Tree-Based Morse Regions: A Topological Approach to Local Feature Detection

Yongchao Xu, Pascal Monasse, Thierry Géraud, and Laurent Najman*

***Abstract*—This paper introduces a topological approach to local invariant feature detection motivated by Morse theory. We use the critical points of the graph of the intensity image, revealing directly the topology information as initial "interest" points. Critical points are selected from what we call a tree-based shape-space. Specifically, they are selected from both the connected components of the upper level sets of the image (the Max-tree) and those of the lower level sets (the Min-tree). They correspond to specific nodes on those two trees: (1) to the leaves (extrema) and (2) to the nodes having bifurcation (saddle points). We then associate to each critical point the largest region that contains it and is topologically equivalent in its tree. We call such largest regions the Tree-Based Morse Regions (TBMR).**

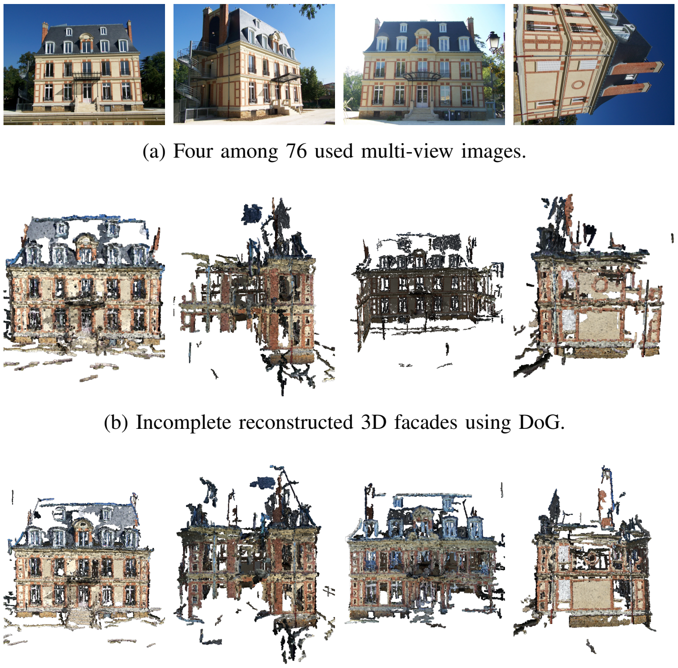
TBMR can be seen as a variant of MSER, which are contrasted regions. Contrarily to MSER, TBMR relies only on topological information and thus fully inherit the invariance properties of the space of shapes (*e.g.*, invariance to affine contrast changes and covariance to continuous transformations). In particular, TBMR extracts the regions independently of the contrast, which makes it truly contrast invariant. Furthermore, it is quasi parameter-free. TBMR extraction is fast, having the same complexity as MSER. Experimentally, TBMR achieves a repeatability on par with state-of-the-art methods, but obtains a significantly higher number of features. Both the accuracy and the robustness of TBMR are demonstrated by applications to image registration and 3D reconstruction.

***Index Terms*—Min/Max tree, local features, affine region detectors, image registration, 3D reconstruction.**

I. Introduction

LOCAL invariant feature detection [1, 2, 3, 4, 5, 6, 7, 8] is an important step in a number of applications such as wide baseline matching, object and image retrieval, tracking, recognition, image registration and 3D reconstruction. The classical process to obtain the features consists in detecting a specific class of interest points, such as corners, together with an associated scale generally obtained from a scale-space. Typical examples of such key locations are the local extrema of the result of difference of Gaussians (DoG) applied in scale-space to a series of smoothed and resampled images. Several crucial invariance properties are required for using such points in applications, such as invariance to image translation, scaling, and rotation, to illumination changes or to local geometric distortion.

In this paper, we propose a topological approach to extract the local invariant features. We first extract some initial "critical" points, based on ideas from the Morse theory [11]: minima, maxima, and saddle points. More precisely, following [12], we propose to choose critical regions in the two trees (called Min-tree and Max-tree [13]) made by the connected components of lower and upper level sets: those critical regions are the leaves and the regions resulting from a fork. For each critical region a scale is selected. Instead of using a scale-space, the scale comes from the tree-based shape space which is built from the image f using the Max-tree and Min-tree: we associate to a critical region R_c the largest region containing it and topologically equivalent in its tree. We call our method Tree-Based Morse Regions (TBMR).


(a) Four among 76 used multi-view images.

(b) Incomplete reconstructed 3D facades using DoG.

(c) The four facades of the 3D reconstruction using TBMR.

Fig. 1. An example of 3D reconstruction using local invariant features. Top: 4 among 76 used multi-view images. Four facades of the PMVS [9] densified sparse 3D reconstruction from the SfM pipeline [10] using DoG (middle: only 62 images are calibrated, and the back facade of the building is missing.) and the proposed TBMR (down: all the 76 images are calibrated, and the whole mansion is reconstructed).

As detailed in Section III-A, the tree-based shape space is invariant to affine contrast changes and to continuous (topological) transformations such as translation, scaling or rotation. As TBMR uses only topological information, TBMR inherit from this shape space, and is thus independent of the image contrast and covariant to these same continuous transformations. As demonstrated in this paper, TBMR is also robust to local geometric distortion. Furthermore, it is essentially parameter-free: only two non-significant parameters are applied, so that we ignore regions that are either too small or too large. Besides, the maximum area parameter is actually

laurent.najman@esiee.fr

not important: large shapes are few and do not influence much the results. And last, but not the least, efficient algorithms with a quasi-linear or a linear complexity are available to compute the TBMR [13, 14, 15].

Qualitative experiments (Section V-A) show the better distribution of TBMR compared to other state-of-the-art methods. Quantitative evaluation, based on the image coverage measurement in Section V-B, confirms the qualitative evaluation. Tests in Section V-C demonstrate that TBMR achieves repeatability score comparable to other state-of-the-art methods with a significantly higher number of correspondences. We evaluate TBMR on two applications in which many matched features are required: image registration (Section V-D) and 3D reconstruction (Section V-E). For these two applications, and as illustrated in Fig. 1, results attest that TBMR improves over the commonly used DoG [16].

II. Related work

There exist a variety of local invariant feature detectors having relatively good performance, as assessed by several evaluation frameworks [5, 6, 7, 8]. The first type is based on scale-space. Harris corners, Hessian based detectors, and the Difference of Gaussians (DoG) are such instances. The Harris corner detector [17] finds the extrema of a corner measure based on the second moment matrix at some fixed scale. A scale-adapted Harris corner detector and its extension Harris-Laplace [18] with scale selection find extrema of the Laplacian of Gaussian (LoG) filter. The Hessian detector [3] extracts the extrema of a feature measure based on the Hessian matrix. Its extension Hessian-Laplace [18] uses the same scale selection as Harris-Laplace. The affine versions of both Harris and Hessian are based on the affine shape estimation using the second moment matrix. Harris based detectors tend to extract corner-like structures, while Hessian based detectors tend to find blobs and ridges. DoG [16] is similar to the Hessian detector in the sense that it approximates LoG by the trace of the Hessian matrix. DoG tends to extract points at isotropic blob structures.

Some recent feature detectors also based on scale space are FAST [19, 20] and its variants: AGAST [21], ORB [22], BRISK [23], and FREAK [24]. The FAST method in [19, 20] proposed an efficient approach for corner detection, based on the comparison of pixel values on a ring centered at a feature point. The AGAST [21] detector is an improved version for accelerated performance of FAST. The recent ORB [22] detector is a rotation-invariant extension of FAST. The detector used in BRISK [23] is a multi-scale AGAST. It searches for maxima in scale-space using the FAST score as a measure of saliency, and estimates the scale of each keypoint in the continuous scale-space. The FREAK [24] method uses the same detector as the one used in BRISK.

The second type of feature detectors is based on tree-based shape space. Although its original definition is quite different, the Maximally Stable Extremal Regions (MSER) [1] are easily understandable using Min-tree and Max-Tree. Indeed, as shown in [25], the MSER algorithm extracts the regions (nodes) that correspond to local minima of a stability function along the path to the root of the tree. The stability function $\mathcal{A}_q$ of a given node $\mathcal{N}$ is given by the difference between the area of some (grand-)parent $\mathcal{N}_\Delta^+$ and some (grand-)child $\mathcal{N}_\Delta^-$, divided by the area of the node itself. It is given by:

$$\mathcal{A}_q(\mathcal{N}) = (|\mathcal{N}_\Delta^+| - |\mathcal{N}_\Delta^-|)/|\mathcal{N}|, \tag{1}$$

where $|\cdot|$ denotes the cardinality, $\mathcal{N}_\Delta^+$ and $\mathcal{N}_\Delta^-$ are respectively the lowest ancestor and the highest descendant such that $|f(\mathcal{N}_\Delta^+) - f(\mathcal{N})| \geq \Delta$ and $|f(\mathcal{N}) - f(\mathcal{N}_\Delta^-)| \geq \Delta$, and Δ is a stability range parameter that fixes the intensity level difference. It is reported in [6] that MSER achieves state-of-the-art repeatabilities and regions accuracies. It is also very efficient. However, the number of detected features are comparatively small, which limits its ability for some applications like image registration and 3D reconstruction. Perdoch *et al.* [26] propose the Stable Affine Frame (SAF), for which only local stability is required. Many more features are obtained with a comparable repeatability score. However, it is much slower than MSER.

