Stable Salient Shapes

P. Martins *, P. Carvalho *, C. Gatta †
*Centre for Informatics and Systems (CISUC)
University of Coimbra, Coimbra, Portugal
{pjmm, carvalho}@dei.uc.pt
† Computer Vision Centre (CVC)
Autonomous University of Barcelona, Barcelona, Spain
cgatta@cvc.uab.es

***Abstract*—We introduce Stable Salient Shapes (SSS), a novel type of affine-covariant regions. The new local features are obtained through a feature-driven detection of Maximally Stable Extremal Regions (MSERs). The feature-driven approach provides alternative domains for MSER detection. Such domains can be viewed as saliency maps in which features related to semantically meaningful structures, e.g., boundaries and symmetry axes, are highlighted and simultaneously delineated under smooth transitions. Compared with MSERs, SSS appear in higher number and are more robust to blur. In addition, SSS are designed to carry most of the image information. Experimental results on a standard benchmark are comparable to the results of state-of-the-art solutions in terms of repeatability score. The computational complexity of the method is also worth of note, as it is lower than those of most of the competing algorithms in the literature.**

I. Introduction

Local image feature detection is a prolific and prominent research topic for the computer vision community. It has proved to be an effective tool in the resolution of a number of vision problems, including stereo vision, image matching and object recognition. The main purpose of local feature detection is to provide a reliable image representation by identifying a sparse set of visually relevant image regions.

Local feature detection is closely related to *feature description*, which aims at summarising the local images patches covered by the detected regions. Some applications require a local feature detection that is covariant with a class of image transformations, e.g., affine ones. A covariant region detection will potentiate an effective invariant description with respect to the same class of image transformations.

In the particular case of affine covariant detectors, many approaches have been suggested in the literature. We will shortly review the most prominent solutions; we refer the reader to the work of Tuytelaars et al. [1] and references within for a more detailed and comprehensive review.

The Hessian-Affine (HESAFF) and the Harris-Affine (HARAFF) detectors [2], [3] are two popular affine covariant solutions based on image derivatives. The former is a combination of a multi-scale Harris-Stephens interest point detector [4] with an *affine shape adaptation* scheme [5]. From initial estimates of interest points detected at their characteristic scales, the algorithm converges to affine covariant locations via an iterative process – the so-called affine shape adaptation–, that estimates elliptical affine covariant regions around the interest points, whose shape is defined by the structure tensor matrix. In the same manner, the Hessian-Affine detector adopts the affine shape adaptation scheme; however, the initial estimates are given by the determinant of the Hessian matrix.

The major shortcoming of the Hessian-Affine and the Harris-Affine detectors is their computational complexity, as they are $\mathcal{O}((s+i)p)$ algorithms, where s denotes the number of scales used in the automatic scale selection, which determines the characteristic scale, i is the number of iterations used in the shape adaptation mechanism, and p is the number of points from the initial estimate.

Affine covariant regions can also be derived from *extremal regions*. In the image domain, an extremal region corresponds to a connected component whose corresponding pixels have either higher or lower intensity than all the pixels on its boundary. Extremal regions hold two important properties: the set of extremal regions is closed under continuous transformations of image coordinates as well as monotonic transformations of image intensities. The Maximally Stable Extremal Regions detector [6] responds to extremal regions that are stable with respect to intensity perturbations. In its original implementation, the MSER detector enumerates extremal regions using a union-find algorithm, whose complexity is $\mathcal{O}(n \log\log n)$, where n denotes the number of pixels. The high repeatability rates shown by the MSER detector in structured images and the suitability of MSERs to be described either by photometric or by shape descriptors [7], [8] have made the MSER detector a prominent reference in the literature.

In this paper, we extend the work presented in [9] to propose a new type of affine covariant features, Stable Salient Shapes (SSS), which are the result of performing a feature-driven detection of Maximally Stable Extremal Regions.

The first step of the algorithm provides saliency maps that will be posteriorly used as domains for MSER detection. In such maps, features that are linked to semantically meaningful structures – such as boundaries and symmetry axes –, are highlighted. Unlike the original edge-driven MSER detection, this extended version contains a combined feature highlighting, which allows the detector to retrieve a higher number of regions without redundancy. Our algorithm overcomes major limitations of a standard MSER detection, namely the sensitivity to image blur, the presence of a reduced number of regions,

and the biased preference for regular shapes. All of this is achieved without significantly increasing the computational complexity of the algorithm.

II. Background and Motivation

The main motivation for our work comes from the well-known advantages in obtaining affine covariant regions from extremal regions as well as the shortcomings that such strategy entails. For a better understanding of our motivation, we will start by providing the basic terminology and the formal definitions around the concepts of extremal regions and MSERs.

A. Preliminaries and Definitions

Let us denote by *image* I the mapping $I : \mathcal{D} \subset \mathbb{Z}^2 \rightarrow \{0, 1, \ldots, M\}$. A *connected component* (or *region*) $\mathcal{Q}$ in $\mathcal{D}$ is a subset of $\mathcal{D}$ for which each pair of pixels $(\mathbf{p}, \mathbf{q}) \in \mathcal{Q}^2$ is connected by a *path* in $\mathcal{Q}$, i.e., there is a sequence $\mathbf{p}, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m, \mathbf{q} \in \mathcal{Q}$ such that $\mathbf{p} \sim \mathbf{a}_1, \mathbf{a}_1 \sim \mathbf{a}_2, \ldots, \mathbf{a}_m \sim \mathbf{q}$, where $\sim$ denotes the equivalence relation defined by $(\mathbf{p} \sim \mathbf{q}) \iff \max\{|p_1 - q_1|, |p_2 - q_2|\} \leq 1$ (*8-neighbourhood*).

