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 in different conditions, is a challenging task. Furthermore, in the analysis of scientific imagery, it must be possible in terms of human perception to appreciate detected local features, thus making the task even more complex. One method, used by ecologists in their population studies and conservation efforts, is photo identification. In addition to identifying individual plants or animals or classifying species, precise phenotypic (appearance) measurements are needed. A renowned generic feature detector, Maximally Stable Extremal Regions (MSER), performs very well on structured images, but has difficulties with blur, lighting and increased resolution. The detected regions do not always correspond to semantically meaningful image structures, and the large number of regions hampers scalability. This paper proposes a Data-driven Morphology Salient Regions (DMSR) detector which overcomes these limitations. We present a new binarization algorithm which uses a threshold derived from the data; the resulting binary image is analyzed for saliency using morphology. DMSR shows transformation invariance and comparable repeatability to MSER on several evaluation benchmarks while obtaining better invariance to lighting, blur and resolution. This is achieved via significantly fewer regions, leading to better scalability. Preliminary results on animal and plant images indicate that DMSR could indeed be a suitable approach for wild-life biometric applications as the detected regions correspond well to the semantic salient image structures. We also introduce OxFrei - a dataset for transformation-independent detection evaluation.**

1. Introduction

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

A decade ago, a performance evaluation paper by the Visual Geometry Group in Oxford compared existing region detectors [4]. A clear conclusion 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: it has been implemented as part of MATLAB, OpenCV, VLFeat, etc. However, 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 in contrast to the robustness to resolution of our proposed *Data-driven Morphology Salient Regions (DMSR)* detector on a detail taken from the 'underground' image in the TNT dataset [5]. While many of the MSER regions are not found when the image resolution increases from 1.5 MPixel to 8 MPixel, the DMSR region detection is consistently resolution invariant. Analyses in geometric scale-space have shown that the formulation of the region stability criterion makes MSER prefer regular shapes [6].

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

1.1. Envisioned scientific applications

While most research has been focused on generic applications, the fields of *animal and plant biometrics* are attracting more attention [7], [8]. Computer vision is becoming a vital technology enabling the wild-life preservation efforts of ecologists. Along with individual or species photo-identification, 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 [9]. Analysis of these microscopic features from anatomical sections of wood are very important for studying the secondary growth and development of trees [10]. 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, which is very important to fight illegal logging [11]. Therefore, the automatic detector of regions corresponding to wood cells has to perform well for the two tasks ecology scientists face: identification of species or individuals followed by phenotypic measurements. The generic detectors do not satisfy this need: they produce redundant overlapping regions which often do not coincide with semantic structures (Figures 16, top and 17, left).

1.2. Evaluation benchmarks

Although crucial for the development of detectors, there is a shortage of region-based evaluation benchmarks, especially for performance analysis that is independent 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]. Each transformation is present only in one test sequence in the structured images and one in the textured images group. The *Freiburg dataset* contains 416 higher resolution images, generated by transforming 16 base images in order to de-tangle transformations from content [12]. 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. This dataset 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*, [13]. 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 limited indoor setup and hence does not capture natural outdoor variations of acquisition conditions. Interestingly, the authors report the performance (recall, not repeatability) of MSER on their dataset as only moderate. Also, the image resolutions present in the datasets are still relatively low compared to the demand of the real world digital image applications.

1.3. Contributions

We propose an integrated solution for both image correspondence and object recognition tasks simultaneously. Our new salient regions detector, DMSR, is related to the *Morphology-based Stable Salient Regions (MSSR)* detector that we developed earlier [14], [15]. DMRS includes a data-driven binarization that is robust to lighting and blur and yields a much smaller number of regions and is more stable across transformations. It has similar or higher repeatability (lighting, blur and increased resolution) compared to MSER, while detecting non-redundant perceptually salient regions. DMSR consistently achieves the lowest number of detected regions, which will lead to the best scalability during the following region descriptors matching step. In addition, we have composed and shared the new OxFrei dataset combining the natural homographies of the Oxford and the higher resolution images of the Freiburg datasets. The dataset has been designed to facilitate transformation-independent detectors performance evaluation. All experimental results, dataset and open-source software are available online to facilitate research repeatability [16].

