DATA-DRIVEN REGION DETECTOR FOR STRUCTURED IMAGE SCENES

Elena Ranguelova

Netherlands eScience Center
Amsterdam, The Netherlands

ABSTRACT

A Data-driven Morphology Salient Regions (DMRS) detector, related to our MSSR detector is proposed. It demonstrates comparable repeatability to the best-known MSER detector on standard structured scenes and better resolution invariance on a high-resolution benchmark. This is achieved via significantly smaller number of detected regions- a much desired property in the big data era. A data-driven binarization algorithm gives compact image representation, subsequently analyzed for saliency using morphology. Also, a new dataset, 'OxFrei', for transformation-independent detection evaluation is introduced. While MSER is an excellent detector for generic applications, the DMSR is geared towards the analysis of scientific imagery (e.g. in emerging domains such as animal and plant biometrics) for detecting precisely semantically meaningful regions. In this paper, DMSR is demonstrated to better detect identifying structures in marine animals and wood microscopy images.

***Index Terms*—** region detection, data-driven, morphology, structured scenes, scientific visual analytics

1. INTRODUCTION

Finding reliably and repeatedly correspondences between two images of the same object or scene, taken under different viewpoints and acquisition conditions, is the first fundamental step in numerous computer vision applications: wide baseline stereo matching, image retrieval, model-based recognition, visual mining, object categorization, etc. An important class of features are the distinct (salient) regions, which correspond to the same image patches, detected independently for each viewpoint. The region detectors must be *covariant* (often called *invariant*) to a class of transformations, usually *affinity* and various photometric distortions.

While the research has been focused on the generic applications, the emerging fields of *animal and plant biometrics*, is attracting more attention of the community [1, 2]. It becomes clear that computer vision is the vital technology enabling the wild-life preservation efforts of ecologists in the big data era. An important question for these scientists, along with the individual or species photo-identification, is to obtain automatic reliable measurements of semantically meaningful regions, extracted from usually highly structured images. The generic region detectors often do not satisfy this need. In addition, there is a shortage of publicly available benchmarks for evaluating the robustness to transformations independently of the image content.

In this paper, we tackle the identified problems with an affine-covariant regions detector for structured images - *Data-driven Morphology Salient Regions (DMRS)* and introducing a new dataset for performance evaluation.

1.1. Related work

A decade ago, a performance evaluation paper by the Visual Geometry Group in Oxford, compared the existing affine-covariant region detectors, [3]. A reference *Oxford dataset*, which has become the standard evaluation benchmark since, has been introduced. A clear conclusion of the comparison was that the *Maximally Stable Extremal Regions (MSER)* is the best performing detector for *structured* scenes, [4]. The MSER has become the de-facto standard in the field, for example as part of the MATLAB Computer Vision Systems Toolbox. Despite its success, the MSER detector has several drawbacks: sensitivity to blur; produces many nested regions (not aligned with the perceptual saliency); the number of image regions is often large (undesirable for the subsequent matching in large datasets) and the performance degrades with up to 25% with the image resolution increase, [5]. Analysis of the MSER features in the geometric scale-space showed that the original formulation of the stability criterion makes MSER biased towards regular shapes, [6].

Many researchers have proposed improvements to MSER, although none of them increased the performance drastically. An MSER color extension, *Maximally Stable Color Region* is proposed in [7]. It outperforms a simple MSER per color channel combination and a color blob detector on the Oxford dataset. Improving the MSER region distinctiveness by morphological dilation operator on the detected Canny edges is proposed in [8]. The improved detector shows better performance than MSER for image classification in a bag of words framework, but the benchmark evaluation of repeatability is not reported. The MSER has been also extended to *Maximally Stable Volumes* and used to successfully segment 3D medical images and paper fiber networks, [9].

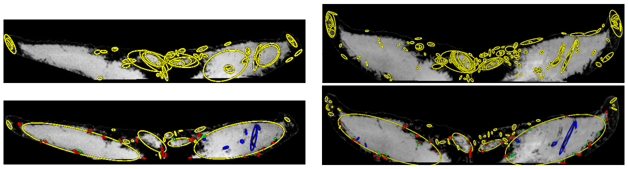
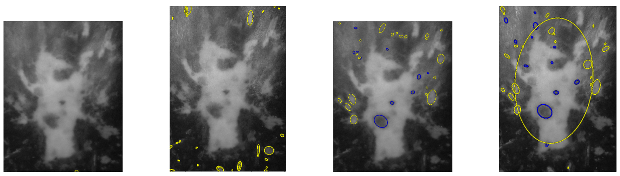


Fig. 1. Region detectors on two images of the tail of the same humpback whale. First row: MSER, second row: DMSRA.



