A Salient Region Detector for Structured Images

Elena Ranguelova
Netherlands eScience Center
Amsterdam, The Netherlands
Email: E.Ranguelova@esciencecenter.nl

***Abstract*—Finding correspondences between two images of the same scene or object, taken from different viewpoints and conditions, is a challenging task. Analyzing scientific imagery often requires the detected local features to coincide with the human perception, thus making the task even more complex. Ecologists use photo-identification methods in their population studies and conservation efforts. In addition to the task of identifying an individual plant or animal or classifying species, precise phenotypic measurements are needed. Generic detectors, such as the renowned Maximally Stable Extremal Regions (MSER) perform very well on structured images, but have difficulties with blur, lighting and increased resolution. The detected regions do not always correspond to semantically meaningful structures and their large number hampers scalability. This paper proposes a Data-driven Morphology Salient Regions (DMSR) detector which overcomes the above limitations. A new binarization algorithm uses a threshold derived from the data and the resulting binary image is analyzed for saliency using morphology. DMSR shows transformation invariance and comparable repeatability to MSER on several benchmarks while achieves better invariance to lighting, blur and on images with increasing resolution. This is achieved via significantly fewer regions, leading to better scalability. Some preliminary results on animal and plant images, indicate that DMSR could be a suitable approach for such wild-life biometric application as the detected regions correspond well to the semantic salient structures. The paper also introduces OxFrei - a dataset for transformation-independent detection evaluation.**

1. Introduction

The first fundamental step in numerous computer vision applications (wide baseline stereo matching, image retrieval, visual mining, etc.) is to reliably and repeatedly find the correspondence between a pair of different images of the same scene or object [1], [2], [3]. One class of methods-*region detectors* find distinct (salient) regions, which correspond to the same image patches, detected independently in each image. The detectors must be *invariant* to, usually, *affine* transformations (viewpoint, scaling, etc.) and various photometric distortions (scaling, lighting, blur, resolution, etc).

A decade ago, a performance evaluation paper by the Visual Geometry Group in Oxford compared existing region detectors [4]. A clear conclusion of the comparison was that *Maximally Stable Extremal Regions (MSER)* is the best performing detector for *structured* scenes, e.g., those containing homogeneous regions with distinctive boundaries [1]. MSER has become the de-facto standard in the field, e.g., it is in the MATLAB Computer Vision Systems Toolbox, OpenCV, VLFeat etc. Despite its success, the detector has several drawbacks: it is sensitive to image blur; it produces nested and redundant regions and its performance degrades with the increase of image resolution [5]. Figure 1 illustrates this degradation and the robustness of our proposed *Data-driven Morphology Salient Regions (DMSR)* detector to increased image resolution on a detail from the 'underground' image from TNT dataset [5]. While many of the MSER regions are not detected when the image resolution increased from 1.5 to 8 MPixel, the DMSR regions are detected consistently invariant to the resolution. Analysis in geometric scale-

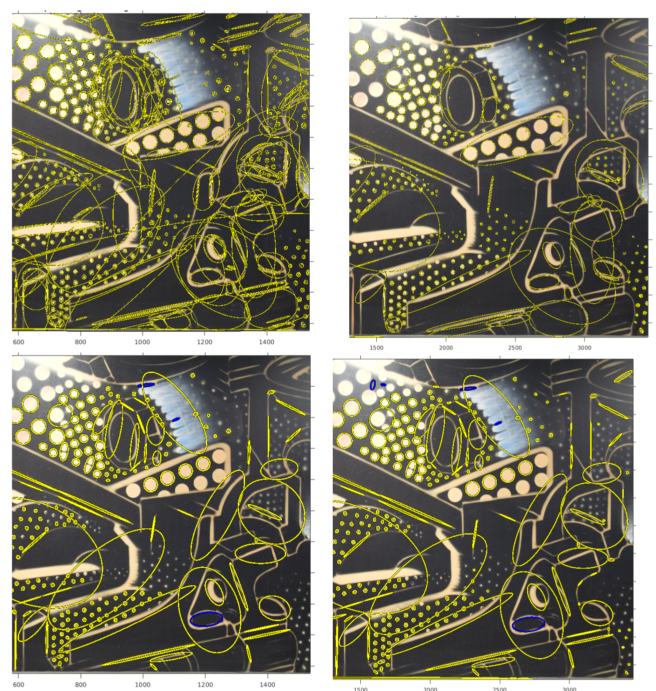

Figure 1. Region detection on the 'underground' image (detail) from the TNT dataset. Top row: MSER, bottom row: DMSR (proposed detector). Left: low 1.5, right: high 8 MPixel resolution

space have shown that the formulation of the region stability criterion makes MSER prefer regular shapes [6].