2. Related work

Many researchers have proposed improvements to MSER, but without drastic increase in performance. An MSER color extension, *Maximally Stable Color Region*, outperforms both an MSER-per-color-channel combination and a color blob detector [17]. Improving the MSER region distinctiveness by morphological dilation on the detected Canny edges is proposed in [18]. The improved detector shows better performance in a classification application, but evaluation of repeatability is not reported. Chen et al. also combine MSER with Canny edges to cope with blur for detecting text in natural images [19]. Kimmel et al. propose several reinterpretations of the stability measure 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 one from a boundary-related features enhanced representation [20], [21]. This is done via "feature-highlighting" of edges and ridges generating saliency maps which are used as domains for the MSER detector. As a result, SSS is less sensitive to blurring, but it also detects even 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 in processing animal biometrics imagery usually obtained with care for minimal occlusion, this 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" [22]), [23]. The authors claim that the completeness property of SSS makes it a suitable detector to solve object recognition tasks, but their software is not shared.

A region detector suitable for object-class recognition, the *Principal Curvature-Based Region (PCBR)*, uses curvilinear structures (ridges) to enhance the regions [24]. PCBR differs from MSER in two aspects: it analyses regions in scale space, thus providing different levels of region abstraction, and it also overcomes the problems caused by local variations within regions by focusing on their boundaries rather than interiors. PCBR is similar to our approach in using morphological operators for detecting robust watershed regions, but this is done on the principal curvature image instead of on the intensity (or binarized) image. The reported average repeatability on the Oxford dataset is worse than that of MSER given that PCBR has been designed for object-class recognition. PCBR in combination with object-class recognition algorithms has been able to distinguish between two related species of stone-fly larvae with very similar appearance. The comparison, however, is with Kadir's salient detector [25] and Hessian-affine, not with MSER. Dzeng et al. point out the importance of designing detectors tailored to object recognition tasks separately from the detectors designed for general applications like wide-baseline stereo matching (such as MSER). Our aim is, however, to addresses both applications with the same technology.

3. Data-driven Morphology Salient Regions Detection

MSER decomposes a gray-scale image into binary cross-sections and evaluates the stability of the connected components across sections to determine maximally stable regions. In contrast, DMSR starts with a data-driven binarization. The single binary image is then analyzed for saliency using binary morphology.

3.1. Binary Salient Regions Detection

We claim 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 types of salient regions grouped into: *Inner Salient Structures (ISS)* and *Boundary Salient Structures (BSS)*. The 2 types of ISS are (1) *holes* – set of connected black pixels entirely surrounded by white pixels, and (2) *islands* – set of connected white pixels surrounded by black ones, i.e. the inverse of holes. A significant connected component (CC) $\mathcal{B}^1$ is defined as a CC with area proportional to the image area by Λ. The 2 BSS are (3) *protrusions*- set of white pixels on the border of a significant CC, which if pinched off from the CC will increase its boundary with no more than $2\pi r$, where r is the radius of the morphological structuring element (SE), and (4) the *indentations*- protrusions inverse. These types, defined in Table 1, also apply to the MSSR detector [15].

The regions are obtained from $\mathbf{B}$ by morphological operations [26]. The *hole filling* operation $\bullet(\cdot)$ on the set of all white pixels $\mathbf{B}^1$ intersected with the set of all black pixels $\mathbf{B}^0$ can be used to detect holes:

$$S^i_{01} = \rho^i_{01}(\mathbf{B}) = \bullet(\mathbf{B}^1) \cap \mathbf{B}^0. \quad (1)$$

TABLE 1. BINARY SALIENCY DEFINITIONS USED IN SECTION 3.1.

