FEATURE-DRIVEN MAXIMALLY STABLE EXTREMAL REGIONS

First Author Name[1], Second Author Name[1] and Third Author Name[2]
[1]*Institute of Problem Solving, XYZ University, My Street, MyTown, MyCountry*
[2]*Department of Computing, Main University, MySecondTown, MyCountry*
{*f_author, s_author*}*@ips.xyz.edu, t_author@dc.mu.edu*

Keywords: Local Feature Detection, Maximally Stable Extremal Regions, Completeness

Abstract: The high repeatability of Maximally Stable Extremal Regions (MSERs) on structured images along with their suitability to be combined with either photometric or shape descriptors to solve image matching problems have contributed to establish the MSER detector as one of the most prominent affine covariant detectors. However, the so-called affine covariance that characterizes MSERs relies on the assumption that objects possess smooth boundaries, a premiss that is not always valid. We introduce an alternative domain for MSER detection in which boundary-related features are highlighted and simultaneously delineated under smooth transitions. Detection results on common benchmarks show improvements that are discussed.

1 INTRODUCTION

Vision tasks such as object recognition, view matching, and tracking, just to name a few, are often performed under the use of local descriptors, which are usually characterized by an invariant response to certain classes of image transformations. A common procedure to obtain the invariant description consists in a preliminary detection of local image regions in a covariant way, which will provide local patches to be described in an invariant manner.

In the particular case of affine transformations, the vision community has been prolific in introducing covariant feature detectors as well as invariant descriptors. Detectors such as the Harris-Affine, Hessian-Affine (Mikolajczyk and Schmid, 2004) or the Maximally Stable Extremal Regions (MSERs) algorithm (Matas et al., 2002) are noteworthy solutions to detect affine covariant features. For a complete review and extensive references on this class of regions, we refer the reader to the work of Tuytelaars and Mikolajczyk (2008) and references within.

The almost linear complexity of the algorithm, the repeatable results and the suitability of MSER to be described by photometric or shape descriptors (Moreels and Pietro, 2007; Forssen and Lowe, 2007) have made the MSER detector one of the popular references in the literature. Here, we bring into the analysis the well-known and not so well-known shortcomings of the above mentioned detection. As a result, we suggest an alternative domain for the detection of MSERs which will allows us to retrieve a substantially higher number of features and improve the robustness to blur. Furthermore, the new domain provides a more reliable summary of relevant image content via MSERs detection.

2 BACKGROUND AND MOTIVATION

Before laying out a detailed description of the motivation behind our study, we will shortly review the basic concepts and terminology around MSER detection.

2.1 Maximally Stable Extremal Regions

Concisely, a MSER is defined as an affine covariant image region which is either brighter or darker than its immediate surroundings and without any particular shape. For a more formal definition, let us consider an *image* as a mapping $I : \mathcal{D} \rightarrow [0,1]$, where $D \subset \mathbb{R}^2$. Let us also denote the *interior* and the *boundary* of a connected component Q_c in $\mathcal{D}$ as ∂Q_c and $\mathrm{int}(Q_c)$, respectively. MSERs correspond to a very specific subset of connected components in the image domain. They are said to be *extremal* because the corresponding pixels have either higher or lower intensity than all the pixels on its boundary: $I\,|_{\mathrm{int}(Q_c)} > c$ or $I\,|_{\mathrm{int}(Q_c)} < c$ for a given $c \in [0,1]$, which is the value on the boundary. Thus, Q_c represents a *maximal connected component*

of a level set. Designating the region as *stable* refers to the fact that a change of $c = c + \Delta$, with $\Delta > 0$, will imply minor changes in the area of the extremal region. Q_c will be coined as *maximally stable* if the *stability measure*

$$\rho(Q_c) = \frac{|Q_c|}{\frac{\mathrm{d}}{\mathrm{d}c}|Q_c|} \quad (1)$$

attains a local maximum at c (Matas et al., 2002; Kimmel et al., 2011).

2.2 MSER detection: known shortcomings

Probably, the most important property of MSER is its affine covariance that can be directly inferred from the covariance of the image level sets with respect to affine transformations in the image domain. On the other hand, the affine invariance of the detector's stability measure guarantees the affine covariance of the regions. However, as addressed and shown by Kimmel et al. (2011), the aforementioned property prevails only if the boundaries of objects are smooth, which is not always the case. Additionally, the strategy adopted for the detection of these local features is based on the assumption that images possess well-defined structures (MSER features are usually anchored on region boundaries (Tuytelaars and Mikolajczyk, 2008)). These shortcomings have been reported in the well-known comparative study on affine covariant features performed by Mikolajczyk et al. (2005). MSERs – along with Hessian-Affine regions – have shown to be the most repeatable features in the overall benchmark. However, the former has evidenced an expected less consistent behavior: reasonably textured scenes or blurred sequences were prone to provide less repeatable results. As an example, Figure 1 illustrates such fragility when a progressive de-focus blur is present: as it increases, the shape of MSERs tends to change and, simultaneously, its number decreases.

Furthermore, the number of features that the MSER detector retrieves is usually low if we compare it to the number of local features detected by other prominent detectors, *e.g.*, Hessian-Affine or Harris-Affine. We can regard this fact as a shortcoming of MSER detection, as it might hinder the robustness of MSERs to occlusions.