A complete review of invariant feature detectors is out of scope of this paper. The interested reader is referred to Tuytelaars and Mikolajczyk's survey [27]. Although these detectors are widely used in many applications in the computer-vision community, direct application of these detectors to other fields can be difficult. For example, surgical navigation has shown significant difficulties due to the free-form tissue deformation and changes of visual appearance of surgical scenes, and alternate solutions have been proposed [28].

III. Tree-Based Morse Regions

In this section, we describe our proposed topology-based local invariant features detector called the *Tree-Based Morse Regions (TBMR)*. The TBMRs are extracted from the shape space built from the image using the Max-tree $\mathcal{T}_M$ and Min-tree $\mathcal{T}_m$. In Section III-A, we briefly review the Morse theory [11, 12] and the original notion of space of shapes. In Section III-B, we show how to extract "interest" regions from these shape spaces. The algorithm of TBMRs extraction based on component trees is described in Section III-C.

A. Tree-based Morse Theory and Shape Spaces

The aim of Morse theory is to describe the topological changes of the (iso)level sets of a real-valued function in terms of its critical points. Recall that a Morse function is a smooth function f whose critical points (*i.e.*, points where $\nabla f = 0$) are isolated. Critical points are minima, maxima, and saddle points of f. The topology of f is directly linked to the analysis of those critical points.

The use of Morse theory is not new in computer vision: see, *e.g.* the use of the contour tree [29, 30] and the Reeb graph [31, 32] for shape matching. However, the Morse function is not an adequate model for an image, as it prevents the existence of plateaux for example. A consequence is that we want to deal with *regional* extrema and saddle *regions* instead of isolated points. The chapter 4 of [12] describes several notions of critical values and prove that they are equivalent to the critical points of Morse theory. In $\mathbb{R}^2$, the Morse structure of an image is extracted from its upper and

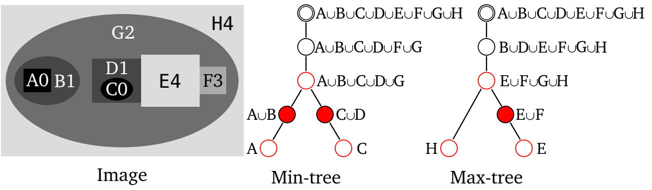


Fig. 2. A synthetic image and the corresponding Min-tree (middle) and Max-tree (right) representation. The critical regions are represented by red circles: (1) nodes having more than on child, and (2) leave-nodes. The filled regions are the TBMRs.

lower level sets, and is equivalent to the notion of critical value used in Morse theory.

For any $\lambda \in \mathbb{R}$, the upper level sets $\mathcal{X}_\lambda$ and lower level sets $\mathcal{X}^\lambda$ of an image $f : \mathbb{R}^2 \to \mathbb{R}$ are defined respectively by $\mathcal{X}_\lambda(f) = \{p \in \mathbb{R}^2 \mid f(p) \geq \lambda\}$ and $\mathcal{X}^\lambda(f) = \{p \in \mathbb{R}^2 \mid f(p) \leq \lambda\}$. Thanks to the inclusion relationship, the connected components of upper (*resp.* lower) level sets can be organized into a tree structure, that is called the Max-tree $\mathcal{T}_M$ (*resp.* Min-tree $\mathcal{T}_m$).

Let us intuitively describe the building of the Max-tree along with the topological changes in the level sets. Imagine that the surface is completely covered by water, and that the level of water slowly decreases. Islands (regional maxima) appear. These islands form the leafs of the Max-tree. As the level of water decreases further, islands grow, building the branches of the Max-tree. Sometimes, at a given level, several islands merge into one connected piece. Such pieces are the forks of the Max-tree, *i.e.*, the nodes of the tree with several children. We stop when all the water has disappeared. The emerged area forms a unique component: the root of the tree representing the whole image.

Following this intuitive descrition, the critical regions we consider in this paper are the extrema of f, corresponding to the leaves of the Max-tree $\mathcal{T}_M$ and of the Min-tree $\mathcal{T}_m$ of f, and the saddle points of f, corresponding to the nodes of these trees having several children.

According to Morse theory, critical points provide essential geometric information. This would lead us to consider extrema and saddle points. Indeed, these are invariant under any increasing contrast change. Unfortunately, critical points are difficult to use directly: (i) many of them are due to noise, and (ii) there is a crucial absence of information about their scale. Typically, most extrema are single pixels, which do not provide a clue about an adapted neighborhood to use as descriptor. Notice that applying a grain filter to the image does not solve the problem. Most regional extrema will then have the size of the grain filter, which is an artifact of the method. For this reason, and as we detail in the next section III-B, we are led to use the largest regions containing the critical region without a topological change. In other words, we are associating scales to critical regions, corresponding to the components containing them just before their merge. However, in this paper, the exact positions of the critical points are ignored, since at the end the centroids of the components are used as feature points.

An example of a synthetic image, together with its Max-tree and its Min-tree, is shown in Fig. 2. In this figure, critical regions are highlighted with a red circle.

Such types of tree-based image representations feature several interesting properties. Specifically, they are invariant to affine contrast changes and covariant to continuous (topological) transformations. Furthermore, efficient algorithms with linear or quasi-linear complexity are available [13, 14, 15].

As described in section II, many feature detection methods are based on a scale-space. The *causality principle* is certainly the most fundamental principle of multi-scale analysis [33]. From this principle, for any couple of scales $\lambda_2 > \lambda_1$, the "structures" found at scale λ_2 should find a "cause" at scale λ_1. Both $\mathcal{T}_M$ and $\mathcal{T}_m$ share such a property: indeed, an extremum that appears at a given level λ_1 gives rise to a whole branch of the tree, which corresponds to a region from very fine to the whole image. Thus, both $\mathcal{T}_M$ and $\mathcal{T}_m$ can be seen as a multi-scale image representation, that we call a shape-space. Such shape-spaces have another interesting property: contrary to scale-space, the contours of a given shape (connected component) correspond to actual contours in the image, without any "blurring" due to convolution with a kernel. These properties of the tree-based image representations make the shape space very appropriate for local invariant features detection. Indeed, these interesting properties of the shape-space made from the connected components of both upper and lower level sets, coupled with an attractive criterion for regions selection (see Section II), are the reasons of the success of MSER. However, the MSER criterion, based on Eq. (1), is not invariant to contrast changes (contrarily to the TBMR topological criterion detailed in the next section).

B. *Feature extraction based on Morse theory and shape space*

The next stage is to associate a scale to each critical region. A critical region corresponds to a change of topology in its tree: either an apparition (leaves corresponding to extrema) or a merge (region resulting from a fork). Thus, on a branch of the tree between two critical regions, there is no topological change in the tree. In other words, a region that is not critical is topologically equivalent to the first critical region we encounter going from the region to the leaves of the tree. Conversely, a critical region $\mathcal{R}_c$ is topologically equivalent to any region that contains $\mathcal{R}_c$ but no other disjoint critical region. As we want as much context as reasonable to encode the region, a good "scale" choice for representing a critical region is the largest region to which it is topologically equivalent in its tree. We call such a region a *Tree-based Morse Region* (TBMR).

In practice, we do not consider the TBMRs that are either too small or too big. Discarding small regions is performed before analyzing the tree structure, which means that they do not contribute to the topological changes of the tree structure. Discarding them eliminates also some noise without modifying other components. In our experiments, we always set this lower bound at 30 pixels. In addition, regions that meet the image border are considered truncated so we ignore them. In Fig. 2, the TBMRs are drawn with a red disk. The evolution of the number of children of a region, starting from a leaf to the root is illustrated in Fig. 3.