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

Islands can be obtained either similarly but from the inverted image or by identifying all non-significant CCs:

$$S^i_{10} = \rho^i_{10}(\mathbf{B}) = \bullet(\mathbf{B}^0) \cap \mathbf{B}^1 = \mathbf{B}^1 \backslash \bigcup_j \mathcal{B}^1_j. \quad (2)$$

The morphological *opening* operator (based on SE E-disk with radius r) $\gamma_E(\mathbf{B})$ generally smooths a CC's contour eliminating thin protrusions. As a consequence, protrusions can be obtained by subtracting an opened CC from the original (this is known as the *white tophat transform* $WTH_E(\mathbf{B}) = \mathbf{B} - \gamma_E(\mathbf{B})$):

$$S^b_{10} = \rho^b_{10}(\mathbf{B}) = \bigcup_j WTH_E(\mathcal{B}^1_j). \quad (3)$$

Morphological *closing* $\phi_E(\mathbf{B})$ tends to narrow breaks and long thin indentations. Therefore, indentations can be picked up by applying the *black top hat (BTH)* transform $BTH_E(\mathbf{B}) = \phi_E(\mathbf{B}) - \mathbf{B}$:

$$S^b_{01} = \rho^b_{01}(\mathbf{B}) = \bigcup_j BTH_E(\mathcal{B}^1_j) \quad (4)$$

or by applying the WTH to the inverted image.

The binary saliency operator ρ is therefore defined via:

$$\rho = \gamma_\lambda \circ (\rho^i_{01} \cup \rho^i_{10} \cup \rho^b_{01} \cup \rho^b_{10}), \quad (5)$$

where the *area opening* operator γ_λ removes isolated regions smaller than λ pixels.

Figure 2 illustrates the exact shaped binary salient regions detected from a synthetic 100×100 binary image with DMSR parameters $\Lambda = 1/100$, $r = 5$ and $\lambda = 10$.

Figure 2. Binary salient regions detection. Color coding: ISS: holes - blue, islands - yellow; BSS: indentations - green, protrusions - red.

The ISS are similar to the definition of the MSER+ and MSER- regions [1]. While MSER detects both types, DMSR (and MSSR) is parametrised with the desired type(s) of salient regions to look for. In this paper, detectors using only ISS, i.e., directly comparable to MSER, are denoted by DMSR/MSSR, while DMSRA/MSSRA are detectors using all region types.

3.2. Data-driven binarization

Any gray-scale image $\mathbf{I} : \mathcal{D} \subset \mathcal{Z}^2 \rightarrow \mathcal{T}$, where $\mathcal{T} = \{0, 1, \ldots, t_{max}\}$ and $t_{max} = 255$ is the maximum gray value $2^n - 1$ encoded by $n = 8$ bits, can be decomposed into cross-sections at every possible level t: $\mathbf{I} = \sum_{t \in \mathcal{T}} CS_t(\mathbf{I})$. Obtaining a section at level t is equivalent to thresholding the image at threshold t: $CS_t(\mathbf{I}) = 1.(\mathbf{I} > t) + 0.(\mathbf{I} < t)$ is a binary image.

Both MSER and MSSR need multiple binary cross-sections: the former uses minimum area change of a CC as stability criteria, while the latter generates cumulative binary saliency masks to determine salient regions. In this paper we argue that it is possible to obtain a single optimum binarization inspired by the notion of feature completeness [22]. Along with robustness (invariance to transformations), speed (detection should be fast) and sparseness (features should be less than the image itself), an important prerequisite for a good detector is completeness - the salient information present in an image should be maximally preserved. We argue that for a detector which should also reflect saliency, preserving a maximum number of CCs is a way of achieving completeness. On the other hand, larger (significant) CCs are perceptually more salient than smaller CCs.