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

Fig. 2. Region detectors on two images of the pineal spot of the same leatherback turtle. (a),(b): MSER, (c),(d): DMSR.

In the context of humpback whale identification, we have developed the *Morphology based Stable Salient Regions (MSSR)* detector, [10, 11]. DMRS gives even smaller number of regions (important for efficiency in subsequent matching) than MSSR. At the same time it pertains the perceptually salient property, in contrast to the often redundant concentric MSER regions, see Fig.1. Figure 2 (a) and (b) illustrates that MSER does not even detect useful regions on the leatherback images, while DMSR detects meaningful regions repeatedly.

Evaluation benchmarks are crucial for the development of region detectors. The standard Oxford dataset is very small for nowadays standards: 48 low resolution images (6 test sequences) with known homographies between the independent photos and evaluation protocol, [3]. The *Freiburg dataset* contains 416 higher resolution images, which unlike the Oxford set, have been generated by transforming 16 base images in order to de-tangle transformations from the image contents, [12]. Its main drawback is the lack of complete documentation. The *TNT dataset* contains versions of test sequences with increasing resolution from 1.5 MB to 8 MB, along with highly accurate homographies. It suitable for evaluation robustness to resolution rather than transformations [5].

1.2. Contributions

In this paper a robust to lighting and blur binarization algorithm is proposed. It is the basis for the proposed salient regions detector, the DMSR. The detector is faster than MSSR, while the regions are also perceptually salient. DMRS produces much smaller and more stable number of regions with comparable or higher (lighting, blur and increased resolution) repeatability to MSER. Its potential for enhancing scientific imagery analytics is illustrated. DMSR is available as open source software, [**?**]. Also, a new 'OxFrei' dataset combining features of the Oxford and Freiburg datasets is introduced.

Fig. 3. Binary salient regions detection. Color coding: holes- blue, islands- yellow, indentations - green, protrusions- red.

ISS	A CC $S^i_{fb} = \{\mathbf{p} \in \mathcal{D}, \forall \mathbf{p} = f,$ $\forall \mathbf{q} \in \partial S^i_{fb}, \mathbf{q} = b, \mathbf{q} \notin \partial \mathbf{B}\ \},$
2 types	S^i_{10} (islands), S^i_{01} (holes); $\mathbf{S}^i = S^i_{01} \cup S^i_{10}$
BSS	$S^b_{fb} : \{\mathbf{p} \in S^b_{fb} \subset \mathcal{B}^f, \forall \mathbf{p} = f,$ $\mathbf{q} \in \partial S^b_{fb} \subset \partial \mathcal{B}^f, \forall \mathbf{q} = b\},$ $\lvert\partial \mathcal{B}^f\rvert - \lvert\partial(\mathcal{B}^f \backslash S^b_{fb})\rvert < 2\pi r$
2 types	S^b_{10} (protr.), S^b_{01} (indent.); $\mathbf{S}^b = S^b_{01} \cup S^b_{10}$
Regions	$\mathbf{S} = \mathbf{S}^i$ (DMSR), $\mathbf{S} = \mathbf{S}^i \cup \mathbf{S}^b$ (DMSRA)

Table 1. Binary saliency definitions used in Section 2.1.

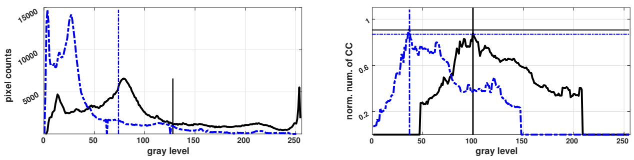
2. DATA-DRIVEN MORPHOLOGY SALIENT REGIONS

While MSER and MSSR process each gray-level of the image searching for stable salient regions, DMRS transforms the gray-scale saliency into binary saliency problem.