We define the *boundary* of a region $\mathcal{Q}$ as the set $\partial\mathcal{Q} = \{\mathbf{p} \in \mathcal{D} \backslash \mathcal{Q} : \exists \mathbf{q} \in \mathcal{Q} : \ \mathbf{p} \sim \mathbf{q}\}$. A connected component $\mathcal{Q}$ in $\mathcal{D}$ is an *extremal region* if $\forall p \in \mathcal{Q}, \mathbf{q} \in \partial\mathcal{Q} : I(\mathbf{p}) < I(\mathbf{q})$ or $I(\mathbf{p}) > I(\mathbf{q})$.

Let $\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_{i-1}, \mathcal{Q}_i, \ldots$ be a sequence of extremal regions such that $\mathcal{Q}_k \subset \mathcal{Q}_{k+1}, k = 1, 2, \ldots$. We say that $\mathcal{Q}_i$ is a *maximally stable extremal region* if and only if the *stability criterion*

$$\rho(k, \Delta) = \frac{|\mathcal{Q}_{k+\Delta} \setminus \mathcal{Q}_k|}{|\mathcal{Q}_k|}, \tag{1}$$

attains a local minimum at i, where Δ is a positive integer denoting the *stability threshold*. As area ratios are preserved under affine transformations, ρ is an affine invariant measure. Consequently, MSERs are covariant with these geometric transformations.

We note that the detection of MSERs is related to image thresholding, since every extremal region is a connected component of a thresholded image [1]. In fact, we can alternatively describe the detection of MSERs as a process that considers all the possible thresholdings of an intensity image. An extremal region is considered maximally stable if it shows a small area change across several thresholdings [10]. By increasing the threshold, we detect MSERs$+$, which correspond to dark regions with brighter boundaries. MSERs$-$, which are brighter regions with dark boundaries, can be obtained through the same process by inverting the input intensity image.

B. Extremal Regions and Maximally Stable Extremal Regions: downsides

Some of the major limitations of MSER detection have been summarised in the comparative study on affine covariant regions performed by Mikolajczyk et al. [11]. In this study, MSERs and regions detected with the Hessian-Affine detector have shown higher repeatability scores. However, the former has shown an inconsistent performance: sequences of images with different levels of blur as well as textured sequences have produced less repeatable MSERs. If we take into account that MSERs tend to anchor on region boundaries — which suggests that the method is more suitable to deal with well-structured scenes [1] –, one can expect less repeatable results on textured scenes. The sensitiveness to image blur is explained by the undermining effect that it has on the stability criterion.

Another important downside of MSER detection is related to the number of regions that the detector retrieves. MSERs tend to be in lower number than other regions retrieved by detectors such as the Hessian-Affine or the Harris-Affine. A reduced number of features may not provide the best coverage of the content, which impairs the robustness of the method to object occlusions.

Kimmel et al. [12] have observed that the affine covariance of MSERs is verified if and only if objects possess smooth boundaries. As the affine covariance of MSERs is an immediate consequence of the covariance of the image level sets with affine transformations of the coordinates, it is required that the point-spread function of camera lenses is small compared to the natural blurring of objects. The authors have also noted that the stability criterion as defined in (1) prefers regular (round) shapes to irregular ones. This bias for regular shapes has been demonstrated by showing that if two regions have the same area and the same intensity along the boundaries, the one with a shorter boundary will yield a lower value of ρ.

C. Refinements on the MSER detector

The MSER detector has been extended to deal with colour images [10] and video sequences [13]. It has also been subject to several refinements. In [8], Forssen introduced an alternative MSER detector that makes use of a multi-scale pyramid representation with one octave between scales. This multi-resolution approach detects MSERs at each resolution and duplicated regions at consecutive scales are removed by discarding fine scale MSERs with similar locations and sizes as regions detected at the next coarser scale. The multi-resolution MSER detector produces a higher number of regions and is more robust to image blur and scale changes. On the downside, it requires a detection at each scale, which increases the computational complexity of the algorithm. The algorithm for the detection of Stable Affine Frames (SAFs) [14] can be regarded as a refinement on the MSER detector. SAFs lie on the boundary of extremal regions. Unlike MSERs, the stability of SAFs with respect to intensity perturbations is measured locally, i.e, we do not require the whole boundary to be stable to intensity changes. This algorithm produces a higher number of features and covers more evenly the image content. Moreover, SAFs are more repeatable in the presence of image blur. On the downside, the method requires a considerably higher computational effort.

Kimmel et al. [12] free the MSER detector from the preference towards regular shapes by presenting several redefinitions of the stability criterion, which prefer irregular shapes and are still affine invariant. The main goal of these reinterpretations is to define more informative shape descriptors.