We define, then, three sets of connected components in $CS_t(\mathbf{I})$: $\mathcal{A}_t$- *all*, $\mathcal{L}_t$- the *large* and $\mathcal{V}_t$- the *very large* CCs. The CC size in the last two categories are defined by $\Lambda_{\mathcal{L}}$ and $\Lambda_{\mathcal{V}}$ fraction of the image area $A_{\mathbf{I}}$. Let us denote the normalized number of elements in a set ($|\cdot|$) by $\|\cdot\| = |\cdot| / \max_{t \in \mathcal{T}} |\cdot|$. Finding the optimal binarization threshold t_{opt} is then defined as:

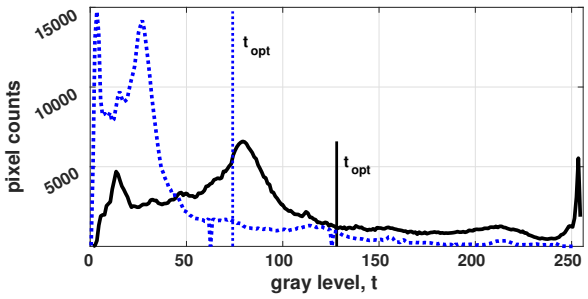
$$t_{opt} = \arg\max_{t \in \mathcal{T}} (w^{\mathcal{A}} \|\mathcal{A}_t\| + w^{\mathcal{L}} \|\mathcal{L}_t\| + w^{\mathcal{V}} \|\mathcal{V}_t\|), \qquad (6)$$

where $w^{\cdot}$ are the weights per set of CCs.

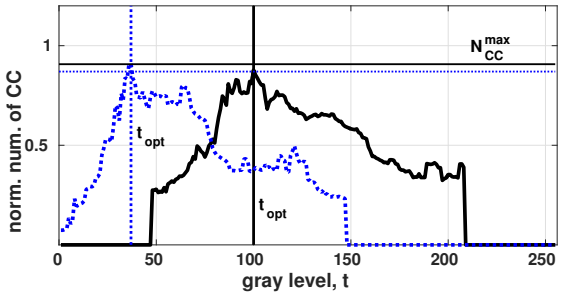
In comparison to the standard Otsu thresholding which does not select a single CS stable across photometric transformations, choosing t_{opt} as defined ensures a transformation-invariant stable number of regions. Figures 3 and 4 show the Otsu and the data-driven thresholding in respect to lighting. When the lighting decreases, the Otsu thresolding produces fewer regions with often smaller extent (e. g. the cars' windows), while the proposed binarisation generates regions with the same spatial extent (Fig. 4).

3.3. 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 3.1 and [14], [15]. As a result, DMSR produces fewer non-overlapping and perceptually salient regions compared to MSER. This is illustrated in Figures 1, 5, 13, 14, 16 and 17 where the regions are visualized by their equivalent ellipses, not by their exact detected shapes.



(a) Otsu



(b) Max number CC

Figure 3. Finding the optimal threshold for two images from the 'Leuven' sequence (Oxford dataset, lighting): the base image- solid black line, the forth image - dotted blue line.

Figure 4. Binarization of two images of the 'Leuven' sequence (lighting). Left: base image, right: forth image; top row: gray scale; middle row: Otsu binarization, bottom row: proposed binarization.

4. Performance Evaluation

We have applied the standard performance evaluation metrics - the *repeatability score* (R) and the *number of correspondences* (N_C) [4]. The maximum overlap error between matching regions is set to 40% as in [4]. The repeata-

Figure 5. Region detectors on the base image of the 'Graffiti' sequence, Oxford dataset. Left: MSER, right: DMSR

bility score between a pair of base and transformed image, $(\mathbf{I}_B, \mathbf{I}_T)$, is the ratio between N_C in the part of the images with common content and the minimum number of regions in the image pair. Since we were intrested in finding correspondencies between salient regions from structured images only, we focused on this subset from each dataset and ignored the textured scenes (e.g. 'Trees', 'Bark' etc. from the Oxford dataset). We considered five region detectors for evaluation: MSER, MSSR(A) and DMSR(A). For MSER we used the original Matas software with its default settings and the (D)MSSR(A) parameters are: $r = 0.02 * \sqrt{A_{\mathbf{I}}/\pi}$, $\lambda = 3r$, $\Lambda_{\mathcal{L}} = 0.001$, $\Lambda_{\mathcal{V}} = 0.01$. All performance plots and detected regions on all data are available online [16].

4.1. Binarization

In order to determine the binarization weights from equation (6), we have tested all possible meaningful combinations of weights ($\sum w^{\cdot} = 1, w^{\cdot} \in [0.1, 0.6]$) for the sequences of the Oxford dataset.