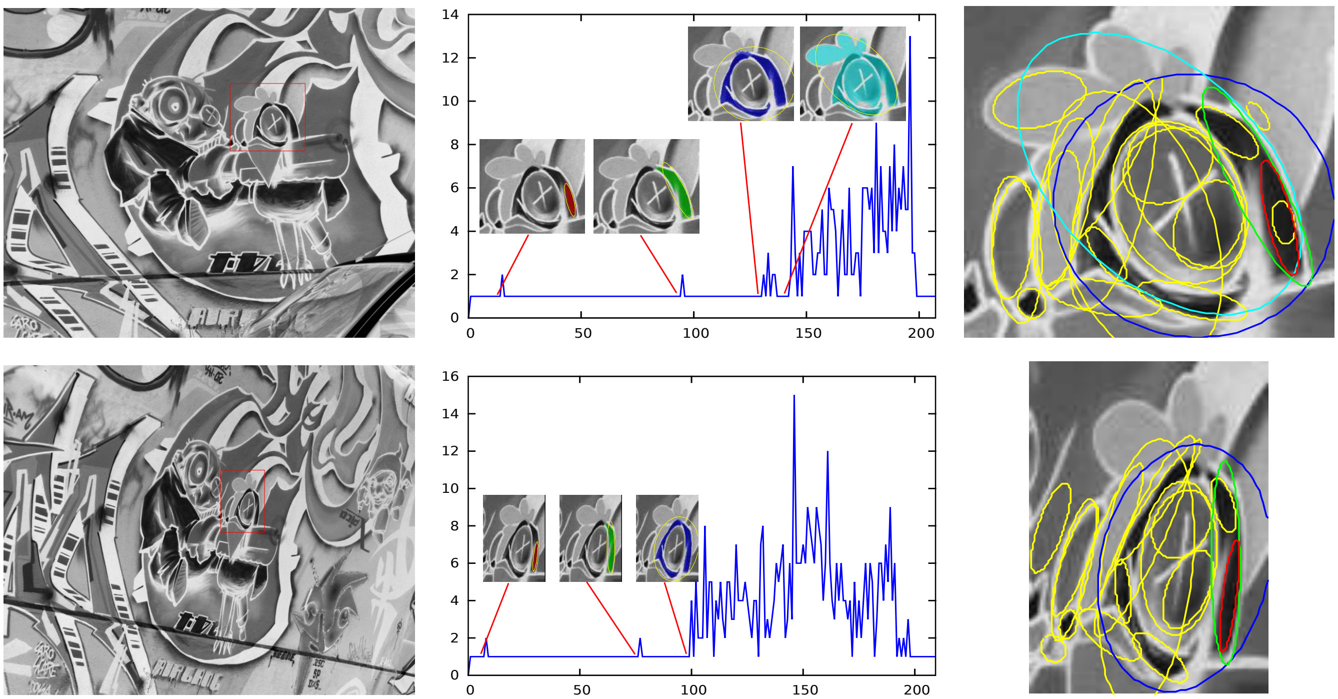


Fig. 3. An example of TBMR extraction. Left: Front view (top) and 30° view (bottom) of “Graffiti” scene [6]. Middle: Evolution of number of children starting from a leaf to the root and some extracted TBMRs (as well as the corresponding ellipses) along the branch; Right: ellipses of the extracted TBMRs inside the corresponding red bounding boxes. Note that only the ellipses of the TBMRs extracted from Min-tree are shown.

As in most shape-space-based methods, we compute the centroids of the selected regions as the final feature points. The ellipse with the same first and second moments as the detected region is then used as the local patch upon which a descriptor is computed.

C. Algorithm of TBMRs extraction

An algorithm that extracts the TBMRs is provided in Algorithm 1. It relies on the computation of the two components trees Max-tree and Min-Tree [13, 14, 34, 15, 35]. The TBMRs are extracted from each one of those two trees independently. We first count the number of children of each node, ignoring the nodes that are too small. Then, the TBMRs are the nodes with a unique child and at least one sibling, removing the nodes that are too big. The corresponding regions provide a (kind of) scale associated to a given critical point, and are approximated by an ellipse.

As shown in Algorithm 1, the extraction of TBMRs is composed of (1) shape space building by Min/Max-tree computation and (2) a linear regions selection from the shape space. There exist many efficient algorithms to compute the Min/Max-tree [35]. The union-find-based approach [14] would take $O(n\alpha(n))$ time for low quantized image (typically 12 bits image or less), where n is the total number of pixels inside an image f and α is a very slow-growing “diagonal inverse” of the Ackermann’s function. Linear algorithms to compute the Min/Max-tree are also available [13, 15]. The selection of TBMRs from the shape would take $O(N)$ time, where N is the total number of nodes of the tree which is no greater than n. In consequence, the complexity of the TBMRs computation algorithm shown in 1 would be $O(n\alpha(n))$ or $O(n)$.

```
TBMR_Extraction($f$, $\lambda_{\min}$, $\lambda_{\max}$)
$\mathcal{T}_M \leftarrow Compute\_MaxTree(f)$
$\mathcal{S} \leftarrow$ TBMR_Tree_Extraction($\mathcal{T}_M$, $\lambda_{\min}$, $\lambda_{\max}$)
$\mathcal{T}_m \leftarrow Compute\_MinTree(f)$
$\mathcal{S} \leftarrow \mathcal{S}\cup$ TBMR_Tree_Extraction($\mathcal{T}_m$, $\lambda_{\min}$, $\lambda_{\max}$)
return $\mathcal{S}$

TBMR_Tree_Extraction($\mathcal{T}$, $\lambda_{\min}$, $\lambda_{\max}$)
$\mathcal{S} \leftarrow \emptyset$
foreach $\mathcal{N}$ in $\mathcal{T}$ do $numChildren(\mathcal{N}) \leftarrow 0$
foreach $\mathcal{N}$ in $\mathcal{T}$ do
    if $area(\mathcal{N}) \geq \lambda_{\min}$ then
        ++$numChildren(parent(\mathcal{N}))$
foreach $\mathcal{N}$ in $\mathcal{T}$ do
    if $area(\mathcal{N}) < \lambda_{\max}$ and $numChildren(\mathcal{N}) = 1$ and
    $numChildren(parent(\mathcal{N})) \geq 2$ then
        $\mathcal{S}$.insert($\mathcal{N}$)
return $\mathcal{S}$
```

Algorithm 1: Extraction of TBMRs. Too small regions (area less than $\lambda_{\min}$) do not contribute to topological changes, and too large regions (area greater than $\lambda_{\max}$) are discarded.

IV. Comparison with some related work

We focus on the comparison with two detector classes, those based on scale-space and those based on tree-based shape space, notably MSER.

A. TBMR versus scale-space feature detection

In spirit, TBMR is very similar to those kinds of approaches. TBMR detects critical points (*i.e.*, extrema and saddle points), but does not rely on a scale-space. As described in section III, it uses the space of shapes provided by the Min-tree and Max-tree representations. This space has the main property of scale-space, namely the causality principle [33].

B. TBMR versus MSER

TBMR can be seen as a variant of MSER, since both rely on Min/Max-tree representations. Indeed, TBMRs are the children of merge nodes (bifurcation). However, TBMRs are never MSERs: indeed, at a merge node, the ratio-of-area criterion proposed in [1] is unstable, and thus the MSER algorithms ignore those nodes.

In practice, the most fundamental difference between MSER and TBMR is related to illumination change, a very common effect in natural images that is reported as an unsolved problem in the literature [8]. Indeed, the MSER stability function shown in Eq. (1) depends on a parameter Δ that fixes the intensity level difference of the (grand-)parent and of the (grand-)child actually used for the ratio. That prevents a true invariance of MSER to illumination change. By contrast, TBMR, being purely topological, is truly invariant to affine illumination change. A less fundamental difference concerns the number of parameters of MSER. As TBMR, MSER uses two parameters to remove too large and too small regions. But MSER also requires in the stability function, on top of the parameter Δ we just described, a threshold to remove unstable regions, and another parameter to group together detected regions that are similar in terms of position and size. The latter is very important, as there are usually too many local minima of the ratio-of-area criterion, and some of them correspond actually to the “same” (very similar) stable regions. Especially in the case of using a very small intensity level difference, which does not have much sense with respect to the notion of stability. Grouping those similar regions makes the extracted regions less accurate (See the registration result presented in Table II).

Such additional parameters are not needed in TBMR. A last minor difference is also related to the definition of MSER at bifurcations: As we mention above, the stability function is not clearly defined in the presence of bifurcations, *i.e.*, when a node has more than one child. That raises a difficulty in trying to reproduce some results: for example, there exist two public implementations of MSER, one from VLFeat [36], the other from OpenCV, each one using a similar but different stability function. The topological definition of TBMR allows for a perfect reproducibility, whatever the chosen algorithm implementation.

C. Comparison of time complexities and of running times

We compare the time complexity of TBMR with the similar method MSER, and the widely used DoG detector. Both TBMR and MSER rely on a computation of Min-tree and Max-tree representation whose time complexity is $O(n\alpha(n))$ or $O(n)$. They differ in the extraction of regions from the Min-tree and Max-tree. As described in Section III-C, the TBMR method requires a computation of number of children which has a time complexity of $O(N)$ and a linear region selection whose time complexity is also $O(N)$. The MSER method relies on the stability function given in Eq. (1). For each node, one have to find its (grand-)parent and the (grand-)child with a level difference of Δ; the time complexity of this step is $O(\Delta N)$. The next step of the MSER method is to select the local minima of the stability function, this has a time complexity of $O(N)$. Consequently, TBMR method and MSER method have a similar time complexity which would be $O(n\alpha(n))$ or $O(n)$. The DoG method is mainly composed of Gaussian blurring, difference of Gaussian, scale-space extrema detection. Since there are just a few number of scales, the time complexity of the naive Gaussian blurring is $O(\omega^2 n)$, where ω is the window size of the kernel of Gaussian blurring. Linear algorithm for Gaussian blurring is also available thanks to recursive filtering [37, 38]. The difference of Gaussian and scale-space extrema detection both have a $O(n)$ time complexity. Finally, the time complexity of DoG is $O(\omega^2 n)$ or $O(n)$.

