

# Shape-Stable Region Boundary Extraction via Affine Morphological Scale Space (AMSS)

Petros Kapsalas  
School of Electrical & Computer  
Engineering, National Technical  
University of Athens  
9, Iroon Polytechniou str.,  
15773 Athens, Greece  
+30-210-7723039  
pkaps@image.ntua.gr

Stefanos Kollias  
School of Electrical & Computer  
Engineering, National Technical  
University of Athens  
9, Iroon Polytechniou str.,  
15773 Athens, Greece  
+30-210-7722488  
stefanos@cs.ntua.gr

## ABSTRACT

In this paper we present a new approach towards the extraction of affine image regions based on detecting shape-stable boundaries from a multi-scale image representation. We construct an affine morphological scale space (AMSS) representation [1], which performs anisotropic diffusion while preserving boundaries and being invariant to affine transformations. We extract the transition boundaries of the diffusivity velocity map and track their evolution at each level of the scale-space. We then determine the stability of the boundary shape through a minimization process over different scales. Unlike most state-of-the-art detectors which use the Gaussian scale space for multi-scale image representation, our approach is intrinsically affine invariant. We evaluate our detector by measuring repeatability of regions in transformed images of the same scene and comparing it to the state-of-the-art region detectors [2].

## 1. INTRODUCTION

In many object recognition tasks, within-class changes in pose, lighting, colour, and texture can cause considerable variation in local intensities. Consequently, local intensity no longer provides a stable detection cue. As such, intensity-based interest operators (e.g., Harris, Kadir [3][4])—and the object recognition systems based on them—often fail to identify discriminative features. An alternative to local intensity cues is to capture semi-local structural cues such as edges and curvilinear shapes [5]. These structural cues tend to be more robust to intensity, colour, and pose variations. As such, they provide the basis for a more stable interest operator, which in turn improves object recognition accuracy. This paper proposes a region detector based on affine image representation. The method introduces a new region extraction scheme that exploits curvilinear structures to reliably detect salient areas. The proposed Shape Stable Region Boundaries (SSRB) detector identifies structures of stable shape within the affine invariant image representation.

Curvilinear structures are lines (either curved or straight) such as roads in aerial or satellite images or blood vessels in medical scans. These curvilinear structures can be detected over a range of viewpoints, scales, and illumination changes. We develop a

process that detects structural regions efficiently and robustly across the levels of the Affine Morphological Scale Space (AMSS). The basic idea of the approach is to exploit topology of transition boundaries between neighbouring image structures. In order to provide sufficient determination of these regions we use a mathematical formalization that utilizes the fundamental property of the AMSS which associates scale-space diffusivity velocity to boundary occurrence. Intra-region smoothing is favored towards inter-region smoothing which means that points close to inter-region boundaries are evolved with higher velocities. The basic idea of the approach is to track the evolution of significant image structures across the pyramid scales and select the structures of interest in terms of shape stability criteria. Thus, the transition boundaries between neighbouring image structures are indicated by locating the maxima of diffusivity velocity. Subsequently, shape features of the extracted boundaries are evaluated. Scale selection is associated to finding shape stable structures, and the stability is determined through the minimization of Euclidean distance regarding structures shape features along consequent pyramid levels.

This work makes two contributions. First, we develop a new operator based on morphological representation for estimating transition boundaries between different structures while we also utilise scale space properties to enhance boundaries' connectivity. Second, we introduce an extended definition of structure stability which is based on estimating the variation of shape features. The latter contribution is employed in order to take advantage of the morphological scale space representation that we have embodied in the extraction scheme and which in turn respects image structures, without disturbing their shape.

Section 2 presents related work, while the proposed method is described in Section 3. An experimental study is presented in Section 4, comparing our approach to the best state-of-the-art methods. Conclusions and description of further work are given in Section 5.