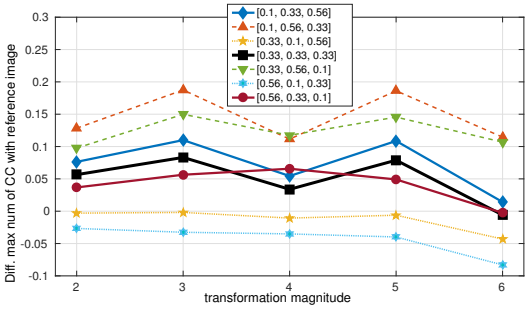


Figure 6. Determining binarization weights. Difference between maximum number of CC for each pair of images (base, transf) for 7 weight combinations. Blur sequence ('Bikes'), Oxford dataset.

Figure 6 shows several of the plots of the difference between the normalized maximum number of CC of the first reference image and each of the 5 transformed images from the blur sequence 'Bikes', Oxford dataset, for several combinations of weights. Those configurations for which the differences are below 0 are not desirable, since the number of CC decreases with the transformation. The combination of weights with the smallest and uniform positive differences and preferably small for the most extreme transformation (the 6th image of the sequence) is preferred. After testing across all transformations, we have concluded that the most constant maximum number of CC is achieved when all weights are equal, i.e. $w^{\cdot} = 0.33$.

4.2. Oxford dataset

Each image sequence of the Oxford dataset consists of one base and five increasingly distorted images [4]. 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 one transformation T. Therefore, it is not possible to separate transformation from image content. On the other hand, the Oxford dataset provides real homographies, unlike many other datasets.

For the viewpoint transformation ('Graffiti'), the best repeatability R is achieved by DMSR with up to 72% for a $40°$ viewpoint change, which is 10% more than the second performing MSER (Figure 7). DMSR performs worse on the scale ('Boat'), but as well as or better than MSER on the blur ('Bikes', Fig. 8) and lighting ('Leuven') sequences [16]. These R scores for all sequences of the dataset are achieved with consistently the smallest number of detected regions by DMSR, making the detector the most scaleable.

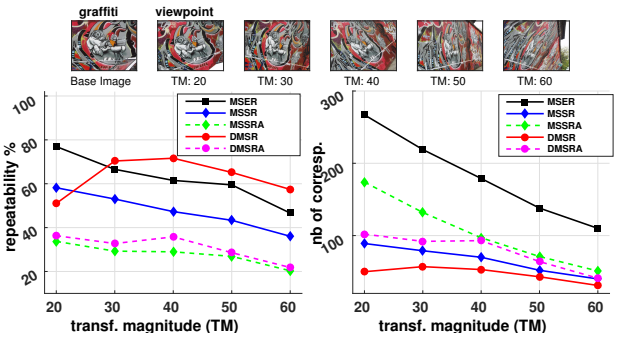


Figure 7. Region detection on 'Graffiti', Oxford dataset.

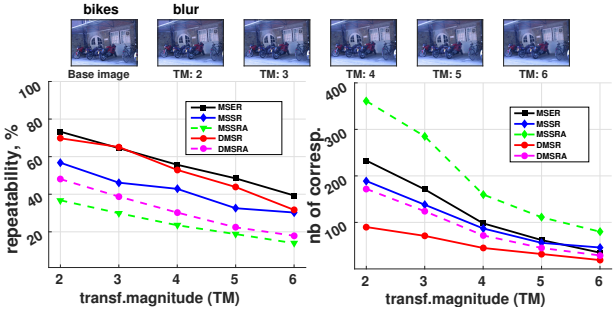


Figure 8. Region detection on 'Bikes', Oxford dataset.

4.3. OxFrei dataset

A major drawback of the Oxford dataset is that one cannot determine whether the detector is robust to a transformation T, as every sequence represents only one T. The creators of the Freiburg dataset separated T from the image content by applying a few transformations (alas not fully documented) to different base images [12]. To address the shortage of evaluation datasets, we have released OxFrei which combines the strong features of the Oxford and Freiburg datasets [16]. We have transformed the Freiburg base images with all homographies of the Oxford dataset. In this way, we have created 54 images in 9 structured scenes each under realistic blur, lighting, scaling and viewpoint transformations.

