

# DATA-DRIVEN REGION DETECTOR FOR STRUCTURED IMAGE SCENES

Elena Rangelova

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,

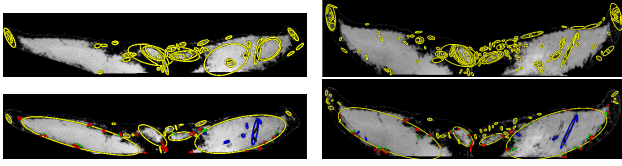
extracted from usually highly structured images. The generic region detectors often do not satisfy this need. In addition, there is a shortage of publicly available benchmarks for evaluating the robustness to transformations independently of the image content.

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.

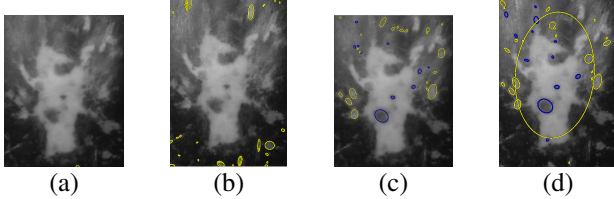
### 1.1. Related work

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. Top row: MSER, bottom row: DMSRA.



**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.

## 2. DATA-DRIVEN MORPHOLOGY SALIENT REGIONS

The DMRS detector offers simpler approach in comparison to MSSR and MSER. While MSER and MSSR process each gray-level image level set in search for stable salient regions, DMRS transforms the gray-scale (or color) image saliency directly into binary salient region detection problem.

### 2.1. Binary Salient Regions Detection

We have explained the morphology-based binary saliency detection in previous publications, the reader is referred to them for more details, [10, 11]. Here, a summary of the main notions and ideas is presented.

The claim is that the perceptual saliency in a binary image of a structured scene  $\mathbf{B} : \mathcal{D} \subset \mathbb{Z}^2 \rightarrow \{0, 1\}$  (1-white, 0-black) is only due to the spatial layout of the regions within the image. 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. A significant connected component (CC)  $\mathcal{B}^1$  is defined as a CC with area proportional to the area of the image by a factor of  $\Lambda$ . Other parameters are  $r$  -the radius of the morphological structuring element and  $\lambda$ - the area opening parameter for filtering too insignificant regions. 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 same notions are used for the MSSR detector and the regions are obtained using the morphology operations *hole filling* and *white/black top hat*. 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 denote detectors using all region types. These definitions are presented formally in Table 1 and the detection illustrated on a synthetic  $100 \times 100$  binary image with parameters  $\Lambda = 100$ ,  $r = 5$ ,  $\lambda = 10$  on Fig.3.

<b>ISS</b>	$A \cap S_{fb}^i = \{\mathbf{p} \in \mathcal{D}, \forall \mathbf{p} = f, \forall \mathbf{q} \in \partial S_{fb}^i, \mathbf{q} = b, \mathbf{q} \notin \partial \mathbf{B}\},$
2 types of ISS	$S_{10}^i, f = 1, b = 0$ (islands) $S_{01}^i, f = 0, b = 1$ (holes)
all ISS	$\mathbf{S}^i = S_{01}^i \cup S_{10}^i$
<b>BSS</b>	$S_{fb}^b : \{\mathbf{p} \in S_{fb}^b \subset \mathcal{B}^f, \forall \mathbf{p} = f, \mathbf{q} \in \partial S_{fb}^b \subset \partial \mathcal{B}^f, \forall \mathbf{q} = b\},$ $card(\partial \mathcal{B}^f) - card(\partial(\mathcal{B}^f \setminus S_{fb}^b)) < 2\pi r$
2 types of BSS	$S_{10}^b, f = 1, b = 0$ (protrusion) $S_{01}^b, f = 0, b = 1$ (indentation)
all BSS	$\mathbf{S}^b = S_{01}^b \cup S_{10}^b$
<b>Salient regions</b>	$\mathbf{S} = \mathbf{S}^i$ (DMSR), $\mathbf{S} = \mathbf{S}^i \cup \mathbf{S}^b$ (DMSRA)

**Table 1.** Binary saliency definitions used in Section 2.1.

## 2.2. Binarization algorithm

### 3. PERFORMANCE EVALUATION

#### 3.1. Standard datasets

3.1.1. *Oxford dataset*

3.1.2. *Freiburg dataset*

3.1.3. *Combined dataset*

3.1.4. *TNT hi-res benchmark*

#### 4. CONCLUSIONS

#### 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,” *International 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 Application in 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. Rangelova 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. Rangelova 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.