A quite less trivial but noteworthy characteristic of MSER detection derives from the stability measure (Eq. (1)), which prefers round shapes to irregular ones. To our knowledge, Kimmel et al. (2011) have been the only ones to address the biased preference for regular shapes. In their paper, the authors have illustrated this bias by showing that if two regions have the same area and the same intensity along their boundaries, the measure ρ tends to prefer the one with a shorter boundary (perimeter). We emphasize the importance of such property as most scenes contain irregular shapes and we surely cannot claim that regular shapes are always more distinctive than irregular ones.

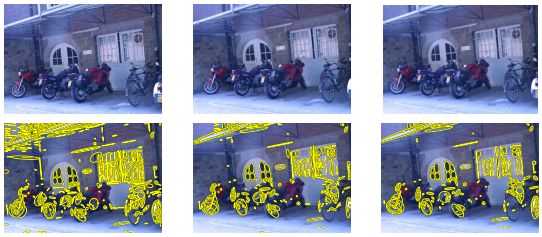


Figure 1: MSER detection (bottom row) on the first three images of the "Bikes" sequence (top row). MSERs are represented by fitted ellipses.

2.3 Related Work: other MSERs

We have highlighted some of the most important shortcomings of MSER detection, namely (i) a relatively high sensitivity to blurring, (ii) a small number of features, and (iii) the preference of the algorithm towards regular shapes.

A straightforward approach to reduce the sensitivity of MSER detection to blur is to apply a low-pass filter to the input image. Such pre-processing will enhance the repeatability of MSERs in the presence of blurred scenes, whereas the number of features will be reduced. A more reliable alternative is to make use of a multi-resolution detection: Forssen and Lowe (2007) suggest an improved version of the MSER detector by building a scale pyramid with one octave between scales. MSERs are detected at each resolution and duplicated features at consecutive scales are removed by comparing their locations. This strategy leads to a higher number of detected features and an improved repeatability under blur and scale changes. However, it requires MSER features to be detected at each resolution, which increases the computational complexity of the method.

With the aim of freeing the MSER detector from the existent preference towards round shapes, Kimmel et al. (2011) propose several reinterpretations of the stability measure in order to define more informative shape descriptors.

3 FEATURE-DRIVEN MSERS

We extend the concept of MSER to an alternative image representation in which certain boundary-related features are highlighted. MSERs detected on this domain will be coined as *feature-driven MSERS*. The proposed domain is less sensitive to blurring and allows the MSER detector to respond to a substantially higher number of features per image. Moreover, irregular shapes will appear more regular in the new domain. In short, we propose a domain that encapsulates the desired properties to yield a reliable MSER detection, namely the existence of regular shapes and well-defined boundaries combined with smooth transitions. The new domain is also designed to improve the *completeness* of MSER features. We can regard a set of features as a complete one if it preserves as much as possible the information contained in the image (Dickscheid et al., 2010). Increasing the number of detected MSERs is not enough to produce a more complete feature set; the regions should also preserve the most relevant image content.

In the process of obtaining a domain characterized by well-defined boundaries, but with smooth transitions, we start by highlighting boundary-related features. The process described herein highlights edges using gradient information, which is a valid source of information to build the domain with the desired properties. Other alternatives can be considered for feature highlighting, *e.g.*, the eigenvalues of the Hessian matrix.

For our purpose, obtaining smooth transitions at the boundaries is as important as the process of feature highlighting and the consequent identification of structural regions. 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. In the domain given as an example, we will 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 the result of convolving I with a Gaussian kernel g with variance σ^2. From the first order derivatives of L at different scales, we obtain a new input for MSER detection:

$$\hat{L}(\vec{x}) = \sum_{i=1}^{N} \sigma_i \sqrt{L_x^2(\vec{x},\sigma_i) + L_y^2(\vec{x},\sigma_i)}, \qquad (2)$$

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 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. Fig. 2 illustrates the suggested detection on a scene with de-focus blur. It is readily seen that the new set of MSERs provides a good coverage of the content and robustness to blur.