III. Stable Salient Shapes

At a first glance, the ideal image for the MSER detector is the one that is well structured, with uniform regions separated by strong intensity changes [1]. However, the affine covariance holds for MSERs if and only if the boundaries of the objects in the scene are smooth [12]. These are the principles in which the SSS detector is based on. As we have stated in the introductory section, we can succinctly describe the construction of Stable Salient Shapes as a feature-driven detection of MSERs, where such features correspond to structures related to objects boundaries or even symmetry axes.

The first step of the proposed algorithm, coined as *feature highlighting*, consists in building a saliency map for each one of the features to be highlighted. These maps are intended to be suitable domains for MSER detection. There are two main reasons for emphasising features related to structures such as boundaries or symmetry axes. First, the semantic meaningfulness of these structures is high as they carry a relevant amount of image information, including fine details [15], [16], [17]. Second, we note that the boundaries of MSERs often correspond to objects boundaries. Since the stability of an extremal region with respect to intensity changes is measured at its boundary, we will be reshaping objects boundaries to yield more stable extremal regions.

The subsequent step of the algorithm consists in detecting MSERs on each of the saliency maps. We cannot expect a full complementarity among the extremal regions detected along the different maps, since they are related to the same structures. The third and final step of the algorithm takes into account the potential overlapping or even the duplication of MSERs and performs a region pruning. Figure 1 depicts the main steps of SSS detection.

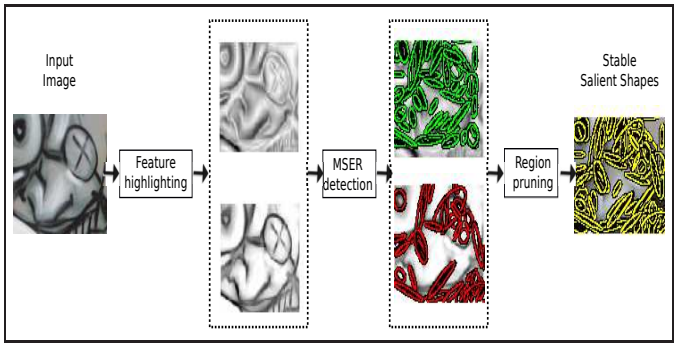

Fig. 1. Algorithm for the detection of Stable Salient Shapes (feature-driven MSERs). The first step of the algorithm produces saliency maps that will be used in the next step as input images for MSER detection. The maps emphasise features that are related to semantically meaningful structures, such as boundaries and symmetry axes.

A. Feature highlighting

We perform the highlighting of edges and ridges by means of two differential-based measures, which yield two different saliency maps. Edges are undoubtedly important features in an image, since they reflect the presence and the shape of objects in a scene. Our algorithm relies on the structural information provided by edges. Ridges are related to symmetry axes as they often correspond to the major axis of symmetry of elongated objects.

The first measure intends to highlight edges and, simultaneously, delineate smooth transitions at the boundaries, whereas the second one highlights ridges while preserving the aforementioned smoothness. The detection of structures at different scales will help us to define smooth transitions. The process of averaging information over scales is the key component to obtain the desired smoothness. For the averaging procedure, normalised Gaussian derivatives [18] are utilised.

1) Edge highlighting: We highlight edges and obtain smooth transitions by making use of the gradient magnitude, computed by means of Gaussian derivatives at several scales. Let $L(:,\sigma)$ be a smoothed version of image I by means of a Gaussian kernel G at the scale σ^2, i.e., $L(\mathbf{x},\sigma) = (G(\sigma) * I)(\mathbf{x})$. The edge strength can be found by measuring the gradient magnitude,

$$|\nabla L(\mathbf{x},\sigma)| = \sqrt{L_x^2(\mathbf{x},\sigma) + L_y^2(\mathbf{x},\sigma)}, \tag{2}$$

where L_x and L_y denote the first-order partial derivatives of L in the x and y directions, respectively. From (2), we obtain our measure for edge highlighting:

$$F_1(\mathbf{x}) = \sum_{i=1}^{N} \sigma_i \left|\nabla L(\mathbf{x},\sigma_i)\right|, \tag{3}$$

where the standard deviation σ_i varies in a geometric sequence $\sigma_i = \sigma_0 \xi^{i-1}$, with $\sigma_0 \in \mathbb{R}^+$, $\xi > 1$, and N denotes the number of scales. The final image is, therefore, the result of averaging gradient magnitude computed at different scales. By doing this, with a reasonable number of scales, smooth transitions at the edges will be obtained.

2) Ridge highlighting: The measure for ridge highlighting derives from the Hessian matrix,

$$\mathcal{H}(\mathbf{x},\sigma) = \begin{bmatrix} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{yx}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{bmatrix}, \tag{4}$$

where L_{xx}, L_{xy} and L_{yy} are the second order partial derivatives of L, a Gaussian smoothed version of image I. The *principal curvature* [19], which highlights curvilinear structures is either given by

$$P_{max}(\mathbf{x},\sigma) = \max(0, \lambda_2(\mathcal{H}(\mathbf{x},\sigma))), \tag{5}$$

or

$$P_{min}(\mathbf{x},\sigma) = \min(0, \lambda_1(\mathcal{H}(\mathbf{x},\sigma))), \tag{6}$$

where λ_1 and λ_2 denote the minimum and maximum eigenvalues, respectively. We note that Eqs. (5) and (6) respond to complementary structures: the former responds to dark lines on a brighter background, whereas the latter detects brighter lines on a dark background. From the principal curvature, we obtain the measure for ridge highlighting:

$$F_2(\mathbf{x}) = \sum_{i=1}^{N} \sigma_i^2 P_{max}(\mathbf{x},\sigma_i), \tag{7}$$

where $\sigma_i = \sigma_0 \xi^{i-1}$, with $\sigma_0 \in \mathbb{R}^+$, $\xi > 1$, and N denotes the number of scales.