## 2. RELATED WORK

Interest operators can typically be classified into two categories: intensity-based detectors and structure-based detectors [2]. Intensity-based detectors depend on analyzing local differential

geometry or intensity patterns to find points or regions that satisfy some uniqueness and stability criteria. The Harris-affine and Hessian-affine detectors [3][6][7] compute maximum determinants of the second moment matrix and the Hessian matrix respectively across scale space and then apply Laplacian-based characteristic scale selection [8] and second-moment-matrix-based shape adaptation [9]. Maximally Stable Extremal Regions (MSER) [11] uses a threshold selection process to detect stable regions that are either brighter or darker than the surrounding region. Shift Invariant Feature Transform [12] (SIFT) (i.e., the Difference of Gaussians (DoG) extrema detector used by Lowe in [11]) finds local extrema across three consecutive difference-of-Gaussian scales and then removes spurious detections via a DoG-response threshold followed by a Harris-like metric to eliminate edge detections. Kadir's salient region detector [4] calculates the entropy of the probability density function (PDF) of intensity values over various scales to find regions with entropy extrema. Other intensity-based detectors include SUSAN (acronym of the method defined by the term Smallest Univalued Segment Assimilating Nucleus) [4] [12][13], Intensity (extrema-)Based Regions (IBR) [12][15], and the work of Moravec [12][15] and Beaudet [12][16].

Structure-based detectors depend on structural image features such as lines, edges, curves, etc. to define interest points or regions. Early structure-based detectors analyze various 2D curves such as the curvature primal sketch or B-splines extracted from edges, ridges, troughs, etc. and then selected high curvature points, line or curve intersections, corners, ends, bumps, and dents as interest points [12][13]. Tuytelaar's Edge-Based Region (EBR) detector [9] fits a parallelogram defined by Harris corner point and points on two adjacent edge contours (extracted by the Canny detector). Scale-Invariant Shape Features (SISF) detects circles at different locations and scales by evaluating salient convex arrangements of canny edges based on a measure that maximizes how well a circle is supported by surrounding edges.

### 3. SHAPE STABLE REGION BOUNDARY

As discussed above, we aim at detecting covariant regions along the multi-scale representation. The first difference with the other state-of-the-art methods is that we consider a truly affine invariant scale-space instead of adapting the linear scale-space. The affine invariant scale space is defined through the following partial differential equation presented below:

$$\frac{\partial u}{\partial t} = |Du| \text{curv} \frac{1}{|\xi|} \quad (1)$$

$$u|_{t=0} = u_0 \quad (2)$$

The implied notation in the above equation is:  $Du$  denotes the gradient  $\nabla u$  and  $\text{curv} \frac{1}{|\xi|}$  denotes the curvature of the level  $u$ . The evolution scheme is based on diffusive interpretation of the equation. Indeed, if  $\xi$  is unit vector such that  $\nabla u \cdot \xi = 0$  and  $u_{\xi\xi}$  the second derivative of  $u$  in the direction  $\xi$  we have:  $\frac{\partial u}{\partial t} = u_{\xi\xi}$ . This

formulation yields an anisotropic operation, in the sense that diffusion takes place only in the  $\xi$  direction depending on the gradient (compare with the isotropic diffusion  $\frac{\partial u}{\partial t} = \Delta u$  i.e., the

heat equation). Equation (2) can be rewritten as  $\frac{\partial u}{\partial t} = \nabla u \cdot \frac{1}{|\xi|} \nabla u \cdot \frac{1}{|\xi|}$

where  $\nabla u \cdot \frac{1}{|\xi|}$  denotes the speed of diffusion. The latter evolution term is considered here as the diffusivity velocity and it is used to discern adjacent structures within the same image. Exploiting the mathematical properties of the diffusivity velocity operator we can conclude that it highlights regions belonging to transition boundaries while smoothing intra-region structures. The operability of our Shape-Stable-Region-Boundary extraction approach can be summarized through the following steps:

1. The AMSS Scale Space representation is constructed.
2. The Diffusivity Velocity Map is extracted at each stage of the Scale-Space (regions of high velocity levels correspond to transition boundaries)
3. Broken boundaries are resolved by a twin thresholding scheme which considers the velocity direction in adjacent points. According to the anisotropic diffusion theory, each image point  $p_i$  is evolved with a velocity factor normal to the tangent at the specific point. Thus, velocity vectors at two adjacent points of the same curve form an angle of  $0^\circ$ . Based on the above property we resolve connectivity issues by considering the minimization of the velocities internal vectors at two adjacent image locations.
4. Boundary shape features (moments) are extracted at each stage of the scale-space and a feature vector is constructed, that should be invariant across a range of geometrical transformations. To achieve this, normalized moments are employed in this approach. Assuming that the shape boundary has been represented as a shape feature signature  $z(i)$ , the  $r^{\text{th}}$  moment  $m_r$  and central moment  $\mu_r$  can be estimated as:
$$m_r = \frac{1}{N} \sum_{i=1}^N |z(i)|^r \quad (3)$$

$$\mu_r = \frac{1}{N} \sum_{i=1}^N |z(i) - m_1|^r \quad (4)$$

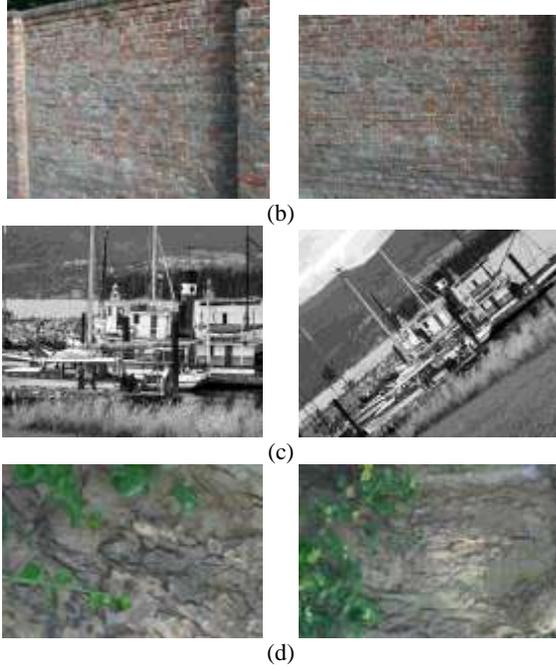
Where  $N$  is the number of boundary points. The normalized moments  $\bar{m}_r = m_r / \bar{\mu}_2^{r/2}$  and  $\bar{\mu}_r = \mu_r / \bar{\mu}_2^{r/2}$  are invariant to shape translation, rotation and scaling
5. The Euclidean distance between feature vectors of the corresponding areas is measured.
6. A region is considered to be stable when its feature vector does not change significantly along consecutive levels of the scale space (the 3 bottom stages of the extraction process will be referred to as propagation stage...)

### 4 EVALUATION

We first summarize the framework for evaluating distinguished region repeatability. A description of the experiments and a discussion of the results follow.



(a)



**Figure 1:** Part of the evaluation dataset. (a) Viewpoint change, (b) viewpoint change, (c), (d) Zoom + rotation

## 4.1 Evaluation Framework

*Repeatability* measures the extent to which regions detected in transformed images of the same scene overlap. We use the evaluation framework presented in [2]. The framework consists of eight images, where each image is subjected to five transformations, resulting in sets of six images. Examples from the image sets are shown in Figure 1. The homographies between the reference images and the other images for each set have been computed, allowing the overlap between distinguished regions in the reference and another image to be evaluated.

In [2], only elliptical distinguished regions are considered, as these are intrinsically produced by four of the six algorithms tested in [2]. For the other algorithms, ellipses approximating the regions are chosen. To be compatible with the framework, we also fit ellipses to the edges of the regions produced by the proposed methods, using the ellipse fitting algorithm in [9].

Repeatability is measured between the reference image and another image from the set. The distinguished regions are detected in both images and those from the second image are projected onto the reference image by using the known homography. Two regions are said to form a region-to-region correspondence if the *overlap error* is sufficiently small— in this paper we use a value of 0.4 as was done for the experiments in [2]. The overlap error is defined as [2]