Image	Run time [ms.]		
	DoG	**MSER**	**TBMR**
Graffiti (800 × 640)	526.6	484.7	471.6
Wall (1000 × 700)	1139.1	1122.9	1124.7
Bark (765 × 512)	579.5	292.1	290.9
Trees (1000 × 700)	1145.4	958.2	1030.4
Leuven (900 × 600)	771.7	524.0	519.6

TABLE I. Averaged computation time for the different detectors applied on several images in the dataset of Mikolajczyk [6]. The corresponding image size is also shown after the name. See text for a discussion.

We have implemented the proposed TBMR method using VLFeat [36]. We compare the running time of TBMR with the public implementation of MSER and DoG already available in VLFeat. The comparison is given in Table I. For each image, each detector is executed 100 times, the average time is shown. All the measurements have been taken on a 3.4GHz/8MB cache Intel core i7-2600, 8GB RAM, Debian 7. The code is compiled with GCC 4.7.2. Note that for the TBMR and MSER methods, the time for region selection is negligible compared to that of the computation of Min-tree and Max-tree. So the two methods have a very similar running time. However, in this public implementation, the accumulated moments for computing the approximated ellipse of each TBMR or MSER region in the Min-tree or Max-tree are computed after the tree construction step. In a more efficient implementation, these moments could have been computed incrementally during the tree construction step. As it is now, the ellipse fitting step takes about 8% of the time of the whole process, depending on the number of moments to compute. TBMR is slower than MSER for the images of *Wall* and *Trees*: indeed, TBMR extracts many more regions than MSER for these two images, hence the ellipse fitting step takes more time. Both TBMR and MSER are faster than DoG. Note that many optimizations are possible for MSER and TBMR. As shown in [39], a more efficient implementation of MSER is $5.6\times$ faster than DoG. An optimized TBMR should have a similar performance.

V. RESULTS

Qualitative and quantitative comparison of the distribution of TBMR with other popular local feature detectors are illustrated in Section V-A and Section V-B. In section V-C, the repeatability assessment of the TBMR is evaluated using the framework of Mikolajczyk *et al.* [6]. Two applications using local invariant features are presented in sections V-D and V-E to compare TBMR with other widely used detectors. The application to image registration in Section V-D highlights the accuracy and the robustness of TBMR. The experiments are conducted on the Stanford Mobile Visual Search (SMVS) Data Set [40]. In Section V-E, the application to 3D reconstruction using structure from motion is first tested on the dataset of Strecha *et al.* [41], providing the ground truth of the camera positions. The baseline error and angular error measurements reveal the accuracy of TBMR. Then the 3D reconstruction experiments are conducted on sets of images taken in a sunny day around some structure. The structure from motion succeeds in reconstructing a complete 3D model using TBMR, whereas only part of the scenes are reconstructed in 3D model when using other detectors.

In all experiments, the parameters of the corresponding method are set with the recommended values found in the literature. More specifically, for Harris-Affine, Hessian-Affine and MSER, we use the Linux binaries[1] given in [6]. For the Harris-Affine detector and the Hessian-Affine detector, the only parameter is the cornerness threshold which is set to 1000 for Harris-Affine, and to 500 for Hessian-Affine. For the MSER method, the parameter Δ in Eq. (1) is set to 10, and the minimum area and maximum area are set to respectively 30 pixels and 1% of the total number of pixels inside the image (we use the same area parameters setting for TBMR). For the DoG method, its public implementation in VLFeat is used. There are five main parameters: the number of octaves is the greatest possible one (*i.e.* roughly $\log_2(\min(\text{width}, \text{height}))$). Each octave is sampled at 3 intermediate scales. The edge threshold is set to 10, which eliminates peaks of the DoG scale space whose curvature is smaller than 10. The peak threshold which filters peaks of the DoG scale space that are too small is set to $255 \times 0.04/3$. In the following, except the explicit "DoG octave 0" whose first octave index is set to 0, the starting octave of DoG method is set to -1. For all the feature matching used in the experiments of image registration in Section V-D and 3D reconstruction in Section V-E, the SIFT descriptors are used. Finally, in all the experiments, the distance ratio of vector angles from the nearest to second nearest neighbor used in the SIFT descriptors matching is set to 0.6.

A. Qualitative features comparison

In this section, we focus on the distribution of the keypoints. This distribution plays an important role for many applications, such as image registration and 3D reconstruction using structure from motion, for which many points covering the object of interest are required.

We compare TBMR with some state-of-the-art local feature detectors (Harris-Affine, Hessian-Affine, DoG octave 0, DoG, and MSER) by visualizing the distribution of the keypoints obtained with each method. In this paper, DoG denotes always DoG starting from octave -1.

For the shape-space-based MSER and TBMR methods, the centroid of the extracted regions is considered as the detected keypoints. The qualitative comparison is conducted on two images taken against the sunlight in a sunny day, and the keypoints distributions are illustrated respectively in Fig. 4 and 5. We can observe that MSER detects few points, which explains the defects in the 3D reconstruction based on MSER in the example shown in Fig. 1. Harris-Affine, Hessian-Affine, and DoG octave 0 extract a reasonable number of points, but only a few of them are located on the object of interest of the scene; that makes them fail in reconstructing the 3D structure in Fig. 1. Using the default option "octave -1" (image of doubled dimensions) for the DoG computation significantly increases the number of detected points. Yet, the additional points are mostly distributed where there were already many points, and not on the object of interest. By contrast, TBMR has a reasonable number of keypoints and they are distributed more uniformly over the whole image. That contributes to its success in obtaining a correct 3D reconstruction; see Fig. 1.

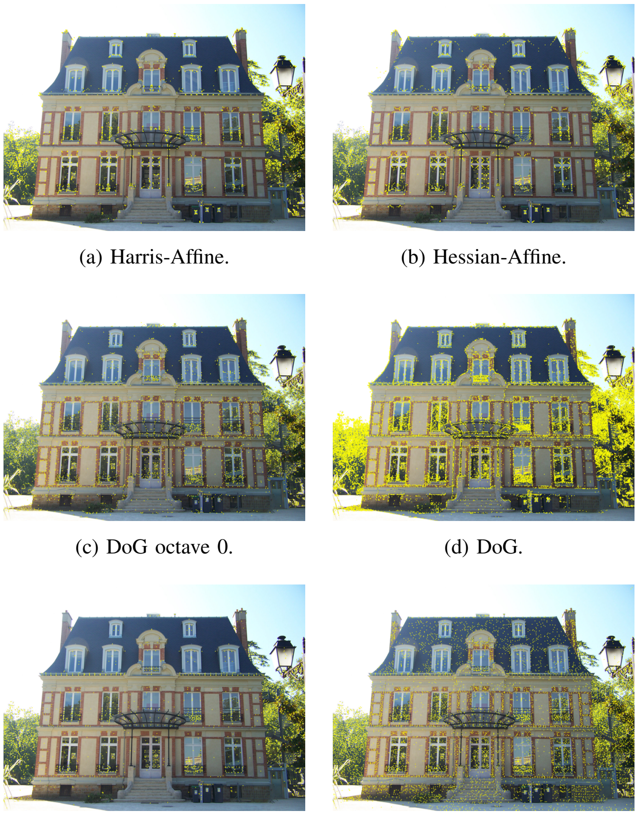
In Fig. 6, we show the qualitative comparison of TBMR with MSER and DoG in the case of a significant change of contrast. MSER and DoG detects very few points in the area with low contrast. By increasing the contrast, MSER and DoG detects some points. On the contrary, TBMRs are perfectly insensitive to contrast change, up to quantization effects. This better performance of TBMR with respect to change of contrast (together with TBMR robustness to viewpoint change, see section V-D below) is one of the main reason for its success in 3D reconstruction; see Fig. 1, Fig. 17, and Fig. 18.

B. Image coverage evaluation

In order to assess the uniformity of those keypoint distributions obtained with different methods, we first measure the distribution of keypoints along the two image dimensions, as well as the number of extracted points for a set of images taken around some scene objects. In Fig. 7, we show the distribution of keypoints position along the horizontal dimension for the images taken around the objects of scene presented respectively in Fig. 5, Fig. 17, and Fig. 18. These images are taken to make the object of interest presented in the middle (horizontally) of the scene. The distributions shown in Fig. 7 are smoothed by taking the average inside a horizontal window (size is set to 21). The total number of keypoints for each coordinate axes can be obtained by multiplying the distribution score with the average number of keypoints shown in Fig. 7. In order to better visualize the comparison between different methods, only DoG, MSER, and TBMR are shown in this figure. MSER extracts few points; TBMR extracts more points than MSER; DoG has many more points. However, TBMR has the largest part of keypoints that cover the objects of interest in those scenes, which is also one of the reason for its success in Fig. 17 and 18.

We also evaluate how well the keypoints cover the image. First, we propose to dilate the extracted keypoints by a 2D window centered at each point with a certain size (*e.g.*, 31).

[1] Available on http://www.robots.ox.ac.uk/~vgg/research/affine



(a) Harris-Affine.

(b) Hessian-Affine.