The dataset allows a transformation-independent robustness study by comparing performance on all data subject to the same realistic T. Figures 9 and 10 show the R score for lighting for one and blur for a few sequences. The standard plots (cf. Fig. 9) are cross-sections along the data dimension of the 3D plots (cf. Fig. 10) for one sequence. The 2D plots for all OxFrei experiments are available online [16]. From all experiments, we concluded that MSER is better for zoom and viewpoint (the latter contrasting with the result on the single Oxford 'Graffiti' sequence), while DMSR is robust to lighting and blur (Fig. 10). Again, the number of detected regions by DMSR is consistently the smallest across different image content and transformations.

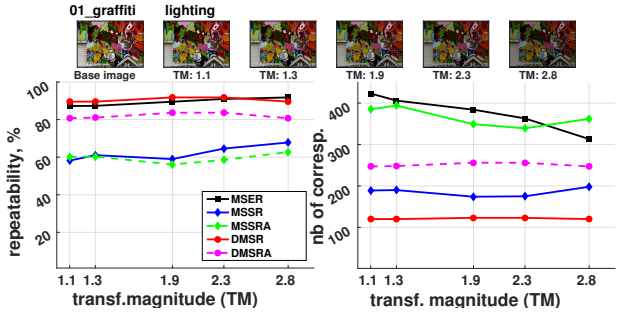


Figure 9. Region detection on '01_graffiti', OxFrei dataset.

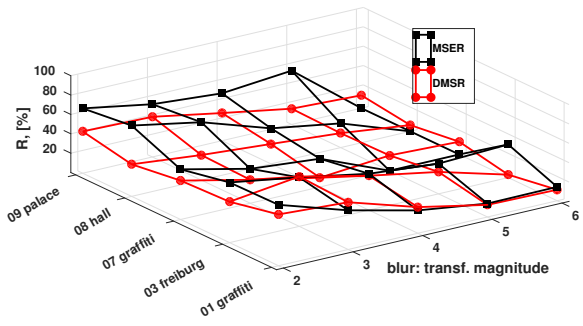


Figure 10. Robustness of region detectors to blur on five sequences of the OxFrei dataset.

4.4. TNT hi-res benchmark

The R score of all detectors from the Oxford evaluation study drops on hi-res images [5]. On the 'underground' sequence from the TNT set, MSER loses up to 25% between 1.5 Mpx (R_1) and 8Mpx (R_4) resolutions. On the contrary, DMSR increases the R score as resolution increases on 'underground' and 'posters' sequences with up to 75%, which is a 15% increase from R_1 to R_4 for a $40°$ viewpoint change (Figure 11). We can even observe that easily by qualitative inspection of the actual detected regions as we already have shown on Figure 1. While MSER increasingly loses features as the image resolution increases, the number and location of DMSR regions remains stable.

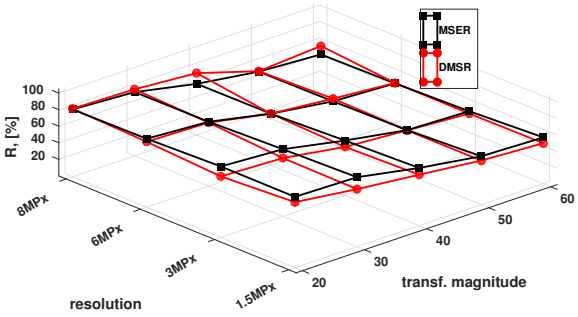


Figure 11. Robustness of region detectors to image resolution and viewpoint. 'Posters', TNT dataset.

