with SIFT. We applied pyramid match kernel (PMK) approach with SVM from [7] with 4 pyramid levels and branch factor equal 20. The PMK was trained for 20 object classes on the training set consisting of 2111 images, and tested on the validation set of 2221 images. As a reference approach, we applied dense feature sampling on a regular grid with the sampling interval of 8, 14, 20, and 26 pixels. This gave 3690 features on average per image. SIFT descriptors for both reference and segmentation based interest points were generated with patch radii of 16, 24, 32, and 40 pixels. Forcing the fourfold scales upon MSER/HE also resulted in their best performance compared to an affine/scale invariant configuration. The MSER, HE, MS, and WA corner features were tested for two different numbers of descriptors per image. In addition, we show results for the Watershed based detector without the anisotropic filter (WS) to demonstrate its advantage. Table 1 shows the mean average precision for all 20 object categories. Experiments performed in [16] explain poor performance of MSER/HE. WA gave the highest scores of 35.61% MAP. It required 1.3x less features than in case of the dense sampling (33.77% MAP). With 2.3x less features, WA was still on a par with the reference approach. This clearly demonstrates the saliency of the segmentation based features and contrasts with [16]. Also, these results validate our observations from section 3.3 on a larger dataset. Note that they are not directly comparable with top scores for such benchmarks in the literature as we used only one kernel and the validation data set for testing.

| features | dense | he | mser | ms | wa | ws |
|---|---|---|---|---|---|---|
| #regions per img | - | 1710 | 1677 | 1674 | **1609** | 1785 |
| MAP (%) | - | 30.78 | 31.37 | 32.51 | **33.76** | 31.78 |
| #regions per img | **3690** | 2417 | 3886 | 2254 | **2841** | 2796 |
| MAP (%) | **33.77** | 31.49 | 33.00 | 33.35 | **35.61** | 33.14 |

Table 1: MAP results for Pascal 2008 recognition benchmark.

## 4 Conclusions

The performed experiments aimed at investigations of the segmentation repetitive features suitable for matching and recognition systems. The best performer for structured scenes was WA while MS was the second best in this category and first best for natural (well textured) images. These two were followed by EGO yielding slightly lower repeatability scores. Region based interest points proved fairly stable, though such detectors would benefit from a selection scheme based on a stability measure similar to the one applied in MSER.

The junctions of segments were proved as very stable features with means of SUSAN. Even in case of poor region-related performance (e.g. under-segmentation), EGO still yielded good results when matching with SIFT. Again, WA and MS turned out two most stable segmentations. It emerged that interest points based on strong boundary curvature are more suitable for both matching and recognition than simple blob based features. It seems that repeatability, matching and recognition benefit from the well- and over-segmentation strategies since they produce higher numbers of stable features. Although using the information carried by segmentation maps may seem a daunting task due to their instability, they were shown to convey high level of similarity information which was captured by interest points.

# 5    Acknowledgements

# References

[1] P. Arbelaez, C. Fowlkes, and D. Martin. The berkley segmentation dataset and benchmark, 2007. `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench`.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[3] F. J. Estrada and A. D. Jepson. Quantitative evaluation of a novel image segmentation algorithm. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2:20Ű–26, 2005.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, , J. Winn, and A. Zisserman. The pascal visual object classes challenge 2008 (voc2008) results, 2008. `http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/workshop/index.html`.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2005.

[6] F. Ge, S. Wang, and T. Liu. Image-segmentation evaluation from the perspective of salient object extraction. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 1: 1146–1153, 2006.

[7] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. *In Proc. of the International Conference on Pattern Recognition*, 2:1458–1465, 2005.

[8] V. Hedau, H. Arora, and N. Ahuja. Matching images under unstable segmentations. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[9] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. Itk software guide, 2005. `http://www.itk.org/ItkSoftwareGuide.pdf`.

[10] D. G. Lowe. Evaluation of interest point detectors. *In Proc. of the IEEE Internatioanl Conference on Computer Vision and Pattern Recognition*, 2:1150–1157, 1999.

[11] T. Malisiewicz and A. Efros. Improving spatial support for objects vis multiple segmentations. *In Proc. of the British Machine Vision Conference*, 2007.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluation segmentation algorithms. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition and Measuring Ecological Statistics*, 2:416Ű–425, 2001.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. *In British Machine Vision Conference*, pages 384–393, 2002.

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. van Goll. A comparison of affine region detectors. *International Journal of Computer Vision*, (65):43–72, 2005.

[16] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. *In Proc. of the European Conference on Computer Vision*, pages 490–503, 2006.

[17] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[18] B. C. Russel, A. A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1605–1614, 2006.

[19] C. Schmid, R. Mohr, and C. Buackhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[20] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2002.

[21] M. Smith and J. M. Brady. Susan-a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[22] T. Tuytelaars and K. Mikolajczyk. A survey on local invariant features. *Foundations and Trends in Computer Graphics and Vision*, (3):177–280, 2008.

[23] H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 2(110):260–280, 2008.