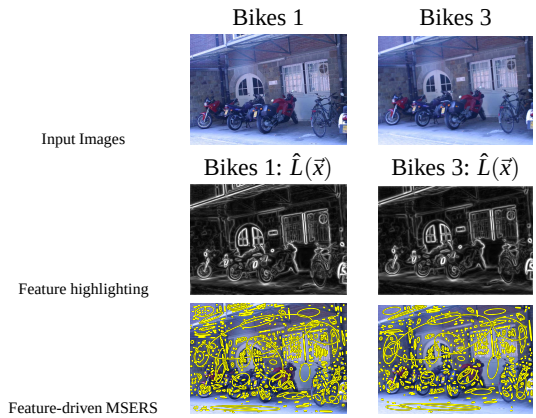
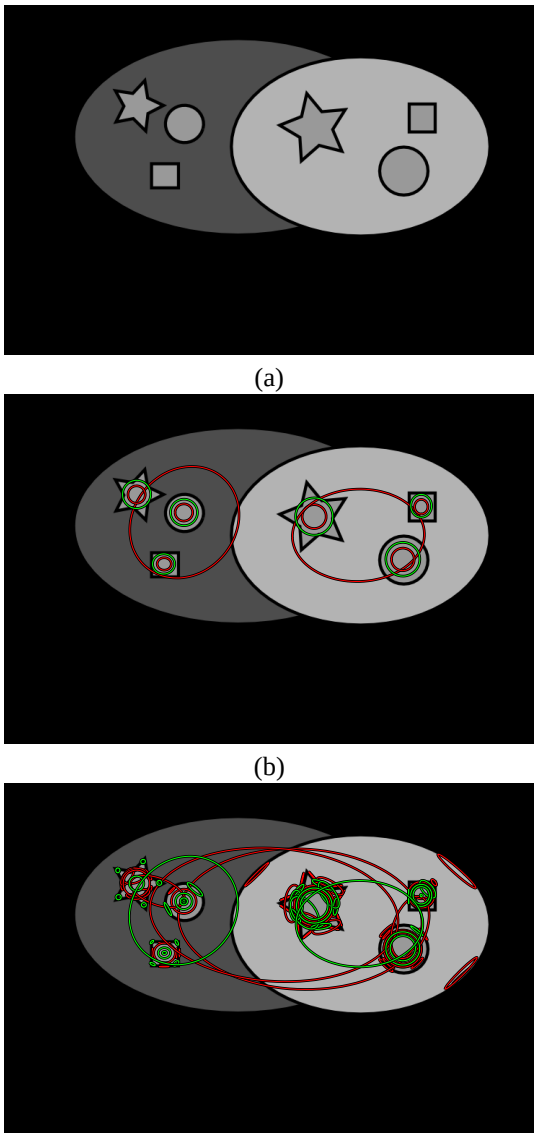


Figure 2: Example of a detection of feature-driven MSERs using the suggested domain. Detected features are represented by fitted ellipses.

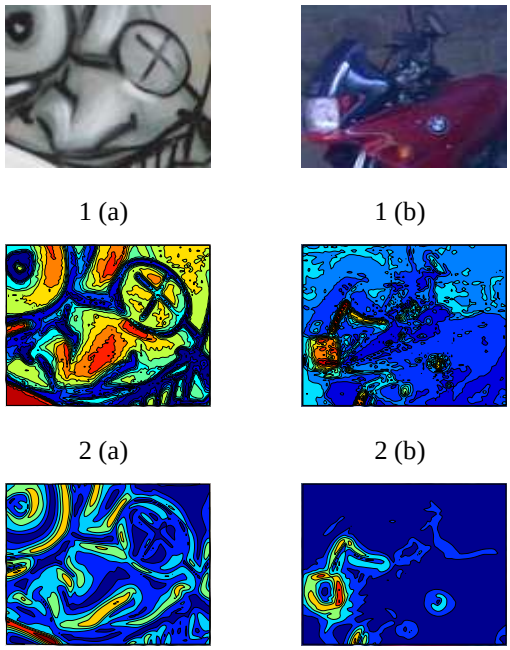
The detection depicted in Fig. 3 serves as a more illustrative example of the difference between original MSERs and the proposed ones. For a better comprehension of the proposed detection and due to the relatively low number of features detected on the synthetic image, we distinguish in this example MSER+ features *i.e.*, dark regions with brighter boundaries, from MSER- features, which correspond to bright regions with darker boundaries. The latter features are obtained by giving the inverted image as input. As can be seen, the use of the proposed domain yields a higher number of MSERs. Furthermore, it allows us to capture informative patterns in the scene such as corners and junctions, which are usually neglected by a traditional MSER detection.

The new domain will also attenuate the so-called irregularity of extremal regions. Figure 4 helps to clearly illustrate the process of attenuation: the boundaries of extremal regions correspond to *isophotes*, *i.e.*, iso-intensity countours. As expected, these curves tend to appear irregular in shape on the luminance channel (see Fig. 4-2(a)-(b)), whereas the new domain provides more regular shapes (see Fig. 4-3 (a)-(b)). The examples in this figure equally demonstrate that the proposed domain tends to highlight relevant structures and present them as potential extremal regions. We can illustrate this fact with the example of the logo on the motorbike: it appears as a scattered object on the luminance channel, whereas the new domain regards it as a relevant structure.



(a) (b) (c)

Figure 3: MSER detection on a synthetic image (green: MSER+, red: MSER-): (a) input image; (b) original detection; (c) suggested detection.



1 (a) 1 (b) 2 (a) 2 (b) 3 (a) 3 (b)

Figure 4: Isophotes and extremal regions: (1) input images; (2) regions delineated by isophotes on the luminance channel; (3) regions delineated by isophotes on the suggested domain.

4 EXPERIMENTAL VALIDATION

We have followed the guidelines given by the benchmark suggested by Mikolajczyk et al. (2005) to assess the repeatability of the redefined MSER on planar scenes. In the literature, repeatability as been regarded as the fundamental criterion to assess the performance of local features. Regardless of its importance, it does not provide a hint on the usefulness and quality of the features. Therefore, our evaluation has taken into account another important criterion that is often neglected: the completeness of features.