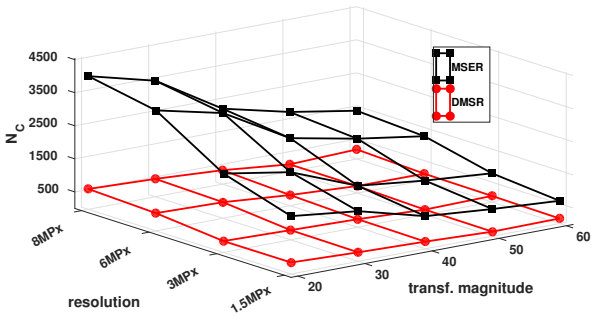


Figure 12. Number of region correspondences versus image resolution and viewpoint. 'Posters', TNT dataset.

Again, the N_C plane of DMSR has the lowest values and the smallest slope of all detectors (Figs. 7, 8, 9, right and Fig. 12 and [16]). The number of detected DMSR regions is up to an order of magnitude lower compared to MSER - crucial for the efficiency of the following matching step on large-scale image datasets.

4.5. Animal and plant biometrics

Challenged by the photo-identification and phenotypical measurement tasks of ecologists, we aimed to test if our detector is able to solve both problems. Photo-identification of individuals or species involves the comparision of a

new image to the exsisting catalogue of known individuals or species, simiar to many generic applications. Phenotypic measurements invole the computation of properties of (exact-shaped) regions with semantic meaning.

Due to the lack of openly accessible large-scale annotated animal or plant biometric datasets, we performed only preliminary experiments on few small datasets, from which we could gain some insights into the performance of the detector. We have qualitatively compared DMSR and MSER on several small animal individual photo-ID datasets (humpback whales, leatherback turtles, newts) and on a wood species identification dataset [15], [27], [28].

In all cases, DMSR produced fewer and perceptually more acurate salient regions, as illustrated by Figures 13, 14, 15 and 16, 17. For some of the leatherback turtles, the MSER did not find any or very few regions, which will make the subsequent matching for photo-identification impossible. On the other hand, the DMSR regions were detected consistently and repeatedly (Fig. 15). For the wood microscopy images, we observed 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 (Fig. 16, 17). Also, we used the exact shapes of the detected regions to compute the wood cell properties (such as size, eccentricity and orientaiton) using those for successful classification of 19 images into 9 wood species [29].

Another advantage of DMSR is the option to selectively detect only some types of salient regions (unlike MSER or any other detector) in which the ecologists are interested. In the wood microscopy images only the light cells are of interest and hence we can chose only to detect 'islands'. Using all 4 types of regions does not seem to improve performance over using only ISS (holes and islands), with the exception of detecting markings on humpback whale tails (Figs. 13, 14).

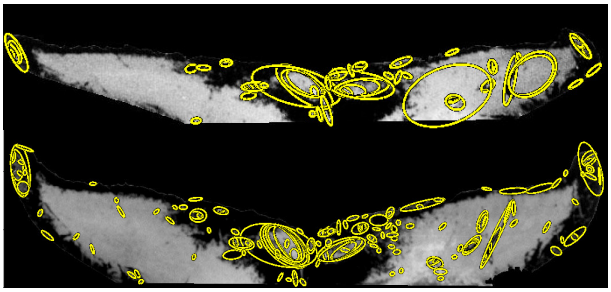


Figure 13. MSER detection on two images of the tail of a humpback whale.

5. Conclusion

Combining data-driven binarization with morphological operations yields a region detector with comparable to superior performance to MSER on various datasets. DMSR produces a much smaller number of regions - a much desired property in large-scale processing. It can cope better with blur, lighting and increased resolution. Furthermore,

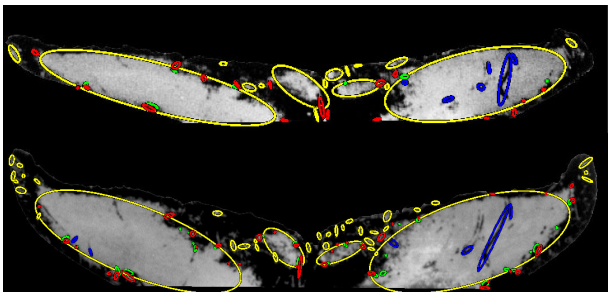


Figure 14. DMSRA detection on two images of humpback whale tail. The salient region types are color-coded like on Fig. 2

