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

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. Our evaluation provides a new insight into suitability of the segmentation methods for generating local features for image retrieval or recognition.

Methodology. This study focuses on gauging performance of Efficient Graph-Based Image Segmentation (EGO), Mean Shift (MS), modified Watershed (WA) and modified Normalised Cuts (NC) in terms of their stability. It also compares them to state-of-the art MSER and Hessian [3] interest point detectors. Inspired by evaluation of affine region detectors [2, 3], we focused on two kinds of key-points 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 [4] is very well tailored to detect corners and junctions on segment boundaries. 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. Finally, non-minima are suppressed, and corners and junctions detected. Figure 1(bottom right) illustrates both types of features.

To quantify performance, we exploited a set of well-known test images from [3]. 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. We also employed two complementary measures based on the homography ground-truth. The region overlap from [3] 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 $\varepsilon_o \leq 0.3$. For the boundary based points the correspondences were considered correct if the distance between the interest point and its nearest projected correspondence satisfied $\varepsilon_n \leq 4$ pixels. The nearest neighbour (NN) repeatability measure was applied to quantify the accuracy of segment boundaries. The goal of adopting the overlap based repeatability [3] was to examine to what extent segments from a given segmentation are roughly preserved over a wide range of transformations. Figure 1 visualises the overlap (bottom third from left) and distance (bottom right) based correspondences.

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) Segmentation maps of EGO, MS, NC, and WA on well-segmented set (left to right). (Bottom) Segment (third from left) and boundary (rightmost) based features (yellow) in the reference image together with their correspondences (black) projected from another image.

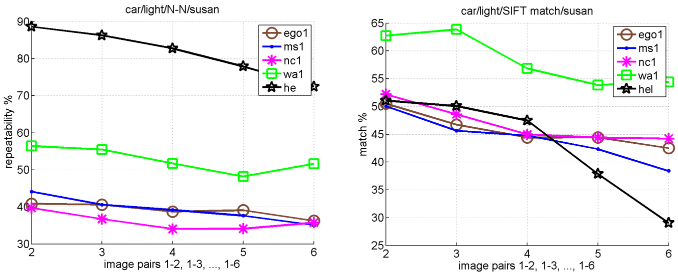

Figure 2: Repeatability of segmentation based SUSAN corners (left) and their matching scores (right) for bike from the over-segmented set.

Results. The repeatability of segmentation corner based features between the original and subsequently distorted images is presented in figure 2(left) and matching results in figure 2(right). This is a single chosen scene. 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 better on structured scenes whilst MS was the clear winner for the natural scenes. Matching of SUSAN corners brought prime results. 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 which showed HE as more repeatable than any of tested methods.

Further, Pascal 2008 data [1] was used to compare segment based corner features to MSER/Hessian [3] and densely sampled points all combined with SIFT. Table 1 shows the mean average precision for all 20 object categories. Note that this is 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. Concluding, the junctions of segments (especially WA based) were proved as very stable features with means of SUSAN. Key-points based on strong boundary curvature seem more suitable for both matching and recognition than simple blob based features.

features	dense	he	mser	ms	wa
#regions per img	**3690**	2417	3886	2254	**2841**
MAP (%)	**33.77**	31.49	33.00	33.35	**35.61**

Table 1: MAP results for Pascal 2008 recognition benchmark.

[1] 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`.

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

[3] 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.

[4] 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.