Pattern Recognition Letters 74 (2016) 9–16

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

On the completeness of feature-driven maximally stable extremal regions ☆

Pedro Martins [a,*], Paulo Carvalho [a], Carlo Gatta [b]

[a] CISUC, Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal
[b] Computer Vision Center, Autonomous University of Barcelona, 08193 Bellaterra, Spain

ARTICLE INFO

Article history:
Received 18 June 2015
Available online 22 January 2016

Keywords:
Local features
Completeness
Maximally Stable Extremal Regions

ABSTRACT

By definition, local image features provide a compact representation of the image in which most of the image information is preserved. This capability offered by local features has been overlooked, despite being relevant in many application scenarios. In this paper, we analyze and discuss the performance of feature-driven Maximally Stable Extremal Regions (MSER) in terms of the coverage of informative image parts (completeness). This type of features results from an MSER extraction on saliency maps in which features related to objects boundaries or even symmetry axes are highlighted. These maps are intended to be suitable domains for MSER detection, allowing this detector to provide a better coverage of informative image parts. Our experimental results, which were based on a large-scale evaluation, show that feature-driven MSER have relatively high completeness values and provide more complete sets than a traditional MSER detection even when sets of similar cardinality are considered.

1. Introduction

Local image feature detection has been a prolific research topic in the fields of computer vision and image analysis, mostly due to the fundamental role it plays in a number of prominent tasks. Local feature detection is often found as the first step of effective algorithms targeted at solving a diversity of problems such as wide-baseline stereo matching, camera calibration, image retrieval, and object recognition. Using a sparse set of locally salient and potentially overlapping image parts – the so-called image features – offers two immediate advantages: (i) the existence of many and possibly redundant patches ensures robustness; (ii) by keeping only informative image parts, a compact image representation is constructed and, subsequently, the amount of data for further processing is reduced. Depending on the application domain, there are other properties that local features should exhibit. For example, for matching tasks, it is fundamental to have repeatable and accurate features. That is, the detector should accurately respond to the “the same” features on two different images of the same scene, regardless of the underlying image transformation. Additionally, features should be distinctive, i.e., the patterns in the immediate surroundings of the features should show a significant degree of variation among themselves. Such property allows local features to be easily distinguished through the use of local descriptors.

Given the importance of repeatability in a wide range of application domains, most studies on local feature detection have been focused on the design of repeatable detectors. Currently, there are various algorithms (e.g., Harris- ,Hessian-Affine [30] or HarrisZ [4]) which are able to detect features with a high repeatability rate even in the presence of severe image transformations, such as viewpoint changes.

The introduction of robust local descriptors (e.g., SIFT [25], SURF [2], or sGLOH [3]) has contributed to set a new paradigm in local feature detection. Besides matching patches on an individual basis, the combination of local descriptors and local features has enabled the construction of robust image representations [39], which is particularly useful to solve problems in which a semantical interpretation is involved, such as the tasks of recognizing objects, classifying scenes, or retrieving semantically equivalent images. Unlike repeatability, the study of robust and compact image representations by means of local features has been overlooked. This could be partially explained by the success of dense sampling-based representations [5,13,36,40] in object and scene recognition, which is a simpler strategy that uses local descriptors densely sampled on a regular grid.

While dense sampling is a well-established and successful strategy for object and scene recognition, there are other application domains, namely emerging ones, that could benefit from robust and simultaneously compact (sparse) image representations

☆ This paper has been recommended for acceptance by Punam Kumar Saha.

* Corresponding author. Tel.: +351 967983202.
E-mail address: pjmm@dei.uc.pt (P. Martins).

http://dx.doi.org/10.1016/j.patrec.2016.01.003

via local features. The development of inexpensive cameras and storage has boosted the creation of significantly large image and video databases. In such databases, search and retrieval mechanisms only make sense if they are efficiently performed [1,26]. The use of local-feature based robust image representations might be a starting point to ensure such efficiency. This scenario emphasizes the importance of analyzing the completeness of features as well as the complementarity between different types of features. Here, completeness means the amount of image information preserved by the local features [10,16], whereas complementarity reflects how different two or more types of features are.

In the large-scale completeness and complementarity test performed by Dickscheid et al. [10], the Maximally Stable Extremal Regions (MSER) [29] showed remarkable overall results: they had significantly higher completeness values compared to other types of local features of similar sparseness. In fact, the MSER completeness values were comparable with the ones of Salient Regions [19], which appear in a substantially higher number.

In this paper, we extend the study on feature-driven Maximally Stable Extremal Regions (fMSER) [27,28], which are a derivation of MSER features and aimed at overcoming some limitations of a regular MSER detection. Our main goal is to analyze the completeness and complementarity of different fMSER. In other words, we are interested in assessing the potential suitability of this type of features for application domains requiring robust image representations via the use of local features. As a result, we present a large-scale evaluation test in which fMSER and MSER are studied in terms of completeness and complementarity.

The remainder of this paper is organized as follows. Section 2 introduces definitions and notations that will be followed throughout this document and presents the motivation behind the construction of feature-driven MSER. Section 3 covers the derivation of several instances of fMSER features from standard MSER. An evaluation of the completeness and complementarity of feature-driven MSER is presented in Section 4. Finally, conclusions and perspectives are given in Section 5.