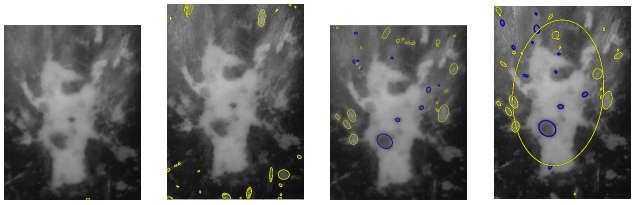


Figure 15. Region detection on two images of the pineal spot of the same leatherback turtle. (a),(b): MSER, (c),(d): DMSR. Note the lack of relevant regions in (a).

it detects semantically meaningful salient regions, which makes it a promising candidate for scientific imagery analytics, especially for the photo-identification and phenotypic measurement tasks. For detection evaluation, high-resolution transformation-independent datasets, like the OxFrei we introduced, should become the standard. In our future work, we plan to formally evaluate the completeness of the DMSR detector and to run extensive validation experiments on large-scale animal and plant biometric datasets.

Acknowledgments

The authors would like to thank Frederic Lens from the Naturalis biodiversity institute in Leiden, Netherlands, for providing the wood microscopy images [28].

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.

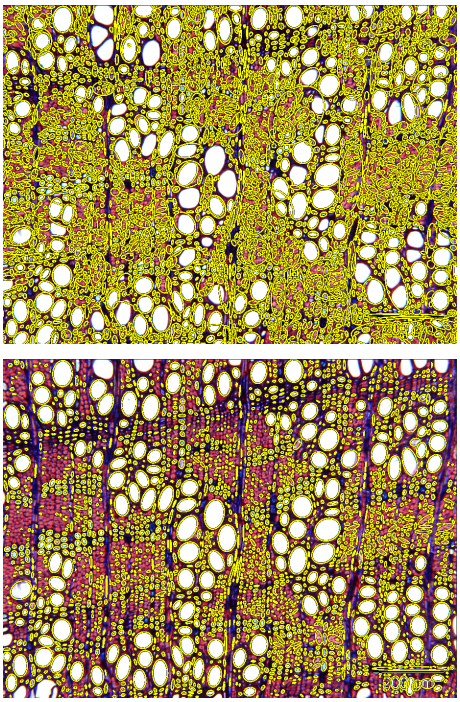


Figure 16. Salient region detectors on microscopy wood (Argania spinosa) image. Elliptic representation. Top: MSER (every second region is shown), bottom: DMSR

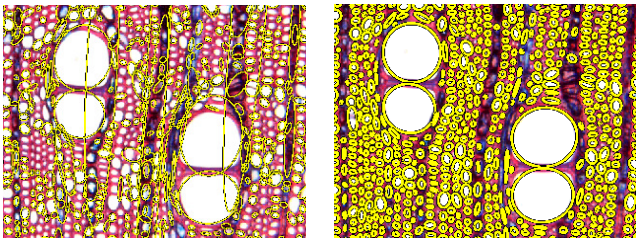


Figure 17. Salient region detectors on microscopy wood (Chrys afr) image (detail). Elliptic representation. Left: MSER (every 3rd region is shown), right: DMSR

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

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

[9] 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.

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

[11] 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.

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

[13] 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.

[14] 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.

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

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

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

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] P. Soille, *Morphological Image Analysis*. Springer, 2003.

[27] E. Pauwels, P. de Zeeuw, and D. Bounantony, "Leatherbacks matching by automated image recognition," in *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects, 8th Industrial Conference, ICDM 2008, Leipzig, Germany, July 16-18, 2008, Proceedings*, 2008, pp. 417–425.

[28] F. Lens, "Microscopy images wood," http://www.naturalis.nl/en/, naturalis Biodiversity Center, Leiden, The Netherlands.

[29] E. Ranguelova, "Wood image classification- initial results. nlesc report." http://knowledge.esciencecenter.nl/books/wood-image-classification/, May 2016.