Since then, many researchers have proposed improvements to MSER with no drastic increase of performance. An MSER color extension, *Maximally Stable Color Region*, outperforms both an MSER-per-color-channel combination and a color blob detector [7]. Improving the MSER region distinctiveness by morphological dilation on the detected Canny edges is proposed in [8]. The improved detector shows better performance in classification application, but evaluation of repeatability is not reported. Chen et al. also combine MSER with Canny edges in order to cope with blur for detecting text in natural images [9]. Kimmel et al. not only point out the preference of the algorithm towards regular shapes, but also propose several reinterpretations of the stability measure in order to define more informative shape descriptors, [6].

Interesting research has been conducted by Martins et al. who propose *feature-driven MSER*, called *Stable Salient Shapes* (SSS) by extending the concept of stable regions from detection using the original image, to a boundary-related features enhanced representation, [10], [11]. This is done via "feature-highlighting" of edges and ridges generating saliency maps for each feature and are then used as domains for the MSER detector. As a result, SSS is less sensitive to blurring, but it also detects more regions per image in comparison to MSER. While the authors consider the latter an improvement since it decreases the detector's sensitivity to occlusion, we argue that in the context of scaling up or when processing animal biometrics imagery usually obtained with care for minimal occlusion, that is a drawback. Their approach also improves the *completeness* of the local features ("the information contained in the image should be preserved by the features as much as possible" as defined in [12]) as explained and studied in detail in [13]. The authors claim that the completeness property of SSS makes it a suitable detector to solve object recognition tasks. Unfortunately they do not share their software implementation of SSS.

A region detector suitable for object class recognition, the *Principal Curvature-Based Region (PCBR)* using curvilinear structures (ridges), has been introduced in [14]. PCBR differed from MSER in two aspects- it analyses regions in scale space, thus providing different levels of region abstraction and also overcomes the problems caused by local variations within regions by focusing on region boundaries rather than interiors. PCBR is similar to our approach in using morphological operators for detecting robust watershed regions, but on the principal curvature image instead of on the intensity (or binarized) image. The reported average repeatability on the Oxford dataset, [4], is worse than that of MSER given that PCBR has been designed for object-class recognition. The PCBR detector in combination with object-class recognition algorithms has been shown to distinguish between two related species of stone-fly larva with very similar appearance better than using other detectors. The comparison, however, is with Kadir's salient detector, [15] and Hessian-affine, not with MSER. Dzeng et al. point out the importance of designing feature detectors more suitable for object recognition tasks different for the detectors designed for wide-baseline matching of scenes as MSER was originally developed for. Our aim is, however, to addresses both applications simultaneously.

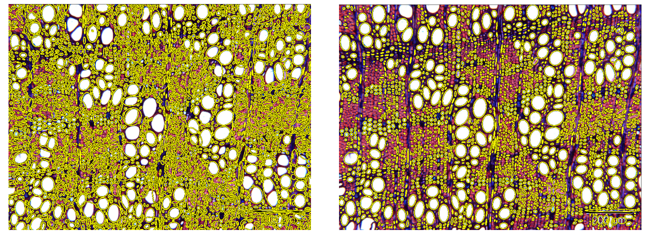

Figure 2. Salient region detectors on microscopy wood images. Left: MSER (every second region is shown), right: proposed detector, DMSR

While most research has focused on generic applications, the emerging fields of *animal and plant biometrics* are attracting more attention [16], [17]. Computer vision is becoming a vital technology enabling the wild-life preservation efforts of ecologists. Along with the individual or species photo-identification, the scientists wish to obtain reliable measurements of meaningful structures from images. For example the automated identification of cell structures is one of the new challenges to be met in studies of the structural biology of plants [19]. Analysis of these microscopic features from anatomical sections of wood are very important for studying the secondary growth and development of trees [18]. Classifying the cell types can be approached by examining their shape, size and spatial distribution. At the same time these characteristics are used for automated wood species classification [20]. Therefore, the automatic detector of regions corresponding to wood cells has to perform well for both tasks ecology scientists face: identification of species or individuals and phenotypic measurements. The generic detectors do not satisfy this need, they produce redundant overlapping regions which often do not coincide with semantic structures. Figure 2 illustrates that it is not possible to obtain accurate statistics on the cell properties using the regions from the MSER detector, while the regions found by our detector would enable such wood anatomy research.

Although crucial for the development of detectors, there is a shortage of region-based evaluation benchmarks, especially for performance analysis independently of the image content. The standard *Oxford dataset* is very small: eight test sequences containing six (one base and five transformed) images of the same scene each. Every pair (base, transformed) is related via a given transformation matrix (homography) [4]. The *Freiburg dataset* contains 416 higher resolution images, generated by transforming 16 base images in order to de-tangle transformations from content [21]. The *TNT dataset* contains versions of the same viewpoint sequences with increasing resolution from 1.5 to 8 MPixel per image. Highly accurate image pair homographies are given. It is suitable for evaluating robustness to resolution rather than to transformations [5]. A more resent and larger dataset is the point feature *DTU Robot Data Set*, [22]. It consists of

60 scenes acquired from 119 positions (executed with a robotic arm), totaling $135,660$ color images of a resolution of 1200×1600. While of a much larger scale and with better precision of correspondences than the Oxford dataset, it is a very limited indoor setup, hence does not capture natural outdoor variations of acquisition conditions. Interestingly, the authors report the perform ace of MSER (measured using Recall and not Repeatability) on their dataset as moderate while still better than the poorly performing Edge-based (EBR) and Intensity-based Regions (IBR) [23].