Furthermore, we would like to determine the level of complementarity between MSERs extracted from the original image and the proposed domain. Special emphasis is given to the performance of MSERs on structured images.

In this section, features that are retrieved using the gradient-based image $\hat{L}$ as the input image for detection will be referred to as f-MSERs.

Implementation details We have made use of the code provided by Vedaldi and Fulkerson (2008) for MSER detection, which has been extended to deal with images intensity values varying in a range different from $\{0,\ldots,255\}$, as the new domain intensity values might be greater than 255.

Parameter settings In the whole set of experiments, the new domain was built with an initial scale $\sigma_0^2 = 0.64$ and $\xi = 1.19$. N, the number of scales, was set to 16. The default value for the variation parameter Δ required by the stability measure is 10, as it is a reasonably low number to give rise to a high number of features. The protocol that we have followed evaluates the repeatability rate and the number of correspondences. A less sparse detection often implies a

drop in the repeatability rate, albeit the potential increase in the number of correspondences. For a reliable evaluation of these two criteria, detectors should respond to a comparable number of features. Thus, in some cases, we report results using additional values for the variation parameter in the stability measure. On one hand, the use of several Δ allows us to have a more insightful analysis and comparison of the main results. On the other hand, it is also advantageous to analyze results for a single (fixed) Δ: we can expect a higher number of f-MSER features but would they be more repeatable than MSERs? Moreover, it is interesting to assess the degree of complementarity between MSER and f-MSER features when there is a discrepancy between the cardinalities of sets.

4.1 Repeatability Evaluation

The benchmark suggested by Mikolajczyk et al. (2005) computes the repeatability between regions detected on two image pairs using 2D homographies as a ground truth. Two regions are deemed as corresponding and, therefore, repeated, with an overlap error of $\varepsilon_0 \times 100\%$ if $1 - \frac{\left|\mathcal{R}_{\mu_1} \cap \mathcal{R}_{(H^T \mu_2 H)}\right|}{\left|\mathcal{R}_{\mu_1} \cup \mathcal{R}_{(H^T \mu_2 H)}\right|} < \varepsilon_0$, where μ_1 and μ_2 denote the approximating ellipses of the two detected regions, H is the homography and $\mathcal{R}_*$ represents the set of points identified by its subscript $*$. The dataset, known as the Oxford dataset, comprises 8 sequences of images, each one with 6 images, whose short description is outlined in Table 1 and depicted in Fig. 5. One of the main motivations to define f-MSERs is to find a more repeatable set of features in structured scenes, even when blur occurs. Figures 7 to 10 depict the repeatability and the correspondence plots for the sequences "Bikes", "UBC", "Leuven" and "Boat" (see detection results in Fig. 6), with an overlap error of 40%. The first image in each sequence is used as the reference. In all the cases, it is clear that f-MSER features are higher in number without showing a significant drop in the repeatability rate. We would like to stress that a substantially higher number of correspondences accompanied with a slight decrease of the repeatability rate is preferable to a minor increase of the repeatability with less features, as in the former the absolute number of repeated features is considerably higher, which might provide a better coverage of the content with a similar repeatability rate. Nonetheless, the sequences of structured images affected and corrupted by blur or JPEG compression have shown an improved repeatability rate with the use of f-MSERs. In the "Leuven" sequence, we have a similar repeatability rate for both types of MSERs, although f-MSERs appear in a higher number. The changes in the lighting conditions produce a similar effect on the detection of MSERs on both domains. For the "Boat" sequence, we report a decrease of the repeatability rate of f-MSERs at higher scales. Nevertheless, the number of repeated f-MSERs at higher scale changes is comparable to the number of repeated MSERs.

For space reasons, we will not display all the results of the experiments in the form of plots. However, Table 2 gives a hint on the behavior of the detected regions for the complete dataset. It provides the repeatability rate and number of correspondences for the third image of each sequence as well as the corresponding average values for each sequence with an overlap error of 40%. By analyzing the table, one can conclude that feature-driven MSERs yield a considerably higher number of correspondences. The lower average repeatability rate of f-MSERs as seen in some sequences is mainly due to the decrease of the repeatability rate in the last images of the sequence. As theory confirms, in well-structured scenes, the new MSERs can yield an increased repeatability. With regard to sequences with viewpoint changes, f-MSERs have shown a satisfactory repeatability. Nonetheless, this repeatability can be improved if we build the domain under an affine scale-space.

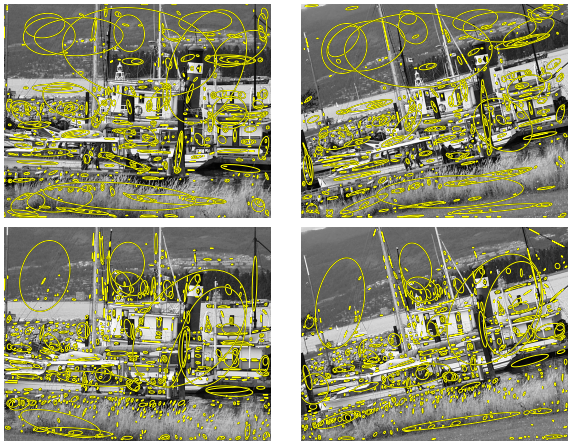


Figure 6: f-MSERs (top row) and MSERs (bottom row) detected on the first and third images of the "Boat" sequence.