2. Background and motivation

Boundary-related semi-local structures such as edges and curvilinear shapes are known for being more robust to intensity, color, and pose variations than typical interest points (e.g., corner points). Some local feature detectors explicitly or implicitly take advantage of this robustness by detecting stable regions from semi-local structures, such as the Edge-based Regions (EBR) detector [37,38], which is based on edge detection, or the Principal Curvature-Based Regions (PCBR) detector [9], which is based on line detection. These two examples use the detection of boundary-related structures to generate the final regions. The Maximally Stable Extremal Regions (MSER) detector [29] implicitly takes advantage of the robustness of boundary-related structures without detecting them. In fact, the MSER detector is in its essence an intensity-based region detector dealing with connected components and extracting extremal regions that are stable to intensity perturbations.

The use of boundary information in the construction of local features is not only advantageous in terms of robustness. The semantic meaningfulness of boundary information is equally relevant, as it allows local features to capture informative image parts, which contributes to the construction of intuitive object representations [21].

When the goal is to ensure a robust image representation in an efficient manner via the use of local features, the MSER detector appears as a suitable option. Extremal regions can be enumerated in almost linear time, which makes the MSER detector one of the most efficient solutions for local feature detection. Additionally, the results from the large-scale completeness evaluation performed by Dickscheid et al. [10] showed that MSER features provide a relatively robust image representation despite their sparseness.

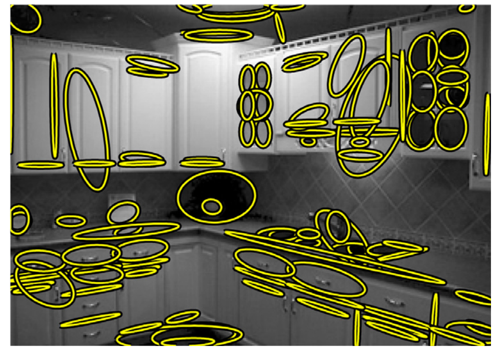

Fig. 1. An example of standard MSER detection. Either the boundaries of objects or other contours are responsible for delineating MSER features. For a better visualization, the original MSER were replaced by fitting ellipses.

Feature-driven MSER [28] were initially proposed as an attempt to overcome the typical shortcomings of a standard MSER detection, namely the lack of robustness to blur, the reduced number of regions, and the biased preference towards round shapes [20]. By detecting more regions on informative parts of the image, feature-driven MSER represent an improvement over standard MSER in terms of completeness.

2.1. Maximally stable extremal regions

Affine covariant regions can be derived from extremal regions. In the image domain, an extremal region corresponds to a connected component whose corresponding pixels have either higher or lower intensity than all the pixels on its boundary. Extremal regions hold two important properties: the set of extremal regions is closed under continuous transformations of image coordinates as well as monotonic transformations of image intensities. The Maximally Stable Extremal Regions detector responds to extremal regions that are stable with respect to intensity perturbations (see Fig. 1). For a better understanding of the MSER detector, we introduce the formal definitions of connected component and extremal regions [33].

A connected component (or region) $\mathcal{Q}$ in $\mathcal{D}$ is a subset of $\mathcal{D}$ for which each pair of pixels $(\mathbf{p}, \mathbf{q}) \in \mathcal{Q}^2$ is connected by a path in $\mathcal{Q}$, i.e., there is a sequence $\mathbf{p}, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m, \mathbf{q} \in \mathcal{Q}$ such that $\mathbf{p} \sim \mathbf{a}_1, \mathbf{a}_1 \sim \mathbf{a}_2, \ldots, \mathbf{a}_m \sim \mathbf{q}$, where $\sim$ is the equivalence relation defined by $(\mathbf{p} \sim \mathbf{q}) \iff \max\{|p_1 - q_1|, |p_2 - q_2|\} \leq 1$ (8-neighborhood).

We define the boundary of a region $\mathcal{Q}$ as the set $\partial \mathcal{Q} = \{\mathbf{p} \in \mathcal{D} \backslash \mathcal{Q} : \exists \mathbf{q} \in \mathcal{Q} : \ \mathbf{p} \sim \mathbf{q}\}$. A connected component $\mathcal{Q}$ in $\mathcal{D}$ is an extremal region if $\forall \mathbf{p} \in \mathcal{Q}, \mathbf{q} \in \partial \mathcal{Q} : I(\mathbf{p}) < I(\mathbf{q})$ or $I(\mathbf{p}) > I(\mathbf{q})$.

Let $\mathcal{Q}_1, \ \mathcal{Q}_2, \ldots, \ \mathcal{Q}_{i-1}, \ \mathcal{Q}_i, \ \mathcal{Q}_{i+1} \ldots$ be a sequence of extremal regions such that $\mathcal{Q}_k \subset \mathcal{Q}_{k+1}, \ \ k = 1, \ \ 2, \ldots$. We say that $\mathcal{Q}_i$ is a maximally stable extremal region if and only if the stability criterion

$$\rho(k, \Delta) = \frac{|\mathcal{Q}_{k+\Delta} \setminus \mathcal{Q}_k|}{|\mathcal{Q}_k|} \tag{1}$$

attains a local minimum at i, where Δ is a positive integer denoting the stability threshold. As the area ratios are preserved under affine transformations, ρ is invariant with respect to affine transformations. Consequently, MSER features become covariant with this type of geometric transformations.

2.1.1. Advantages

As already mentioned in the introductory section, MSER tend to provide a good coverage of informative image parts, despite