This paper contributes to solving the identified problems. We propose a new regions detector, DMSR, and made the software available as open source [24]. It is related to the *Morphology-based Stable Salient Regions (MSSR)* detector that we have developed in the context of humpback whale identification [25], [26]. DMRS includes a data-driven binarization, robust-to-lighting-and-blur, that yields a much smaller number of regions and is more stable across transformations. Unlike SSS, which consists of two separate detection steps a feature saliency map followed by MSER detection, we propose an integrated solution for both image correspondence and object recognition (or identification) tasks simultaneously. It has similar or higher (lighting, blur and increased resolution) repeatability compared to MSER, while detecting non-redundant perceptually salient regions.Also, we compose and share an openly available dataset, OxFrei, combining the natural homographies of the Oxford and the higher resolution images of the Freiburg datasets [24].

2. Conclusion

The conclusion goes here.

Acknowledgments

The authors would like to thank...

References

[1] 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.

[2] Foncubierta-Rodrguez *et al.*, "Region-based volumetric medical image retrieval," in *SPIE Medical Imaging: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, 2013.

[3] S. Escalera, P. Radeva, and O. Pujol, "Complex salient regions for computer vision problems," in *CVPR*, 2007.

[4] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, November 2005.

[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 *et al.*, "Are MSER Features Really Interesting?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2316–2320, 2011.

[7] P.-E. Forssén, "Maximally Stable Color Regions for Recognition and Matching," in *Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[8] S. Wang *et al.*, "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] H. Chen *et al.*, "Robust Text Detection in Natural Images with Edge-enhanced Maximally Stable Extremal Regions." in *ICIP 11*, 2011.

[10] P. Martins, C. Gatta, and P. Carvalho, "Feature-driven maximally stable extremal regions." in *VISAPP (1)*, G. Csurka and J. Braz, Eds. SciTePress, 2012, pp. 490–497.

[11] P. Martins, P. Carvalho, and C. Gatta, "Stable salient shapes," in *2012 International Conference on Digital Image Computing Techniques and Applications, DICTA 2012, Fremantle, Australia, December 3-5, 2012*, 2012, pp. 1–8.

[12] 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.

[13] P. Martins, P. D. Carvalho, and C. Gatta, "On the completeness of feature-driven maximally stable extremal regions," *Pattern Recognition Letters*, vol. 74, pp. 9–16, 2016.

[14] H. Deng, W. Zhang, E. N. Mortensen, T. G. Dietterich, and L. G. Shapiro, "Principal curvature-based region detector for object recognition." in *CVPR*. IEEE Computer Society, 2007.

[15] T. Kadir, A. Zisserman, and M. Brady, *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. An Affine Invariant Salient Region Detector, pp. 228–241.

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

[17] N. *et al.*, Kumar, "Leafsnap: Computer Vision System for Automatic Plant Species Identification," in *The 12th European Conference on Computer Vision (ECCV)*, October 2012, pp. 502–516.

[18] G. Brunel and other, "Automatic identification and characterization of radial files in light microscopy images of wood," *Annals of Botany*, vol. 114, no. 4, pp. 829–890, 2014.

[19] P. Quelhas, J. Nieuwland, W. Dewitte, A. M. Mendonça, J. Murray, and A. Campilho, *Image Analysis and Recognition: 8th International Conference, ICIAR 2011, Burnaby, BC, Canada, June 22-24, 2011. Proceedings, Part II*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Arabidopsis Thaliana Automatic Cell File Detection and Cell Length Estimation, pp. 1–11.

[20] P. Gasson, "How precise can wood identification be? wood anatomys role in support of the legal timber trade, especially cites," *IAWA Journal*, vol. 32, no. 2, pp. 137–154, 2011.

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

[22] H. Aans, A. L. Dahl, and K. S. Pedersen, "Interesting interest points - a comparative study of interest point performance on a unique data set." *International Journal of Computer Vision*, vol. 97, no. 1, pp. 18–35, 2012.

[23] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vision*, vol. 59, no. 1, pp. 61–85, Aug. 2004.

[24] E. Ranguelova, "Large scale imaging: Data, software, results," DOI: http://dx.doi.org/10.5281/zenodo.45156, Jan. 2016.

[25] 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.

[26] ——, "Saliency Detection and Matching for Photo-Identification of Humpback Whales," *International Journal on Graphics, Vision and Image Processing*, 2006.