2.1. Binary Salient Regions Detection

The claim is that the perceptual saliency in a binary image of a structured scene $\mathbf{B} : \mathcal{D} \subset \mathcal{Z}^2 \rightarrow \{0, 1\}$ (1-white, 0-black) is only due to the spatial layout of the image regions. There are 4 possible types of salient regions. The 2 types of *inner salient structures (ISS)* are called *holes* - set of connected black pixels entirely surrounded by white pixels and the dual *islands*- set of connected white pixels surrounded by black ones. A significant connected component (CC) $\mathcal{B}^1$ is defined as a CC with area proportional to the area of the image by Λ. The radius of the morphological structuring element is r and the area opening parameter for filtering of noise-like regions is λ. The 2 *boundary salient structures (BSS)* are the *protrusions*- set of white pixels on the border of a significant CC, which if pinched off from the CC, its boundary will increase with no more than $2\pi r$ and the dually defined *indentations*.

The definitions are valid also for the MSSR detector. The regions are obtained from $\mathbf{B}$ by morphology operations: hole filling, top hat and area opening, for more details see [10, 11]. The ISS are similar to definition of the MSER+ and MSER- regions, [4]. In this paper, detectors using only ISS (e.g. directly comparable to MSER) are denoted by DMSR/MSSR, while DMSRA/MSSRA are detectors using all region types. These definitions are summarized in Table 1 and the exact shape regions are illustrated on a synthetic 100×100 binary image with parameters $\Lambda = 100$, $r = 5$ and $\lambda = 10$ on Fig.3.



(a) Otsu (b) Max number of CC

Fig. 4. Finding the optimal threshold for 2 images from the 'Leuven' sequence (Oxford dataset, lighting): the base image - solid black line, the 4th image - dotted blue line.

(a) (b) (c)

Fig. 5. Binarization of 2 images of the 'Leuven' sequence (lighting). First row- base image, second row- 4th image; (a) gray scale; (b) Otsu binarization, (c) proposed binarization.

2.2. Data-driven binarization

The MSER and MSSR detectors decompose a gray-scale image into binary cross-sections and evaluate the stability of the resulting CCs or accumulate evidence into saliency masks, respectively. In contrast, the proposed DMSR starts with a data-driven binarization producing single binary image which can also be used for image compression.

Any gray-scale image $\mathbf{I} : \mathcal{D} \subset \mathcal{Z}^2 \rightarrow \mathcal{T}$ ($\mathcal{T} = \{0, 1, ..., t_{max}\}, t_{max} = 2^n - 1$ is the maximum gray value encoded by n bits; usually $n = 8$, $t_{max} = 256$) can be decomposed into cross-sections at every possible level t: $\mathbf{I} = \sum_{t \in \mathcal{T}} CS_t(\mathbf{I})$, $(CS_0(\mathbf{I}) = \mathbf{I})$. Obtaining a cross-section at level t is equivalent to thresholding the image at threshold t (setting all pixels with values below t to 0 and the ones above-to 1), $CS_t(\mathbf{I})$ is a binary image. Three groups of CC are defined: $\mathcal{A}_t$- *all* CC in $CS_t(\mathbf{I})$, $\mathcal{L}_t$- the *large* CC in $CS_t(\mathbf{I})$ and $\mathcal{V}_t$- the *very large* CC in $CS_t(\mathbf{I})$. The size of each CC is categorized using $\Lambda_{\mathcal{L}}$ and $\Lambda_{\mathcal{V}}$ fraction of the image area $A_{\mathbf{I}}$. Lets denote the normalized number of elements in a set by $\|\cdot\| = |\cdot| / \max_{t \in \mathcal{T}} |\cdot|$ Finding the optimal threshold t_{opt} is then defined as $t_{opt} = \arg\max_{t \in \mathcal{T}} (w^{\mathcal{A}} \|\mathcal{A}_t\| + w^{\mathcal{L}} \|\mathcal{L}_t\| + w^{\mathcal{V}} \|\mathcal{V}_t\|)$, where $w^{\cdot}$ are the weights per set.

The standard gray-level histogram and Otsu threshold does not provide criterion for selecting a single CS with stable salient regions, while choosing the threshold t_{opt} ensures stable number of regions across image transformations. Figures 4 and 5 illustrate the invariance to lighting.