Our ridge highlighting measure relies on the use of the principal curvature measure to detect darker lines on a bright background. However, the measures defined in (5) and (6) can be used interchangeably, as we perform the enumeration of extremal regions in the resulting saliency map as well as in the inverted saliency map.

In Fig. 2, we depict the saliency maps that are the result of the proposed edge and ridge highlighting, using a license plate as the input image. It is readily seen that both saliency maps preserve the structural information of the image and add some smoothness to the scene. While the edge highlighting mainly captures and accentuates the objects boundaries, the ridge highlighting provides a clearer structural sketch of the scene [19]. For the purpose of MSER detection, we can regard the map that emphasises edges as a more suitable domain, as it generates uniform regions separated by heavy intensity changes and is less sensitive to noise. However, the second map provides us complementary regions, whose detection is important to improve the coverage of the content.

B. MSER detection and region pruning

Apart from the input image, there are no differences between the feature-driven MSER detection and the original one, i.e., we assess the stability of extremal regions using the original stability criterion. Figures 3 and 4 help to illustrate the advantages of our feature-driven MSER detection over the standard one. In Fig. 3, we depict two well-structured scenes and the corresponding isophotes on the luminance channel and on both of the saliency maps. Both maps show a higher number of extremal regions, which can be increased with a higher number of scales. Due to the Gaussian smoothing, the irregularity of extremal regions is attenuated, which will compensate the preference of the stability criterion for regular shapes.

Figure 4 compares standard MSER detection with SSS detection in the presence of Gaussian blur. As blur increases, both types of regions decrease in number. However, this reduction is more significant for MSERs. For $\sigma = 3$, the number of MSERS decreases 19%, whereas the number of SSS decreases 11%. For $\sigma = 10$, the number of MSERs decreases 93%, while the number of feature-driven MSERs decreases 53%. The example also shows that SSS are in higher number and cover the most informative parts of the scene, regardless of the amount of blur.

The two saliency maps do not provide fully complementary regions. Thus, we eliminate regions that are duplicated. To find duplicates, we compare the centroid distance. If this distance is lower than 0.1, we compute the overlap error between the corresponding fitted ellipses. If this error is less than 10%, we discard the region with higher ρ. Figure 5 depicts the proposed region pruning. A previous pruning, which removes regions based on the area or the stability measure ρ, is performed on each map. We will describe it in the upcoming section.

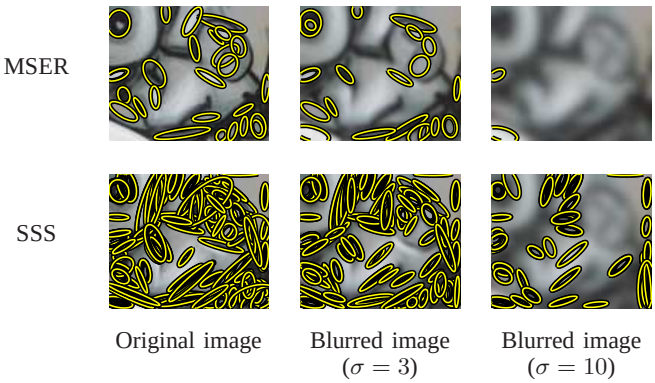

Fig. 4. MSER and SSS detection on images with Gaussian blur (original image, $\sigma = 3$ and $\sigma = 10$). Top row: MSER detection. Bottom row: SSS detection. The detected regions are replaced by fitted ellipses.

Fig. 5. An example of the proposed region pruning. The detected regions are replaced by fitted ellipses.

To conclude this section, we present the results of different detectors on a Siemens star (see Fig. 6). The detection includes the scale covariant SFOP regions [20] and affine covariant features, such as MSERs, Principal Curvature-based Regions (PCBRs) [19], SSS, Harris-Affine, and Hessian-Affine regions. In this example, SSS cover the most informative content without the presence of redundant regions. Moreover, most of the content covered by other regions, is also covered by SSS features .

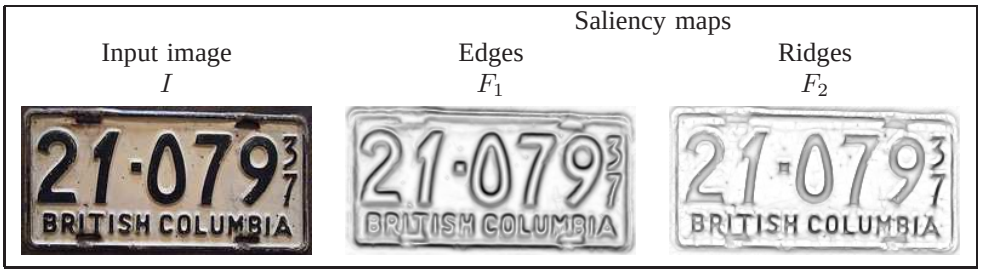

Fig. 2. An example of the proposed feature highlighting. Darker structures in the saliency maps are the most salient ones. To obtain the final saliency maps – F_1 and F_2 –,12 scales were used. The parameters σ_0 and ξ were set to 1 and $\sqrt[4]{2}$, respectively.