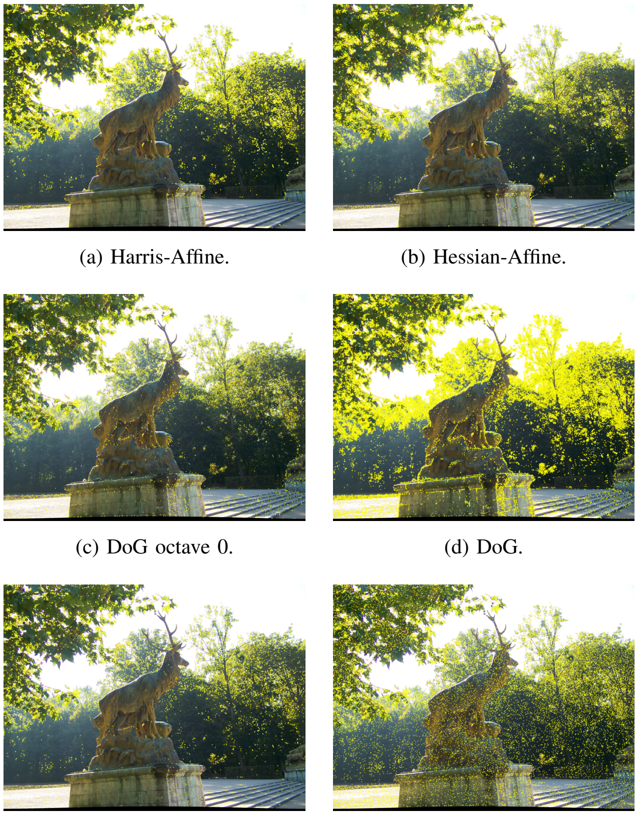
(c) DoG octave 0.

(d) DoG.

(e) MSER.

(f) TBMR.

Fig. 4. Qualitative comparison of TBMR with other widely used local feature detectors applied on an image taken against the sunlight (used in Figure 1). Yellow points in the image are the detected keypoints.



(a) Harris-Affine.

(b) Hessian-Affine.

(c) DoG octave 0.

(d) DoG.

(e) MSER.

(f) TBMR.

Fig. 5. Qualitative comparison of TBMR with other widely used local feature detectors applied on an images taken against the sunlight. Yellow points in the image are the detected keypoints.

Then we compute the rate of area covered by the dilated region. Note that for two close keypoints, their dilated regions may have a large part in common, but the common regions count only once. As shown in Fig. 8, MSER covers a small part of the image because of a few extracted points. TBMR covers the image better than the others having a comparable or much larger number of detected points, which confirms the qualitative observation in Section V-A.

C. Repeatability evaluation

To assess the performance of TBMR, we compare it with some other affine detectors: Harris-Affine and Hessian-Affine, defined on a scale-space, and MSER, defined on the shape space.

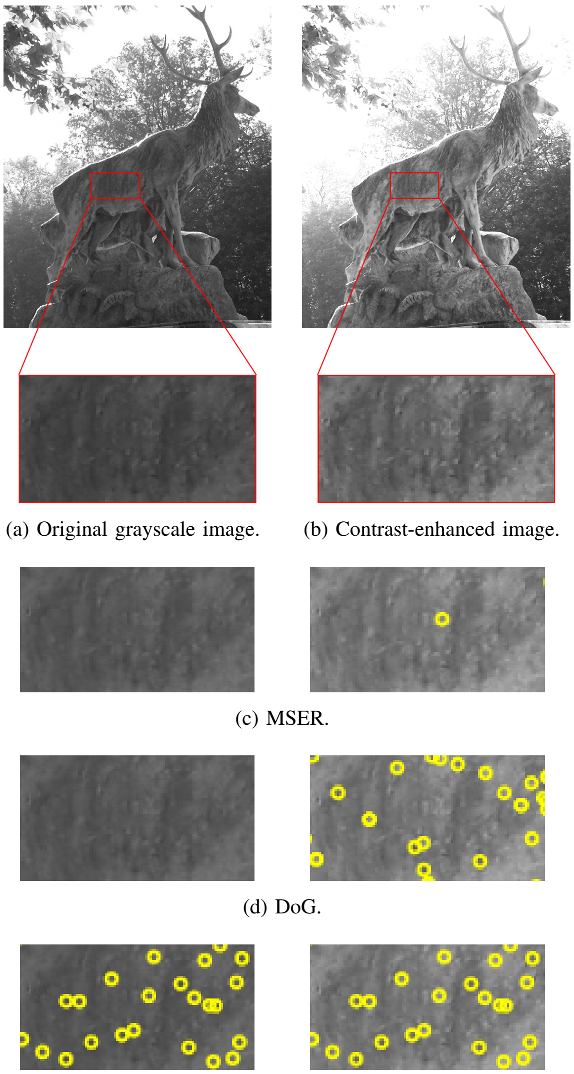
We perform the same tests as in [6], and evaluate the *repeatability score* based on the *overlap error* ϵ:

$$\epsilon(R_{E_1}, R_{E_2}) = 1 - \frac{R_{E_1} \cap R_{H_{21}^T E_2 H_{21}}}{R_{E_1} \cup R_{H_{21}^T E_2 H_{21}}}, \qquad (2)$$

where R_E represents the elliptic region (*i.e.*, local patch of each extracted feature) defined by $x^T E x \leq 1$, and H_{21} is the ground truth homography between the test and reference image. The repeatability score for a pair of images is then defined as the ratio between the number of region-to-region correspondences established under a certain overlap error (*e.g.*, 40% is used in this paper) and the smaller number of regions in the compared images. Another evaluated measurement is the absolute number of correspondences. A high repeatability score and a large number of correspondences are normally desired.

Some results applied to the sequences *Wall* (viewpoint change), *Bark* (scale change), *Trees* (blur), *Leuven* (light change) [6] are illustrated in Fig. 9. Compared to the scale-space-based approaches (*i.e.*, Harris-Affine and Hessian-Affine), the TBMR achieves a competitive repeatability score and a significantly higher number of correspondences, except for the blur sequence *Trees*. The explanation is that the topology of the image is damaged by the blur. For the same reason, the performance is poor on the UBC sequence (not shown here) dedicated for testing robustness to strong JPEG compression artifacts. Such defects (blur, JPEG artifacts) are better handled by the scale-space-based methods. Compared with MSER, which is also based on the shape space, TBMR has a comparable repeatability score, but a significantly higher number of correspondences thanks to the contrast independent property of TBMR. More extensive tests can be found in the supplementary material accompanying this article.

Some extra experiments on other datasets (such as the dataset of DTU [8]) that contains more images, will be considered as a future work.



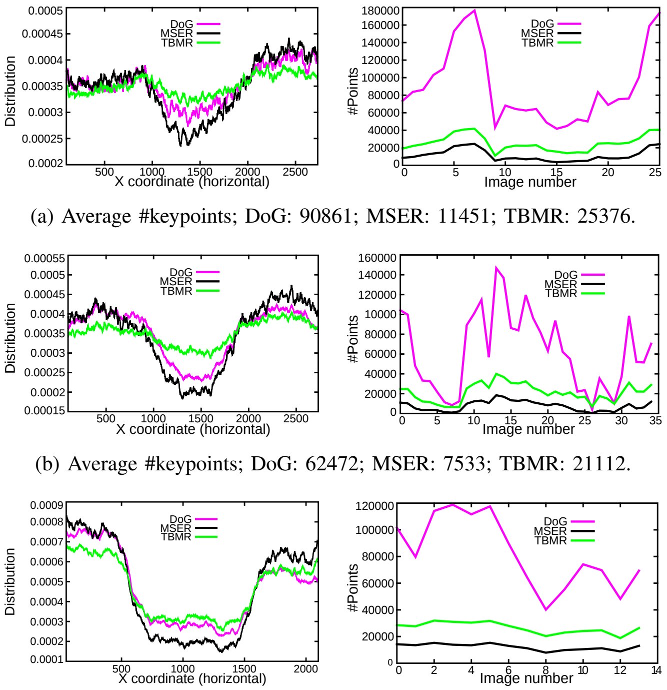
(a) Original grayscale image. (b) Contrast-enhanced image.

(c) MSER.

(d) DoG.

(e) TBMR.

Fig. 6. Qualitative comparison of TBMR with MSER and DoG applied on images with increasing change of contrast. Yellow points in the image are the detected keypoints. Left: Keypoints extracted from the zoomed part of the original image (a); Right: Keypoints extracted from the zoomed part of the contrast-enhanced image (b).



(a) Average #keypoints; DoG: 90861; MSER: 11451; TBMR: 25376.

(b) Average #keypoints; DoG: 62472; MSER: 7533; TBMR: 21112.

(c) Average #keypoints; DoG: 82455; MSER: 12222; TBMR: 26508.

Fig. 7. Horizontal distribution of the keypoints (left) and number of extracted keypoints (right), for the multi-view images taken around the objects of scene presented in respectively Fig. 5, Fig. 17, and Fig. 18 (top to bottom).

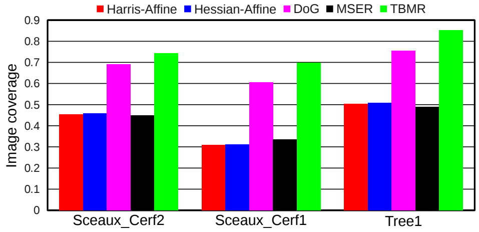


Fig. 8. Image coverage evaluation by dilating a keypoint with a 2D window having size 31×31; Vertical axis represents the average rate of area covered by the dilated region for multi-view images presented in Fig. 5, Fig. 17, and Fig. 18.