Fig. 6. Salient region detectors on the base image of 'Graffiti' sequence, Oxford dataset. Left: MSER, right: DMSR

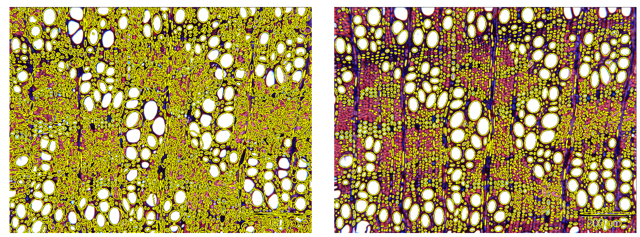


Fig. 7. Salient region detectors on microscopy wood images. Left: MSER (every second region is shown), right: DMSR

After the data-driven binarization, the DMSR detector finds the set of affine-covariant regions $\mathbf{S}$ from the single binary image $CS_{t_{opt}}$ as described in section 2.1 and [10, 11]. DMSR produces smaller number of non-overlapping and perceptually salient regions (visualized by their equivalent ellipses, not exact shapes) compared to MSER, Figs. 6 and 7.

3. PERFORMANCE EVALUATION

The *repeatability score* (R) and the *number of correspondences* (N_C) are the main performance evaluation measures, [3]. The maximum overlap error between matching regions is 40%. The R score between a pair of images $< base, transformed >$ ($< \mathbf{I}_B, \mathbf{I}_T >$) is the ratio between N_C in the common image part and the smaller number of regions in the pair. Only the structured scenes from each dataset are considered and 5 detectors are evaluated: MSER, MSSR(A) and DMSR(A). The MSSR(A)/DMSR(A) parameters used are: $r = 0.02 * \sqrt{A_{\mathbf{I}}/\pi}$, $\lambda = 3r$,$\Lambda_{\mathcal{L}} = 0.01$, $\Lambda_{\mathcal{V}} = 0.1$ and $w^{\cdot} = 0.33$. The original MSER software is used with default settings.

3.1. Oxford dataset

Each image sequence of the Oxford dataset consists of a 1 base and 5 increasingly distorted images, [3]. They are obtained independently of each other and the homographies between each pair $< \mathbf{I}_B, \mathbf{I}_T >$ are the provided ground truth. Each sequence can be used to test only 1 image transformation. The following sequences have been used: 'Boat'-scale (rotation and zoom), 'Bikes'- blur (Fig. 8), 'Graffiti'-viewpoint (Fig. 6), 'Leuven'- lighting (Fig. 5).

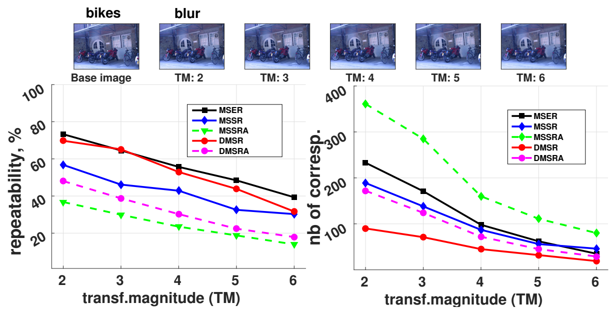


Fig. 8. Region detectors on 'Bikes', Oxford dataset.

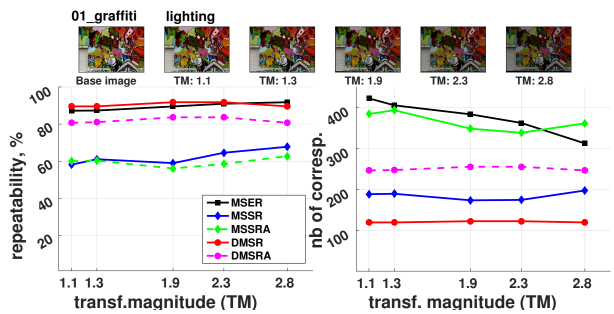


Fig. 9. Region detectors on '01_graffiti', OxFrei dataset.

3.2. OxFrei dataset

The creators of the Freiburg dataset addressed the issue of separating transformation from image content by applying all possible image test transformation to different base images, [12]. Evaluating detectors on the Freiburg dataset proved difficult due to the lack of full transformation parameters documentation. To address the lack of easy to use public dataset, a new set, 'OxFrei', combining the strong features of the Oxford and Freiburg dataset, is released, [**?**]. The Freiburg base images have been transformed with all the naturally obtained homographies of the Oxford dataset. In this way, 9 structured scenes have been created, each subject to 4 transformations (blur,lighting,scaling, viewpoint), in total 26 sequences.