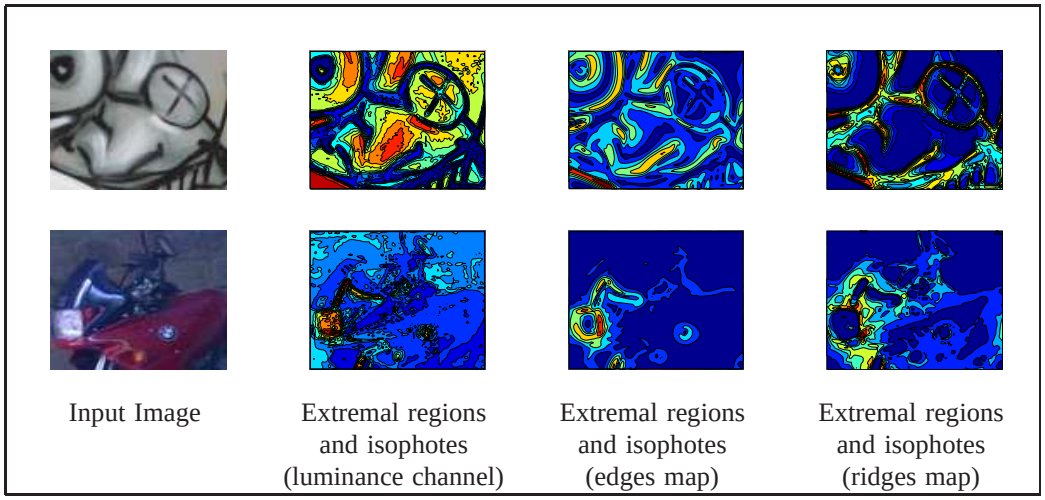

Fig. 3. Regions delineated by isophotes on the different domains.

IV. Experimental Validation

To experimentally assess the performance of the proposed detector, we have followed the guidelines of the standard evaluation protocol proposed by Mikolajczyk et al. in [11], which comprises image sequences with gradually increased effects of geometric and photometric transformations, such as viewpoint changes, scale changes, blur, and JPEG compression. Each sequence contains six images and a ground-truth homography is provided between the first image of the sequence and each one of the five remaining images.

For the sake of clarity, the experimental validation has considered state-of-the-art affine covariant detectors that have been made publicly available. The list includes: the MSER detector, the Hessian-Affine and the Harris-Affine detectors, as well as the PCBR detector. We note that the latter has not been suggested as a generic detector; Deng et al. suggest it for object recognition. However, due to the fact that it detects affine covariant regions from structural information using a multi-scale approach, we have decided to include it in the evaluation. Apart from the MSER detector, all the implementations that we have used are the ones provided and maintained by the authors. For the SSS and MSERs detectors, we have made use of the code provided by Vedaldi and Fulkerson [21]. In the case of the SSS detector, this code has been modified to deal with images whose intensity values vary in a range different from $\{0, \ldots, 255\}$, since the saliency maps intensity values might be greater than 255.

a) Parameter settings: We have built the saliency maps with $\sigma_0 = 1$, $\xi = \sqrt[4]{2}$, and $N = 12$. The stability threshold Δ was set to 20. For the MSER detector, this parameter was set to 10. The minimum and maximum region area were set to 30 and 1% of the image area, respectively, for both of the detectors. In addition, we have only considered MSERs and SSS whose ρ was lower than 0.7. As for the remaining detectors, we have used the default parameter settings.

A. Repeatability evaluation

Repeatability is often regarded as the fundamental criterion to assess the performance of local features. Mikolajczyck et al. defined the repeatability score as the ratio between the number of region correspondences and the smaller number of common regions in a pair of images of the same sequence. As the shapes of affine covariant regions are diverse, each region is replaced by an approximating ellipse. The use of ellipses allows us to define a repeatability score for a given overlap error.

We have performed our evaluation on five sequences (depicted in Fig. 7). The first one is the Graffiti sequence, which contains a well-structured pattern under viewpoint changes. The Wall sequence shows a textured sequence (brick wall) under viewpoint changes. The Boat sequence depicts a relatively textured scene composed of natural and man-made