4.1.1 Additional experiments

We have performed additional experiments on images from the Caltech Image Categories dataset. The test images have been blurred (Figs. 11 and 12) and MSERs and f-MSERs have been detected on them. The repeatability results are summarized in Figs. 13 and 14. Once again, f-MSERs are retrieved in a higher number. When the effect of blurring is more severe,

Table 1: Oxford dataset sequences.

Sequence	Graffiti	Wall	Boat	Bark	Bikes	Trees	Leuven	UBC
Texture		✓		✓		✓		
Transformation	Viewpoint	Viewpoint	Scale	Scale	Blur	Blur	Light	JPEG

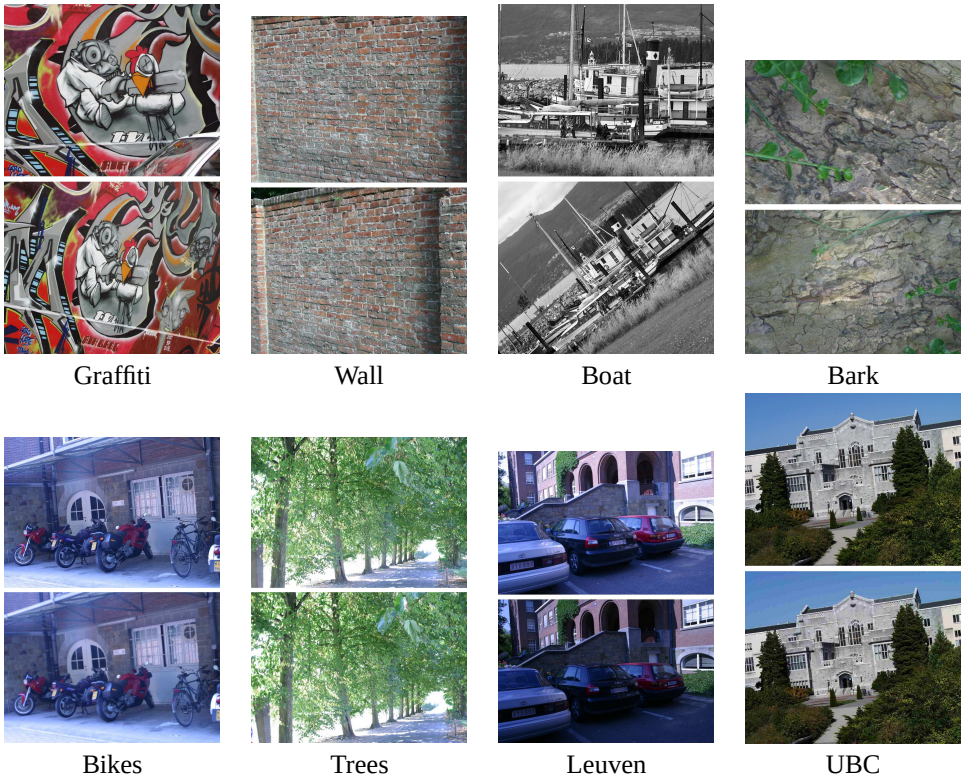


Figure 5: Oxford dataset sequences. Top row: reference images; bottom rows: third images.

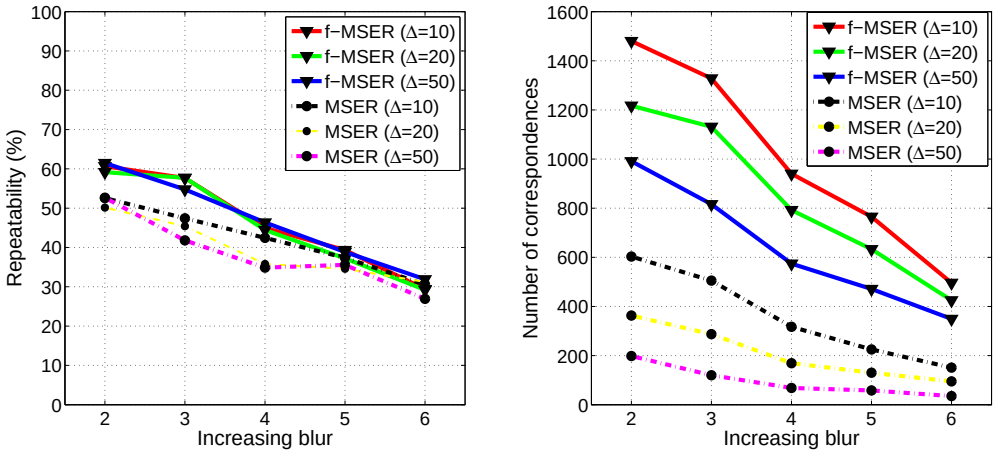


Figure 7: Repeatability rate and number of correspondences for the “Bikes” sequence (overlap error of 40%).