D. Image registration

Image registration methods use the local features to establish a correspondence between a number of interest points (*e.g.*, the centroids of the detected elliptical regions) in images. These one-to-one correspondences are then used to estimate the transformation, thereby establishing point-by-point correspondence between the reference image and the target image. Hence, the accuracy and robustness of the local features is crucial to the quality of image registration results. There should also be enough pairs of matched points so that the estimation of the parameters of the transformation between images is possible.

For these experiments, we use the work of Moisan *et al.* [42]. It is based on the Optimized Random Sampling Algorithm (ORSA) proposed by Moisan and Stival [43], a variant of RANSAC algorithm introducing an *a contrario* criterion [44] to avoid to set thresholds for inlier/outlier discrimination. ORSA is used to estimate the homography registering both images. When the homography is assessed, a panorama is built by stitching the images in the coordinate frame of the second image.

1) Quantitative benchmark on dataset of Mikolajczyk [6]: We first benchmark TBMR with MSER, Harris-Affine, Hessian-Affine, and DoG on the public dataset of Mikolajczyk *et al.*, where the ground truth of homography H is available. For each sequence of images in this dataset, the homography between the first image and the others are provided. We have conducted the image registration experiments on each sequence of images in this dataset. The first image is used as reference image, and the other images with different level of transformations (*e.g.* viewpoint change, scale change, etc.) are considered as target images.

For each point p in the reference image (the first image in each sequence) which belongs to the inlier matched points

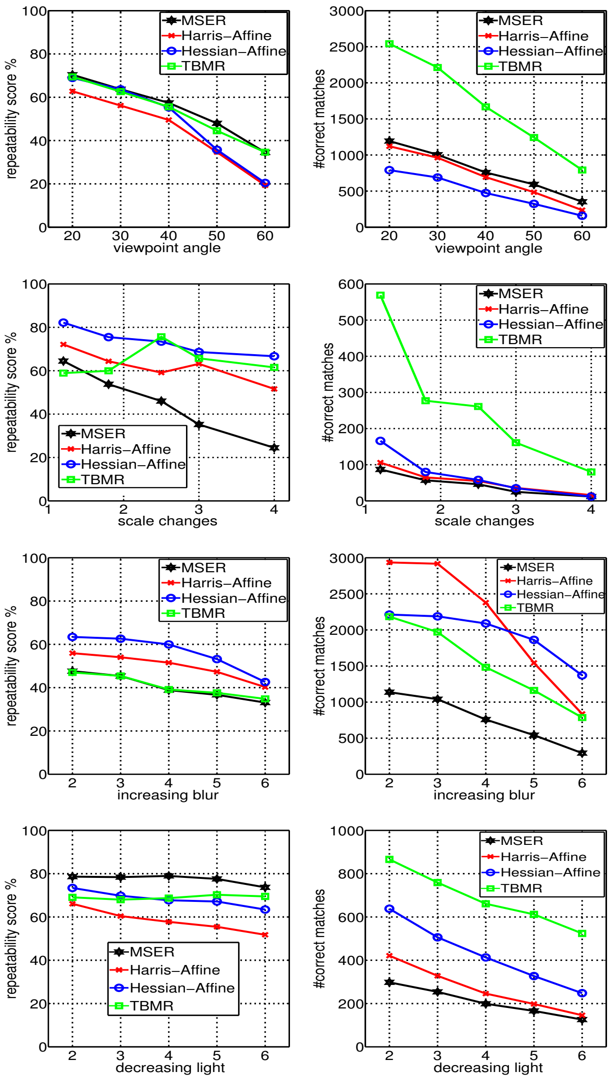


Fig. 9. Repeatability score (left) and number of correspondences (right) for the sequences *Wall, Bark, Trees, Leuven* (top to down).

used to estimate the homography H', we measure the euclidean distance between the points $H \times p$ and $H' \times p$ in the registered image. This distance measures the registration errors: the smaller the distance, the more precise the registration.

Distributions of distances in image registration by homography on "Graffiti" (viewpoint changes) image pairs are shown in Fig. 10 (a - e). The average of distances for each image pair is shown in Fig. 10 (f). In general, TBMR performs better than the others.

2) Qualitative results: The experiments are conducted on the CD-covers of the Stanford Mobile Visual Search (SMVS) Data Set [40]. As the CD-covers are planar scenes, they are amenable to homography registration. We experimented Harris-Affine, Hessian-Affine, MSER, DoG, and the proposed TBMR to compute point correspondences between images using the SIFT descriptors of Lowe [4]. The point correspondences are the input of ORSA.

For the images of Fig. 11 and Fig. 13, Harris-Affine, Hessian-Affine and MSER all fail in estimating the homography due to insufficient number of correspondences. DoG also fails for images in Fig. 13. In Fig. 11, although DoG achieves a homography, the registration is inaccurate at the top left corner, whereas the TBMR results in a meaningful homography in all cases. A chessboard mix of the two registered images for Fig. 11 and Fig. 13 are given respectively in Fig. 12 and in Fig. 14, from which the qualitative results can be better visualized. Note that for these image registration examples, there is no ground truth for these images and finding a relevant metric for such poor quality images is a challenge in itself, so only visual inspection is left to the reader's appreciation.

Method	Canon	Palm
MSER	65	48
MSER2	93	81
Harris-Affine	92	91
Hessian-Affine	97	88
DoG	98	97
TBMR	97	91

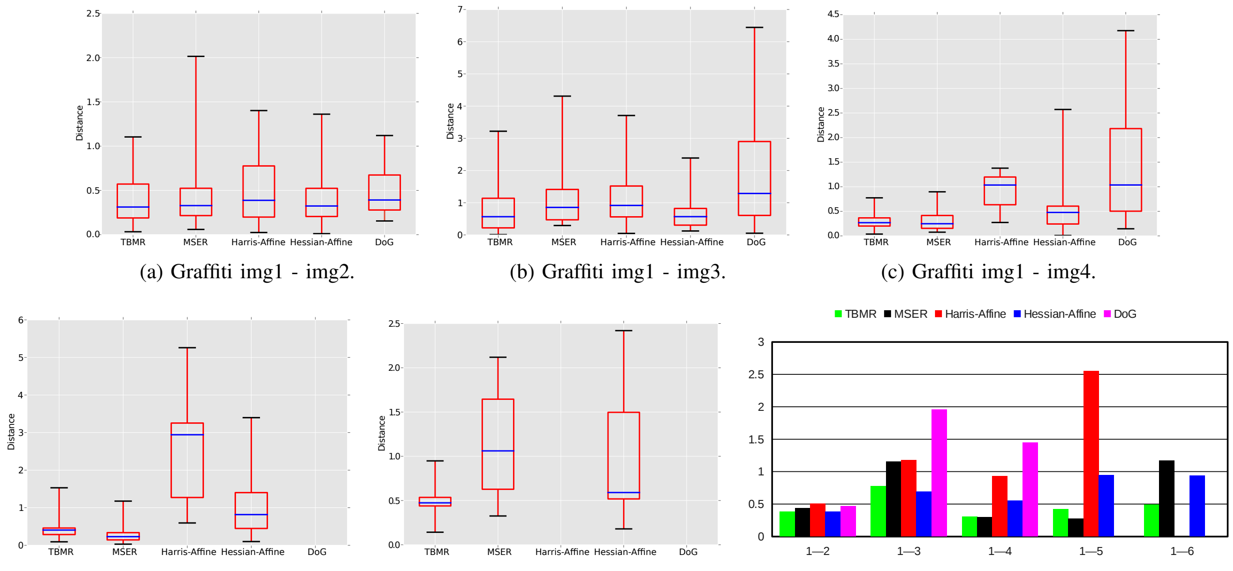
TABLE II. Benchmark of image registration results using different local feature detectors on images taken with "Canon" and "Palm" cameras in SMVS dataset; MSER2 represents the MSER method with the margin value set to 2.

There are 100 different CD-covers in the tested dataset. We have conducted the image registration experiments on the images taken respectively with "Canon" and "Palm" cameras, and the images of reference. The images taken with "Canon" are less blurred than the ones with "Palm". In Table II, we show the number of images having a registration result, obtained using different local feature detectors. MSER with standard parameters (margin value Δ is set to 10) fails for many images; By lowering considerably the margin value ($\Delta = 2$), the performance gets better; In general, TBMR performs better than MSER (with the standard parameters and a very small margin value), Harris-Affine, and Hessian-Affine. TBMR is on par with DoG when the blur is not very important. In addition, among the tested images, we observe that for those images having registration results, the ones using TBMR are more precise in location than those compared methods. But when the blur is important, DoG is better, because topology of the image is damaged by the blur. However, as shown in [45], the multi-resolution detection improves the performance of MSER under blur. We would expect the same improvements by applying a multi-resolution analysis.