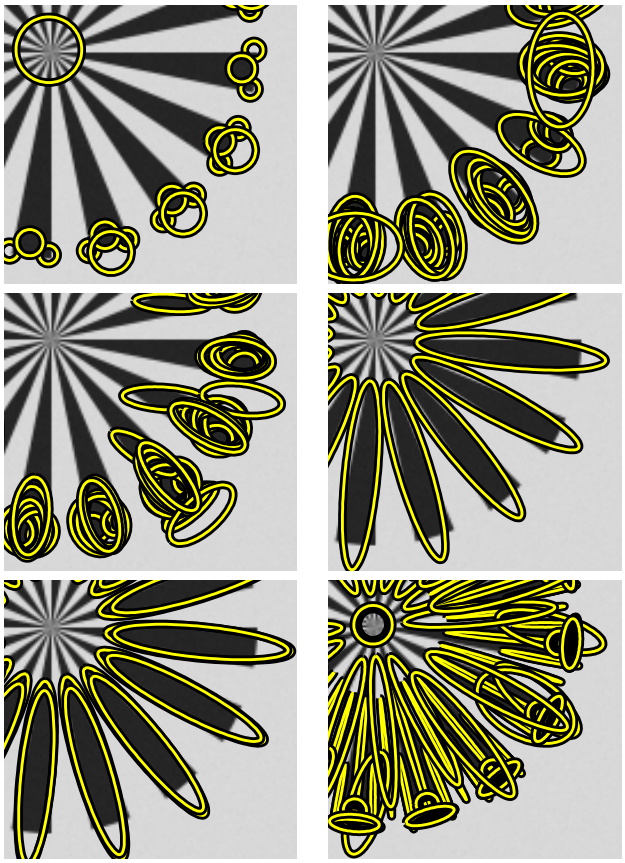

Fig. 6. Local feature detection on the beams of a Siemens star [20]. Top left to bottom right: SFOP, Harris-Affine, Hessian-Affine, MSER, PCBR and SSS. The proposed detection responds to most of the structures detected by the remaining detectors. Only SFOP and SSS detect the centre of the star.

elements under zoom and scale changes. The Bikes sequence shows a scene with man-made objects under de-focus blur. Finally, the UBC sequence depicts a scene composed of man-made and natural objects, which has been corrupted by JPEG compression. Figure 8 shows the output of SSS detection for the reference images of two sequences (Graffiti and UBC).

Figure 9 depicts the repeatability results. One can readily see that SSS tend to appear in higher number. Regardless of the sequence, the SSS detector yields the highest number of correspondences, while its repeatability score is comparable to those of its counterparts. In comparison with MSERs, SSS are more robust to blur and JPEG compression. MSERs exhibit a slightly higher repeatability score for viewpoint changes in the Graffiti sequence as well as for the zoom and rotation variations in the Boat sequence. However, the number of correspondences between MSERs is considerably lower. We note that a substantially higher number of correspondences accompanied by a slight decrease of the repeatability rate is often preferable to a minor increase of the repeatability with less regions, since in the former the absolute number of repeated regions is considerably higher, which might provide a better coverage of the content with a similar repeatability score. Apart from viewpoint changes, the PCBR detector

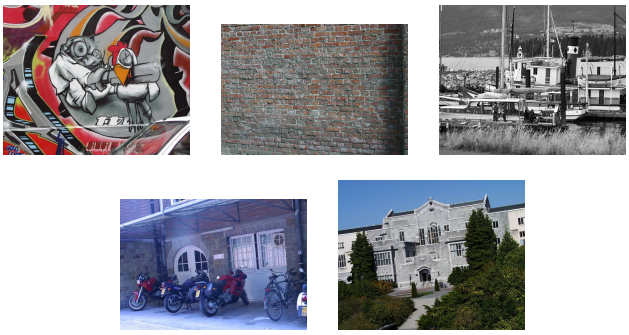

Fig. 7. First images (reference images) of the sequences used in the experiments. First row, from left to right: Graffiti, Wall, Boat. Second row, from left to right: Bikes, UBC.

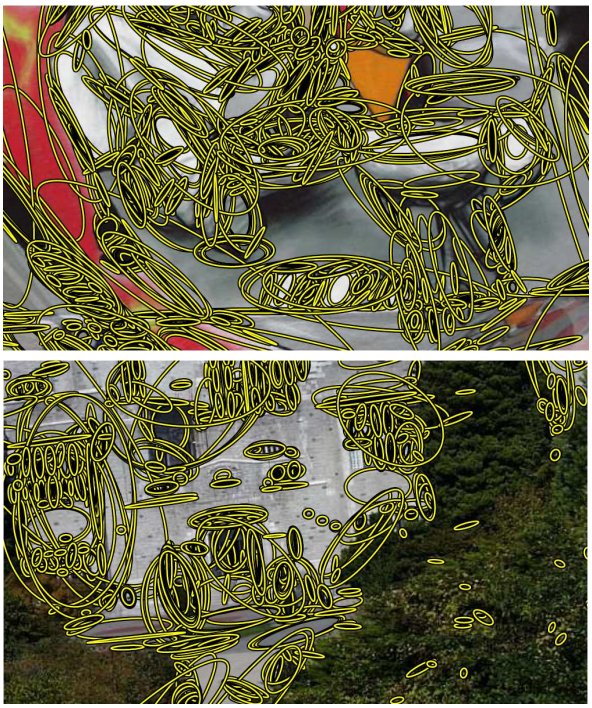

Fig. 8. Examples of SSS detection. Top row: Grafitti reference image (detail), bottom row: UBC reference image (detail).

shows the worst performance, either in terms of repeatability score, or in terms of the number of correspondences.