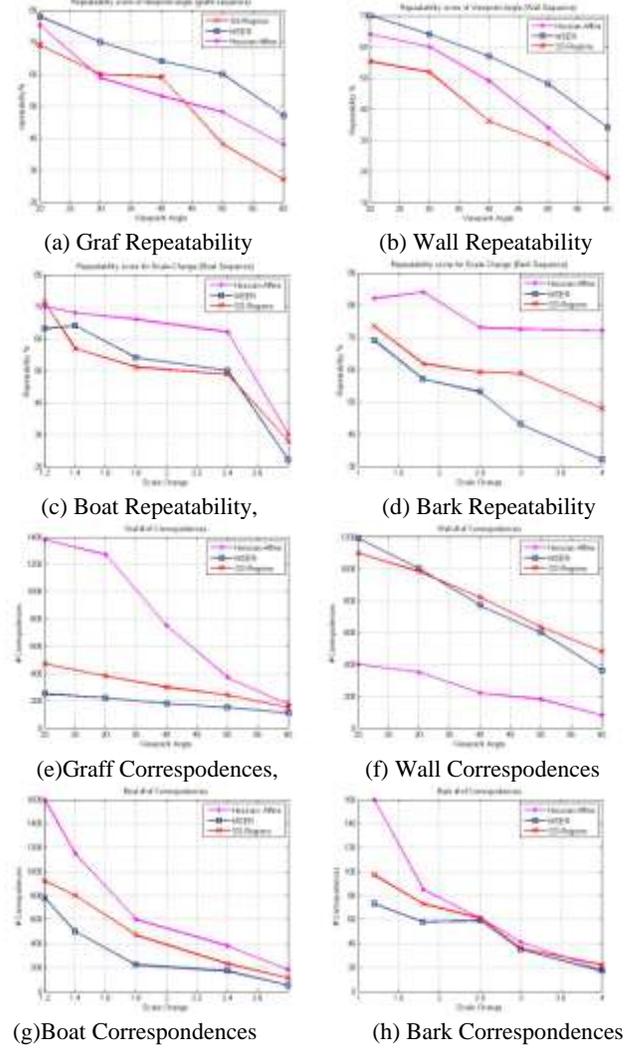
$$1 - \frac{R_{\mu a} \cap R_{\mu b} \mathbf{H}^{-1}}{R_{\mu a} \cup R_{\mu b} \mathbf{H}^{-1}} \quad (5)$$

where  $R_{\mu}$  is the region enclosed by the ellipse defined by  $x^T \mu x = 1$  and  $H$  is the homography relating the images. The *repeatability score* for a pair of images is the ratio between the number of region-to-region correspondences and the smaller number of regions in the pair of images. Only regions located in the part of the scene present in both images are counted. In

addition, the regions are transformed to have a normalised size before calculating the overlap, to avoid the problems with regions of different sizes discussed in [2].

## 4.2 Experiments & Discussion

The results of the repeatability and – number of correspondences – tests for the four groups of six images in the dataset are shown in Figures 2(a)-(h). Curves corresponding to six methods are shown in each graph. The curves labelled *MSEER* and *Hessian-affine* correspond to the two best performing methods of the six tested in the evaluation of affine covariant region detectors in [2]



**Figure 2:** The repeatability and number of correspondences (a, b, e, f) viewpoint angle variation, (c, d, g, h) variation of scale change

As has already been pointed out in [2], different algorithms perform better for different transformations, as can be seen by the repeatability results for the MSEER and Hessian-affine detectors. The problem of comparing detectors producing different densities of regions is discussed in [2]. They point out that for detectors that produce few regions, the thresholds can be set so that the performance is often better than average. For detectors that produce many regions, the image may be so cluttered with regions

that some get matched by accident. The Hessian-affine detector produces the largest number of correspondences for each image sequence except Wall, indicating that the density of the distinguished regions is higher.

For the viewpoint changes (Figure 2(a),(b)), the MSER detector has the highest repeatability. For the Graffiti, Bark and Boat images the SSBR algorithm provided good repeatability scores which are comparable to MSER ones, while for the case of the boat sequence, SSBR outperforms MSER. In an effort to explain the characteristics of SSBR response we should initially exploit its potential to extract reliable structures in images presenting significant edge structures. (graffiti, boat sequence). A further issue regards the SSBR invariance towards viewpoint angle change. The experimental results show that the method provides quite considerable repeatability scores when larger image structures are encountered in the image. One of the drawbacks of the evaluation framework used is that the difference in the number of regions (region density) extracted by each algorithm is not taken into account, which could affect the repeatability results

In an effort to provide visual evaluation of our proposed region extraction approach we illustrate SSBR detected in some images obtained from the Caltech database. Figure 3 shows the symmetrical detections in several images. We can see that the detected regions are quite accurate and distinctive, providing a valuable cue for the detection and recognition of symmetrical objects.