E. 3D reconstruction

Structure from Motion (SfM) is a popular process of estimating a three-dimensional structure from a sequence of two-dimensional images. The SfM algorithms take multi-view stereo images along with the internal camera calibration information as input, and yield a sparse 3D point cloud, camera orientations and poses in a common 3D coordinate system. The feature points are used in the phase of model estimations, including homography, fundamental and essential matrices, and camera poses. These estimations are crucial to the quality of 3D reconstruction. Therefore, the accuracy and robustness of the local invariant features detectors is a defining aspect of the 3D reconstruction result.

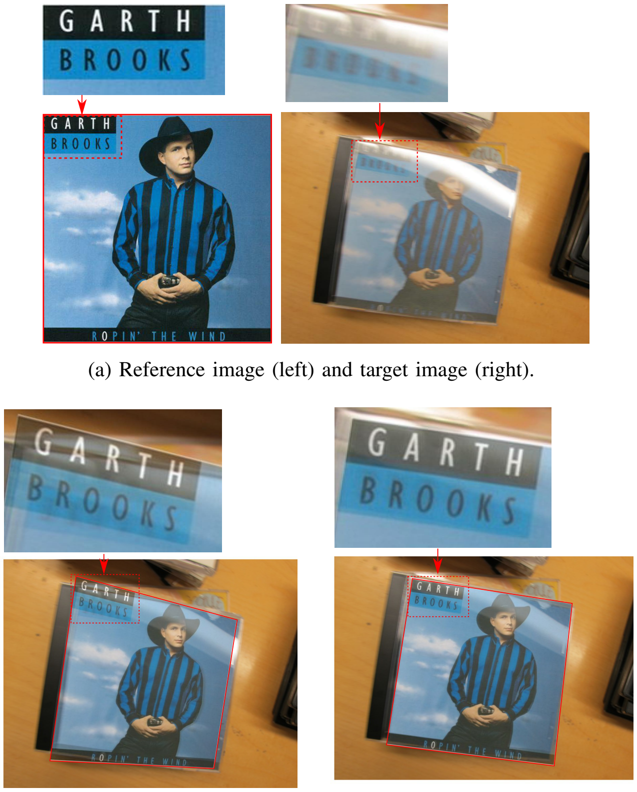
To assess the performance of the feature detectors in this



(a) Graffiti img1 - img2. (b) Graffiti img1 - img3. (c) Graffiti img1 - img4.

(d) Graffiti img1 - img5. (e) Graffiti img1 - img6. (f) Average of all pairs.

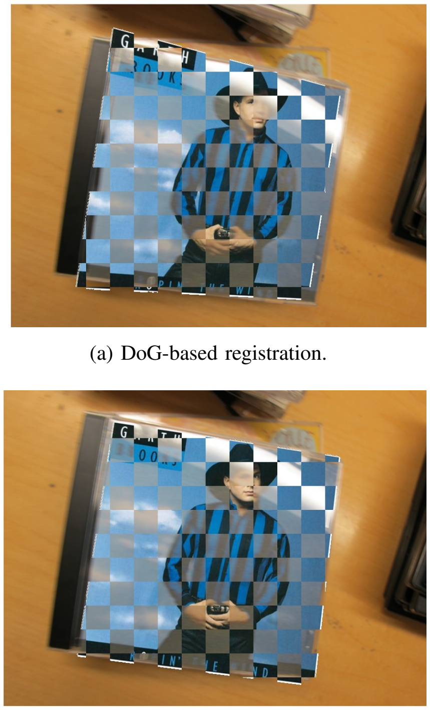
Fig. 10. Distributions of errors in image registration by homography on “Graffiti” image pairs, which comes with a ground truth. DoG fails registering pairs (1,5) and (1,6); Harris-Affine fails on (1,6). There is a slight advantage for MSER on pair (1,5), while on all other pairs TBMR performs better.



(a) Reference image (left) and target image (right).

(b) DoG-based registration. (c) TBMR-based registration.

Fig. 11. Homographic registration of a pair of images. The result obtained using DoG (b) is not as accurate as the one based on TBMR (c); see the zoomed top left corner. Harris-Affine, Hessian-Affine, and MSER all fail in registering this pair of images.



(a) DoG-based registration.

(b) TBMR-based registration.

Fig. 12. Chessboard mix of the two registered images using respectively DoG for (a) and TBMR for (b).

application, we use the software of Moulon *et al.* [10] which relies on an *a contrario* criterion [44] instead of setting thresholds for model estimation in SfM pipelines. In other respects, their pipeline is similar to the one of the popular Bundler software [46]: some initial pair of images are selected and a two-view 3D reconstruction is performed; the other images are then sequentially added, each time refining the 3D scene with bundle adjustment, an iterative optimization method

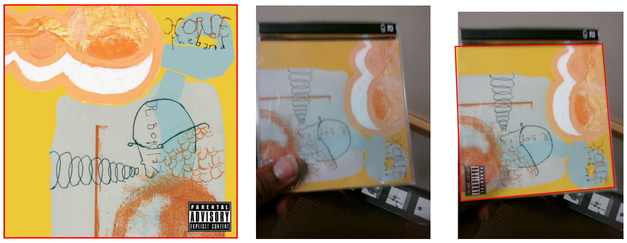


Fig. 13. Homographic registration of a pair of images. Left: reference image. Middle: target image. Right: registration result using TBMR. None other tested detector yields a correct registration.

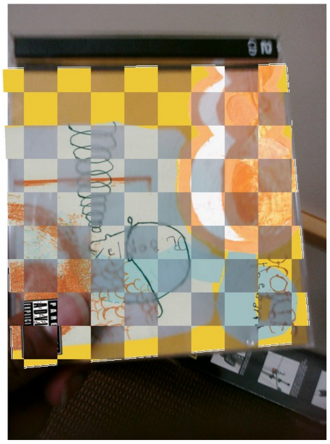
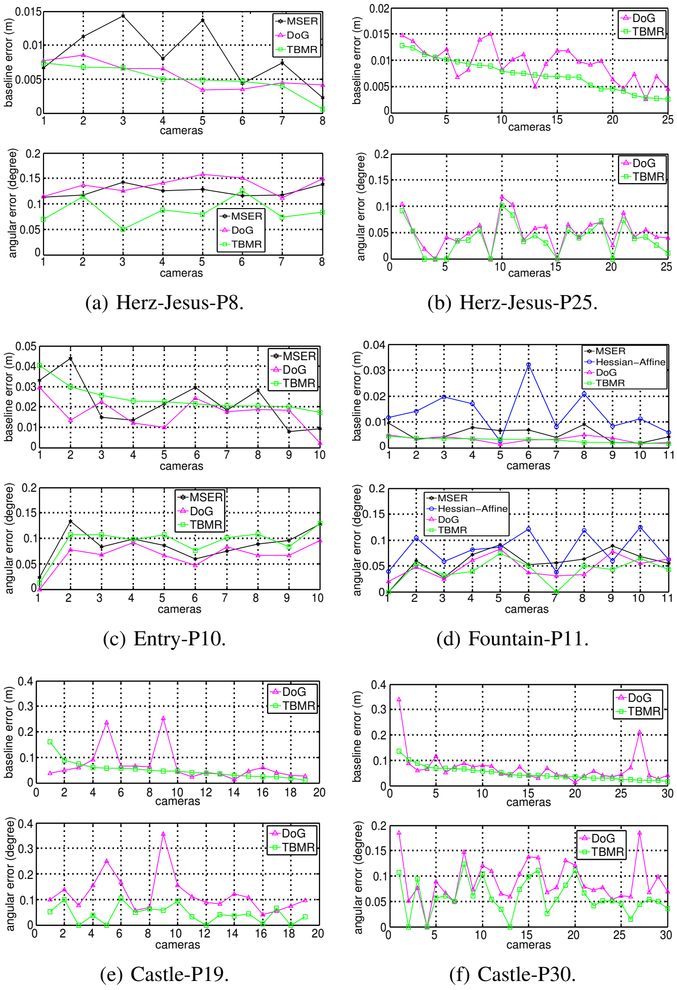


Fig. 14. Chessboard mix of the two registered images using TBMR.

based on Levenberg-Marquardt algorithm. It is shown in their work that the adaptive SfM outperforms the state-of-the-art methods with significant precision improvements. The pipeline may stop prematurely if not enough point correspondences are found when adding extra images.

1) Quantitative benchmark: We first benchmark TBMR with Harris-Affine, Hessian-Affine, MSER, and DoG on the public dataset of Strecha *et al.* [41], where the ground truth of the camera orientations and poses are available. The SIFT descriptors of Lowe [4] are again used to establish for each detector the one-to-one correspondences between the feature points. The quality of the 3D reconstruction is tested in terms of the precision of estimated camera orientations and poses. The baseline errors and angular errors compared to the ground truth are illustrated in Fig. 15. For each sequence, the absence of the curve corresponding to some detectors means that it fails to calibrate all the cameras, or that the baseline and angular errors are too high compared with others. Harris-Affine fails or the measurements are too high for all the sequences. Hessian-Affine fails or the measurements are too high for five sequences, while MSER works for three sequences. DoG and TBMR succeed for all the sequences. Compared to Harris-affine, Hessian-affine, and MSER, TBMR is more robust and behaves better in terms of baseline errors and angular errors. Compared to DoG, in most cases (especially for the sequences of *Herz-Jesus-P8* and *Castle-P19*), TBMR performs better based on the baseline and angular errors. In Fig. 16, we show also the recall rate of good tracks, the ratio between the amount of final maintained tracks, and the number of input tracks found between the feature points. The absolute number of final maintained tracks is presented as well. The highest recall rate and the largest number of 3D points obtained with TBMR also reveal its robustness.