We have also analysed the accuracy of the detectors. The last two columns in Fig. 9 show the number of correspondences and the repeatability score as a function of the overlap error. A higher overlap error yields more correspondences and a higher repeatability score. We verify that SSS is an accurate detector as well as MSER and PCBR detectors. HARAFF and HESAFF detectors tend to improve their ranking as the overlap increases, which means that the regions retrieved by these detectors are the less accurate among the different types of affine covariant regions.

We have also analysed the trade-off between the number of correspondences and the repeatability score for different stability thresholds as in [14]. This analysis has been extended

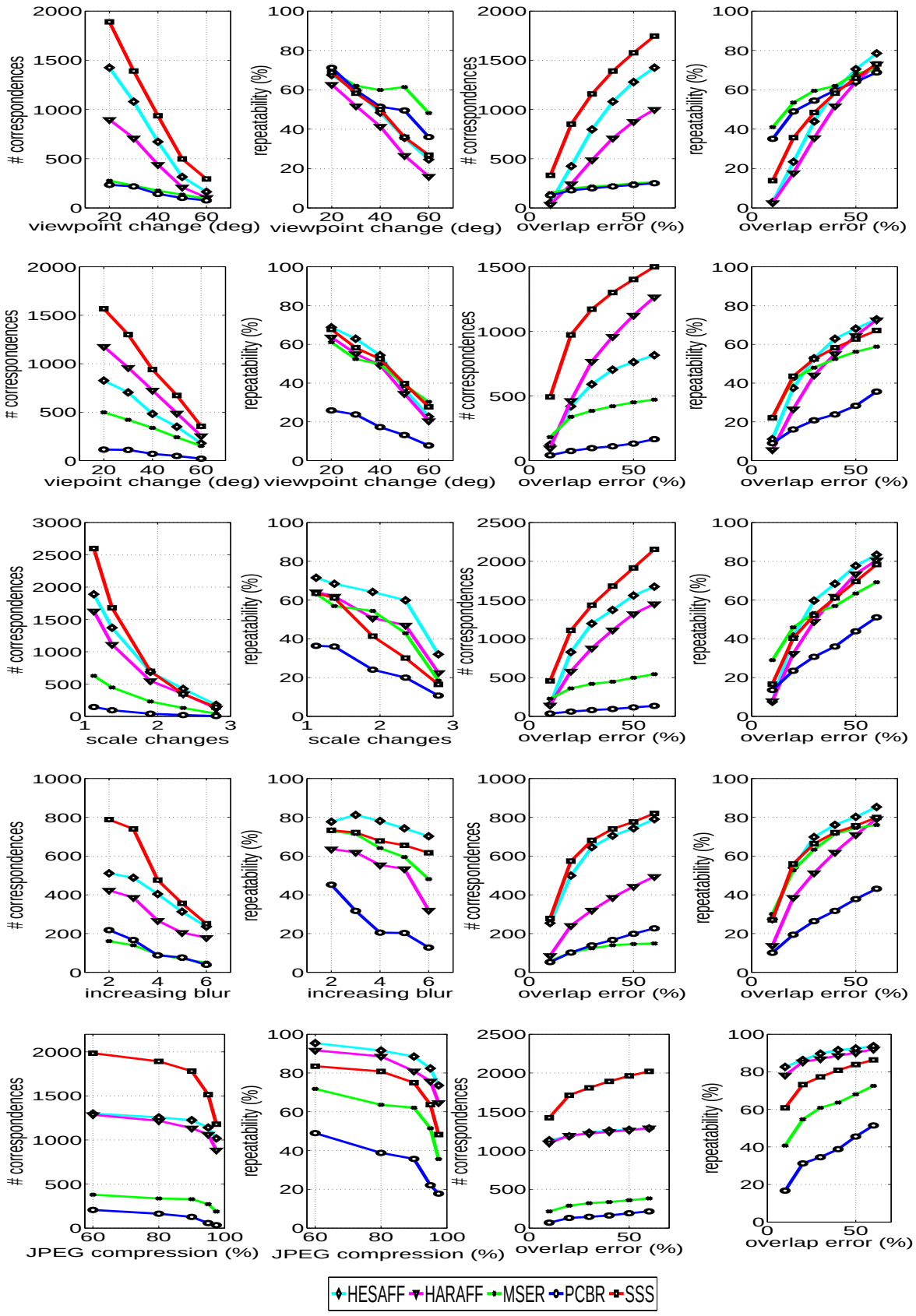

Fig. 9. Absolute and relative repeatability results. First row: Graffiti, second row: Wall, third row: Boat, fourth row: Bikes, fifth row: UBC. The first and second columns show the number of correspondences and the repeatability score, respectively, with an overlap error of 40%. The third and fourth columns depict the number of correspondences and the repeatability for the third image of each sequence as the overlap error varies in the range 10%–60%.