**Figure 3:** Salient region extraction using SSBR

## 5. CONCLUSIONS AND FURTHER WORK

This paper has presented a new structure-based interest region detector called Shape Stable Region Boundaries (SSBR) and has demonstrated its successful application to several tasks. The SSBR interest operator detects stable region boundaries within the affine morphological scale space representation that describes both edge and curvilinear structures. Anisotropic diffusion properties to track the evolution and enhance the connectivity of shape stable boundaries along with a new shape stability estimation scheme are introduced in this work. Further, SSBR achieves robust detection across multiple scales by selecting stable regions across consecutive scales.

Experiments measuring the repeatability of the extracted regions for different types of image transformations are presented. The obtained repeatability falls into the range of the repeatability of the 2 most effective algorithms tested in [2], without surpassing

them but providing quite comparable results. One of the drawbacks of the evaluation framework used is that the difference in the number of regions (region density) extracted by each algorithm is not taken into account, which could affect the repeatability results.

We have so far used shape features to measure the stability of the extracted regions. However, the derived results could be enhanced by re-considering some other image structures characteristics such as region extent or entropy. A further direction that could benefit the extraction of stable regions is to associate region tracking to image evolution as the latter could be derived by the AMSS evolution law.

## 5 REFERENCES

- [1] Alvarez, L., Guichard, F., Lions, P. L., Morel, J. M., Axioms and Fundamental Equations of Image Processing, *J. of Arch. Rational Mech. Anal.*, 123 (1993), pp. 199-257.
- [2] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005, A comparison of affine region detectors. *International Journal of Computer Vision*, (2005), 65(1/2), 43-72
- [3] Harris, C., and Stephens, M., 1988, A combined corner and edge detector. *Alvey Vision Conf.*, (1988), 147-15
- [4] Kadir, T., and Brady, M., 2001, Scale, saliency and image description, 2001, *IJCV*, 45(2):83-105..
- [5] Steger, C., 1998, An unbiased detector of curvilinear structures.(1998), *PAMI*, 20(2):113-125.
- [6] Mikolajczyk, K. and Schmid, C., 2002. An affine invariant interest point detector, 2002., *ECCV*, 1(1):128-142,
- [7] Mikolajczyk K., and Schmid, C., 2004, Scale and affine invariant interest point detectors. *IJCV*, (2004), 60(1):63-86
- [8] Lindeberg, T., 1998, Feature detection with automatic scale selection. (1998) *IJCV*, 30(2), 79-116..
- [9] Lindeberg, T., and Garding, J., 1997, Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. (1997), *Image and Vision Computing*, pages 415-434.
- [10] Baumberg, A., 2000, Reliable feature matching across widely separated views. (2000) *CVPR*, pages 774-781.
- [11] Matas, J., Chum, O., Urban, M., and Pajdla. T., 2004, Robust widebaseline stereo from maximally stable extremal regions, (2004), *Image and Vision Computing*, 22(10):761-767.
- [12] Lowe, D. G., 2004, Distinctive image features from scale-invariant Keypoints, (2004), *IJCV*, 60(2):91-110.
- [13] Smith, S., and Brady, J. M., 1997, Susan—a new approach to low level image processing, (1997), *IJCV*, 23(1), 45-78
- [14] Tuytelaars, T., and Gool, L. V., 2000, Wide baseline stereo matching based on local, affinely invariant regions., (2000), *BMVC*, pages 412-425.
- [15] Moravec, H., 1977, Towards automatic visual obstacle avoidance, (1977). *International Joint Conf. on Artificial Intelligence*, page 584.
- [16] Beaudet, P., 1978, Rotationally invariant image operators. (1978), *ICPR*, pages 579-58