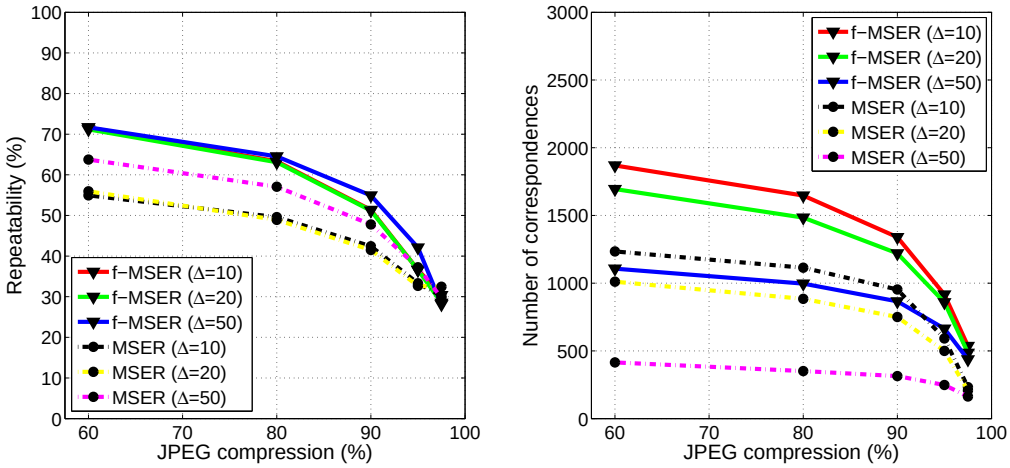


Figure 8: Repeatability rate and number of correspondences for the "UBC" sequence (overlap error of 40%).

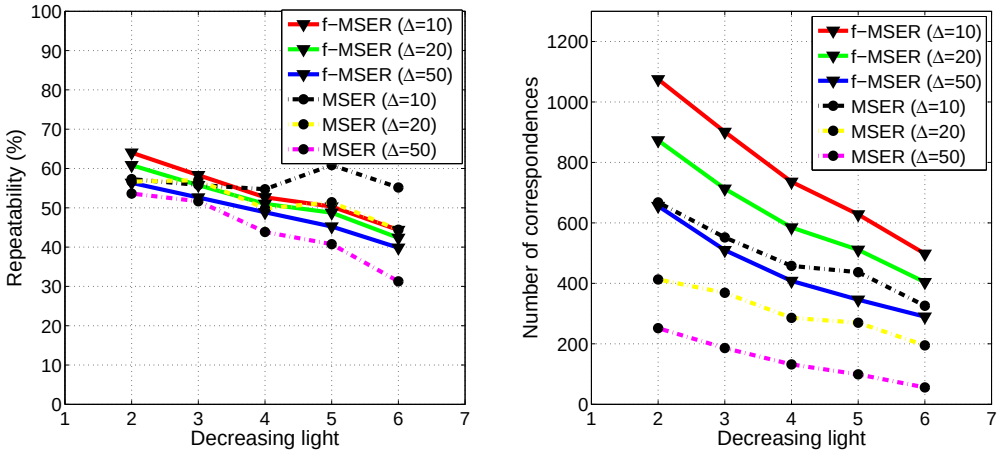


Figure 9: Repeatability rate and number of correspondences for the "Leuven" sequence (overlap error of 40%).

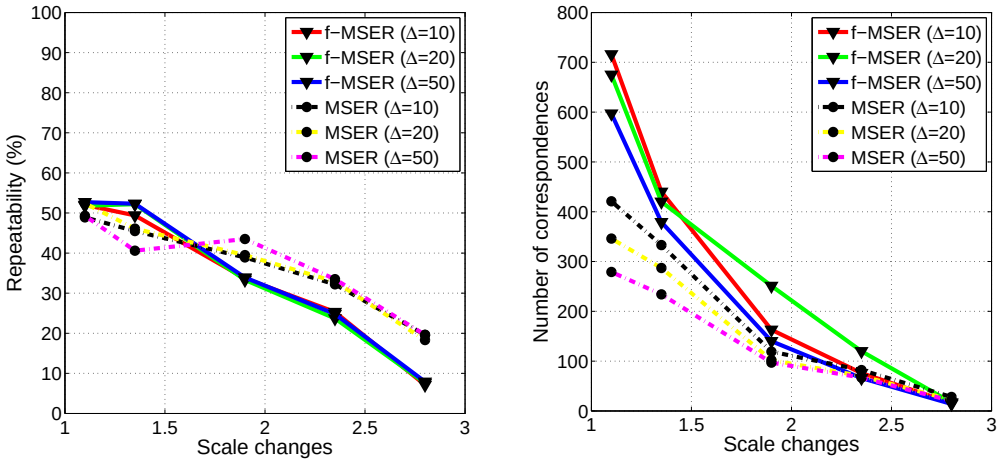
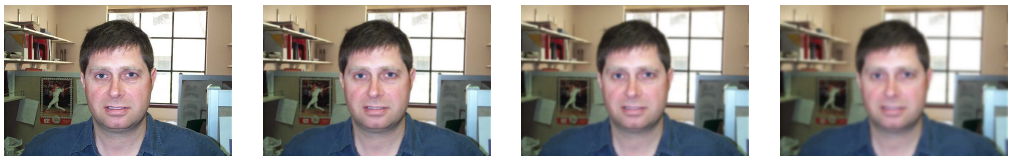


Figure 10: Repeatability rate and number of correspondences for the "Boat" sequence (overlap error of 40%).