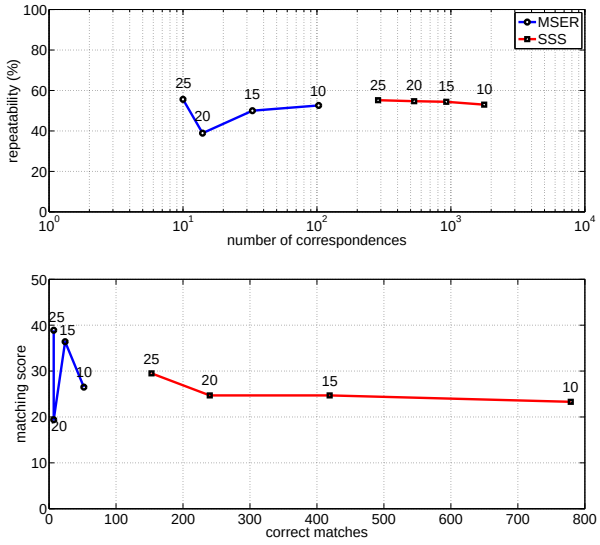

Fig. 10. Repeatability and matching results for the Bikes sequence (third image), with an overlap error of 20%. Top row: number of correspondences *vs.* repeatability score for different stability thresholds (10, 15, 20, 25), bottom row: number of matches *vs.* matching score for different stability thresholds (10, 15, 20, 25).

to equally assess the trade-off between the matching score and the number of matches using the SIFT descriptor [22]. Figure 10 shows these curves for the Bikes sequence, with an overlap error of 20%. The matching score is computed as the ratio between correctly matched regions and the number of corresponding regions. With the SIFT descriptor, MSERs and SSS tend to exhibit similar matching scores. However, as one expected, the latter provides a considerably higher number of correct matches.

V. Conclusions and Future Work

We have introduced a novel type of affine covariant features, Stable Salient Shapes (SSS), which are the result of performing a feature-driven MSER detection. The major motivation for our work came from the well-known advantages and the inherent drawbacks in obtaining affine covariant regions from extremal regions. The goal of the first step of the algorithm is to provide saliency maps that will be used as domains for MSER detection. These maps are characterised by the highlighting of features related to semantically meaningful structures, e.g., boundaries, and the simultaneous presence of smooth transitions at the boundaries. Our algorithm overcomes major limitations of a standard MSER detection, namely the sensitivity to image blur, the presence of a reduced number of regions, and the biased preference towards regular shapes. The experimental validation on a standard benchmark has shown that SSS are comparable to the most prominent affine covariant regions in terms of repeatability score. Concerning the absolute repeatability, our algorithm compares favourably to state-of-the art solutions. Moreover, our solution is efficient; it combines an already efficient MSER detection with a computationally inexpensive image filtering. Our detector was designed to preserve the most relevant image content, which makes it suitable to solve object recognition tasks. Thus, a future research direction is the evaluation of the performance of the SSS detector on object recognition problems.

References

[1] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[2] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European Conference on Computer Vision (ECCV'02)*, vol. I, 2002, pp. 128–142.

[3] ——, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[4] C. Harris and M. Stephens, "A combined corner and edge detector," in *4th ALVEY Vision Conference*, 1988, pp. 147–151.

[5] T. Lindeberg and J. Garding, "Shape-adapted Smoothing in Estimation of 3-d depth Cues from Affine Distortions of Local 2-d Structures," *Image and Vision Computing*, vol. 15, 1997.

[6] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *British Machine Vision Conference 2002 (BMVC'02)*, 2002, pp. 384–393.

[7] P. Moreels and P. Pietro, "Evaluation of Features Detectors and Descriptors based on 3D Objects," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[8] P.-E. Forssen and D. Lowe, "Shape Descriptors for Maximally Stable Extremal Regions," in *IEEE International Conference on Computer Vision (ICCV'07)*, Oct. 2007, pp. 1–8.

[9] P. Martins, C. Gatta, and P. Carvalho, "Feature-driven maximally stable extremal regions," in *7th International Conference on Computer Vision Theory and Applications (VISAPP'12)*, 2012, pp. 490–497.

[10] P.-E. Forssén, "Maximally stable colour regions for recognition and matching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, IEEE Computer Society. Minneapolis, USA: IEEE, June 2007.

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, vol. 65, no. 1/2, pp. 43–72, 2005.

[12] R. Kimmel, C. Zhang, A. Bronstein, and M. Bronstein, "Are MSER features really interesting?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2316–2320, 2011.

[13] M. Donoser and H. Bischof, "3d segmentation by maximally stable volumes (msvs)," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1, 0-0 2006, pp. 63–66.

[14] M. Perd'och, J. Matas, and S. Obdržálek, "Stable affine frames on isophotes," in *IEEE International Conference on Computer Vision (ICCV'07)*, 2007.

[15] M. Nielsen and M. Lillholm, "What do features tell about images?" in *Scale-Space and Morphology in Computer Vision*, ser. Lecture Notes in Computer Science, M. Kerckhove, Ed. Springer Berlin / Heidelberg, 2001, vol. 2106, pp. 39–50.

[16] I. Kokkinos and A. Yuille, "Scale invariance without scale selection," in *Proc. of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, June 2008, pp. 1–8.

[17] T. Dickscheid, F. Schindler, and W. Förstner, "Coding images with local features," *International Journal of Computer Vision*, vol. 94, no. 2, pp. 154–174, 2011.

[18] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[19] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro, "Principal curvature-based region detector for object recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, 2007.

[20] W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *IEEE International Conference on Computer Vision (ICCV'09)*, Kyoto, Japan, 2009.

[21] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[22] D. G. Lowe., "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.