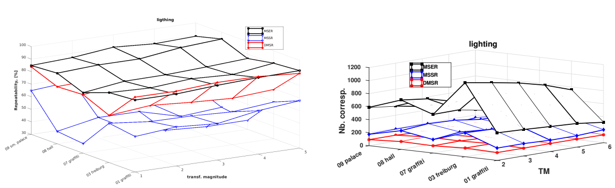


Fig. 10. Robustness of region detectors to lighting.

3.3. TNT hi-res benchmark

3.4. Animal and plant biometrics

The DMSR detector has been designed to address the problems of the scientists in the animal and plant biometrics [1]. The detector has been compared to MSER on several small animal individual photo-ID datasets (humpback whales, leatherback turtles, newts) and on a wood species identification dataset. In all cases DMSR produces less and perceptually moreacurate salient regions, figures 1, 2 and 7. For the wood microscopy images, it is important to measure the geometric properties of the cells, which is achievable with the DMSR regions, unlike the very large number of non-precise and overlapping MSER regions.

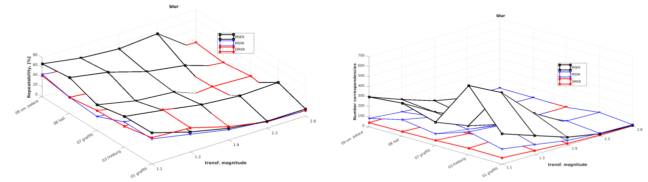


Fig. 11. Robustness of region detectors to blur.

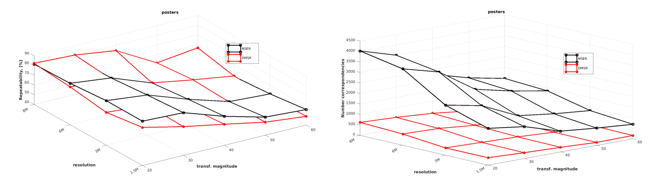


Fig. 12. Robustness of region detectors to image resolution and viewpoint.

4. CONCLUSIONS

Using data-driven binarization threshold and simple binary morphological operations leads to salient region detector, DMSR, with comparable to superior performance to MSER. It produces much smaller number of regions, a very desired property in the big data era. DMSR is better in dealing with blur, lighting and increased resolution. For large-scale detector bench-marking, high-resolution and transformation independent datasets should become the standard. DMSR gives perceptually salient regions, which makes it a good choice for scientific imagery analytic tasks.

5. REFERENCES

[1] H. Kuehl and T. Burghardt, "Animal Biometrics: quantifying and detecting phenotypic appearance," *Trends in Ecology & Evolution*, vol. 28, pp. 432–441, 2013.

[2] N. Kumar, P. Belhumeur, A. Biswas, D. Jacobs, W. Kress, I. Lopez, and J. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification," in *The 12th European Conference on Computer Vision (ECCV)*, October 2012, pp. 502–516.

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Inter-*

national Journal of Computer Vision, vol. 65, no. 1-2, pp. 43–72, November 2005.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *Proceedings BMVC*, 2002, pp. 36.1–36.10.

[5] K. Cordes, B. Rosenhahn, and J. Ostermann, "High-Resolution Feature Evaluation Benchmark," in *The 15th International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2013, pp. 327–334.

[6] 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.

[7] Per-Erik Forssén, "Maximally Stable Colour Regions for Recognition and Matching," in *CVPR*, 2007, pp. 1–8.

[8] Sh. Wang, W. Wang, D. Liu, F. Gu, and B. Dickson, "Enhanced Maximally Stable Extremal Regions with Canny Detector and Applicationin Image Classification," *Journal of Computational Information Systems*, vol. 10, no. 14, pp. 6093–6100, 2014.

[9] M. Donoser and H. Bischof, "3D Segmentation by Maximally Stable Volumes (MSVs).," in *ICPR (1)*, 2006, pp. 63–66.

[10] E. B. Ranguelova and E. J. Pauwels, "Morphology-Based Stable Salient Regions Detector," in *Proceedings of International Conference on Image and Vision Computing New Zealand*, 2006, pp. 97 – 102.

[11] E. B. Ranguelova and E. J. Pauwels, "Saliency Detection and Matching for Photo-Identification of Humpback Whales," *International Journal on Graphics, Vision and Image Processing*, 2006.

[12] P. Fischer, A. Dosovitskiy, and T. Brox, "Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT," *CoRR*, vol. abs/1405.5769, 2014.