(a) (b) (c) (d)

Figure 11: "Faces" sequence (Gaussian blur): (a) original; (b) $\sigma = 5$; (c) $\sigma = 10$; (d) $\sigma = 15$.

Table 2: Repeatability rate and number of correspondences for each sequence (overlap error of 40%).

	3rd image		Sequence (average)	
	MSER	f-MSER	MSER	f-MSER
Grafitti	**56%**(310)	48%(**538**)	**48%**(244)	39%(**402**)
Wall	53%(2093)	**57%**(**2484**)	46% (1595)	46% (**1786**)
Boat	48%(658)	**56%**(**861**)	**41%**(388)	39%(**556**)
Bark	**52%**(276)	48%(**355**)	**60%** (268)	39%(**295**)
Bikes	47%(505)	**58%**(**1328**)	42%(360)	**46%**(**1002**)
Trees	**38%** (1767)	36% (**1796**)	**33%**(1295)	30%(**1401**)
Leuven	57%(668)	**58%**(**901**)	**57%** (488)	54%(**768**)
UBC	50%(1114)	**63%**(**1647**)	42% (825)	**51%**(**1262**)

MSERs tend to be more repeatable but with a noticeably low number of correspondences.

4.2 Completeness Evaluation

How much of the image information is preserved by local features? Feature sets should capture the most relevant image content in order to provide a trustworthy summarized description of the image. This requirement is known in the literature as completeness. A measure of completeness has been proposed by Dickscheid et al. (2010). Succinctly, the *incompleteness* of a detection corresponds to the following *Hellinger distance*: $d(p_H, p_c) = \sqrt{\frac{1}{2}\sum_{\vec{x}\in\mathcal{D}}(\sqrt{p_H(\vec{x})} - \sqrt{p_c(\vec{x})})^2}$, where p_H corresponds to an *entropy density*, computed from local image statistics and p_c denotes a *feature coding density*, inferred from the set of features. We emphasize that the completeness measure is not a simple way of evaluating the image coverage of feature sets; it penalizes local feature sets that contain less informative patterns despite of their coverage. For the purpose of the evaluation, we defined a dataset comprising images from the Oxford dataset sequences. Each sequence was represented by its third image. The parameter settings for the detection coincide with the ones used in the previous subsection. Table 3 outlines the results of the completeness evaluation given by the Hellinger distance between the two aforementioned distances, using feature coding densities computed from MSER and f-MSER feature sets as well as a combination of both sets. The analysis of the latter feature set results allows us to infer on the complementarity of MSER and f-MSER features. Additionally, we have analyzed the completeness of f-MSER+ and f-MSER- features. These results are summarized in the same table.

It is important to note that MSER-like features are not among the most complete ones (Dickscheid et al., 2010). It is readily seen that homogeneous regions are easily entitled to be classified as extremal regions, which, in most of the cases, means excluding the most informative content. Furthermore, the number of features detected by the MSER algorithm tends to be lower than the ones given by other prominent detectors, such as the Hessian-Affine or the Harris-Affine.

One important conclusion to be drawn from Table 3 is that f-MSER detection will provide us a more complete set of features than the one comprised of standard MSERs. The incompleteness values for f-MSERs range from 0.1542 ("Wall" image) to 0.3573 ("Leuven" image), which reflects the high level of completeness of these features. MSER features are less complete and, in some cases, even less complete than f-MSER+ or f-MSER- feature sets. The combination of MSER and f-MSERs features gives us the most complete feature sets for each sequence. However, the complementary between both feature sets is practically non-existent; the incompleteness values for MSER $\cup$ f-MSER features are comparable to the ones for feature-driven MSERs sets. This result helps us to conclude that the relevant content preserved by standard MSERs is also preserved by f-MSERs.

5 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

We have addressed the shortcomings of MSER detection as well as the desired properties of an image in order to provide a reliable MSER detection. As result, we have introduced an alternative domain for MSER detection. This domain is mainly characterized by the highlighting of certain boundary-related features and the simultaneous presence of smooth transitions at the boundaries. The detection of MSERs on this domain

(a) (b) (c) (d)

Figure 12: "Cars" sequence (Gaussian blur): (a) original; (b) $\sigma = 5$; (c) $\sigma = 10$; (d) $\sigma = 15$.

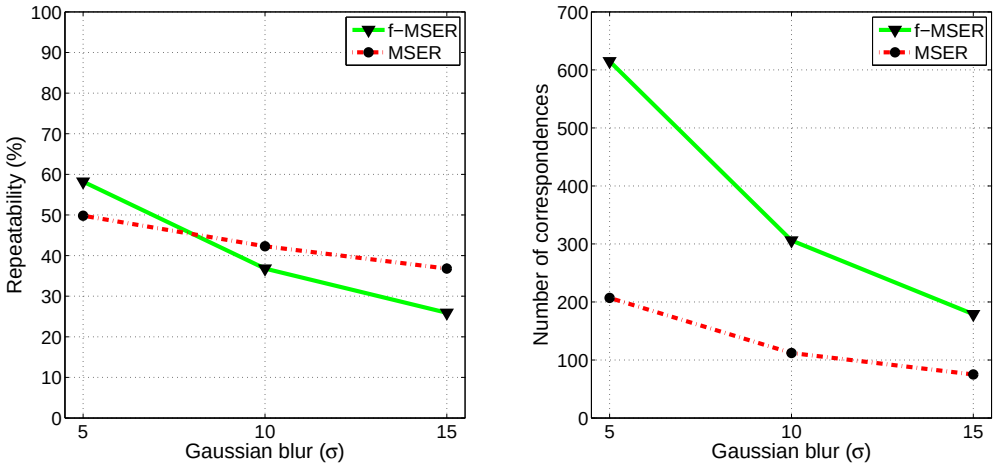


Figure 13: Repeatability rate and number of correspondences for the "Faces" sequence (overlap error of 40%).

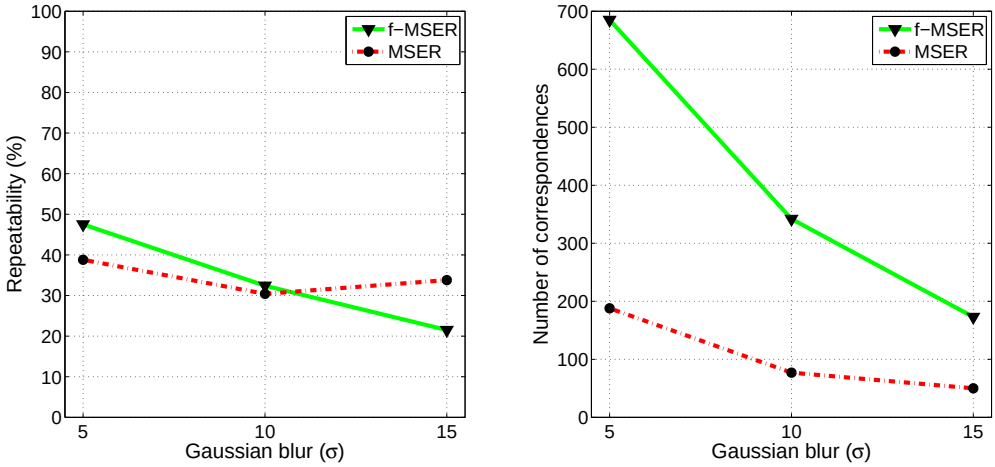


Figure 14: Repeatability rate and number of correspondences for the "Cars" sequence (overlap error of 40%).

Table 3: Dissimilarity measure $d(p_H, p_c)$ and number of regions for the third image of each sequence.

	MSER	f-MSER	f-MSER+	f-MSER-	MSER ∪ f-MSER
Graffiti	0.3828 (1085)	**0.2653** (2216)	0.3317 (1130)	0.3565(1086)	0.2630
Wall	0.2060 (4433)	**0.1542**(4981)	0.2402 (2711)	0.2568(2270)	0.1361
Boat	0.3532 (2319)	**0.2782** (2720)	0.3634 (1393)	0.3474 (1327)	0.2694
Bark	0.2862 (1940)	**0.1921** (2802)	0.2729 (1553)	0.31 (1249)	0.1817
Bikes	0.4827(1123)	**0.3170** (2388)	0.4042 (1311)	0.4318 (1077)	0.3204
Trees	0.2359 (4821)	**0.1874** (5299)	0.2871 (2529)	0.2573 (2770)	0.1755
Leuven	0.4539 (1017)	**0.3573**(1597)	0.4386 (869)	0.4557 (728)	0.3451
UBC	0.2929 (2431)	**0.2636** (2619)	0.3479 (1400)	0.3523(1219)	0.2378

responds to a higher number of regions than the one that uses the luminance channel as input. The feature-driven MSERs tend to be more robust to blur than traditional ones. Furthermore, the new set of features is more complete as it preserves more of the relevant image content. The high level of completeness shown by the feature-driven MSERs suggests that they can be used to solve object recognition problems.

The new domain can be built with any function that responds to boundary-related features, which gives several options to work with. In this paper, we have used the gradient magnitude to build it and, as future research directions, we intend to exploit other measures to derive the domain. Future work also includes assessing the performance of feature-driven MSERs on object recognition problems.

REFERENCES

Dickscheid, T., Schindler, F., and Förstner, W. (2010). Coding images with local features. *International Journal of Computer Vision (in press)*.

Forssen, P.-E. and Lowe, D. (2007). Shape Descriptors for Maximally Stable Extremal Regions. In *Proc. of IEEE 11th Int. Conf. on Computer Vision*, pages 1–8.

Kimmel, R., Zhang, C., Bronstein, A., and Bronstein, M. (2011). Are MSER features really interesting? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(11):2316–2320.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *Proc. of the British Machine Vision Conference 2002 (BMVC 2002)*, pages 384–393.

Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2005). A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1/2):43–72.

Moreels, P. and Pietro, P. (2007). Evaluation of Features Detectors and Descriptors based on 3D Objects. *Int. J. Comput. Vision*, 73(3):263–284.

Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Found. Trends Comput. Graph. Vis.*, 3(3):177–280.

Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`.