(a) Herz-Jesus-P8. (b) Herz-Jesus-P25.

(c) Entry-P10. (d) Fountain-P11.

(e) Castle-P19. (f) Castle-P30.

Fig. 15. Evaluation of camera calibration based on baseline error and angular error applied on the dataset in [41].

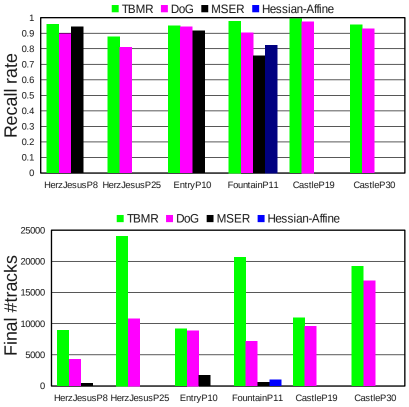


Fig. 16. Recall rate (top) of the tracks and the absolute number (bottom) of final maintained tracks used to yield the sparse 3D points.

2) Qualitative results: We have also tested the SfM method with different detectors on some other set of images taken in a sunny day. Since the SfM produces a sparse 3D point cloud, not a dense 3D reconstruction, the PMVS software [9] is used to densify the 3D points. Note that PMVS relies on the interest points of DoG and on the Harris corners to reconstruct the 3D structures based on the estimated model. Consequently, this may result in a lack of 3D points in a region (*e.g.* the surface near the bottom of the 3D structures on the right side of Fig. 18 using the TBMR). By integrating the TBMR in PMVS, we would expect the reconstructed 3D structures to be denser thanks to its contrast independent property. In Fig. 17, the DoG detector fails to reconstruct the half of the scene where shadows are present. In Fig. 18, the DoG detector behaves similarly, whereas TBMR works well in all cases: the complete 3D structures are reconstructed. Note that Harris-Affine, Hessian-Affine, and MSER perform even worse than DoG, so the corresponding results are not presented in Fig. 17 and 18. Those examples also confirm the importance of the invariance to illumination changes, as pointed out by Aanæs *et al.* [8]. The resulting 3D reconstruction can be better appreciated in the video in the supplementary material accompanying this article.

VI. Conclusion

We have introduced a topological approach to local feature detection motivated by Morse theory. It relies on the critical points (*i.e.*, minima, maxima, saddle points) of the image and on the shape space given by the Max-tree and the Min-tree built from the image. More precisely, we use the critical regions that are the leaves and nodes with bifurcation in the Max- and Min-trees. To each critical region, we propose to associate the largest region from the shape space that contains it but does not contain any disjoint critical region. We have shown that the proposed method, called TBMR, is truly contrast independent and almost parameter-free. Besides, TBMR is fast to compute with a linear or quasi-linear complexity.

Experimentally, we showed on standard data [6] that the developed TBMR achieves a reasonable repeatability and a significantly higher number of features extracted in images. We have also conducted some experiments on two applications relying on local features using public data sets. The homographic registration results and 3D reconstruction results demonstrated the accuracy and robustness of TBMR compared to other state-of-the-art detectors. Its contrast-invariance property and its robustness to viewpoint change make TBMR a method of choice in numerous practical situations.

In the future, we plan to explore the choice of the associated region from the shape space for each critical point. In the definition of TBMR, instead of extracting the largest region, we can imagine selecting the most meaningful region based on some significance measure (*e.g.* the average of gradient's magnitude). Yet such a measure should be designed to be invariant to illumination changes and affine transformations. We also would like to introduce an option to control the number of extracted TBMRs through the topological persistence [47]. Another future work is to assess the usefulness of TBMR for applications such as object recognition and tracking.

Supplementary material is available on http://laurentnajman.org/index.php?page=tbmr. It contains some additional experimental results, a video, and Linux executables. Source code will be available when the paper is published.

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. of British Machine Vision Conference*, 2002, pp. 384–393.

[2] T. Tuytelaars and L. V. Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[3] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 79–116, 1998.

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[5] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[6] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, Nov. 2005.

[7] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[8] H. Aanæs, A. L. Dahl, and K. Steenstrup Pedersen, "Interesting interest points," *International Journal of Computer Vision*, vol. 97, no. 1, pp. 18–35, 2012.

[9] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.

[10] P. Moulon, P. Monasse, and R. Marlet, "Adaptive structure from motion with a contrario model estimations," in *Proc. of Asian Conference on Computer Vision*, 2012.

[11] J. Milnor, *Morse Theory*, ser. Annals of Mathematics Studies. Princeton University Press, 1963, vol. 51.

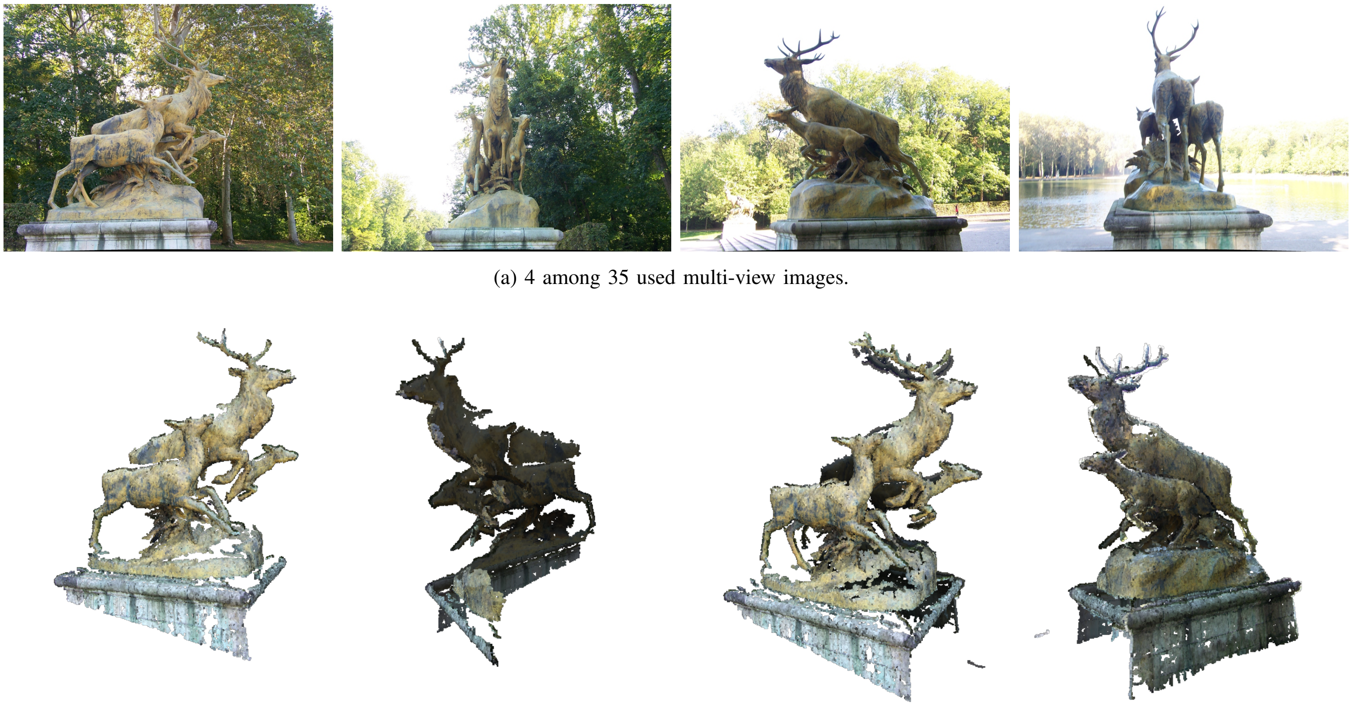
[12] V. Caselles and P. Monasse, *Geometric Description of Images as Topographic Maps*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[13] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.

[14] L. Najman and M. Couprie, "Building the component tree in quasi-linear time," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3531–3539, 2006.

[15] D. Nistér and H. Stewénius, "Linear time maximally stable extremal regions," in *Proc. of European Conference on Computer Vision*, 2008, pp. 183–196.

[16] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. of IEEE International Conference on Computer Vision*, 1999, pp. 1150–1157.



(a) 4 among 35 used multi-view images.

(b) 3D result using DoG, 17 images are calibrated.

(c) 3D result using TBMR, 27 images are calibrated.

Fig. 17. Densified 3D reconstruction. (a) Some input images. (b) and (c) show the left and right side of the 3D structures reconstructed using respectively DoG and TBMR. The use of DoG misses the right side of the scene; the use of TBMR results in a correct 3D reconstruction; other detectors behave worse than DoG.

[17] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[18] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[19] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. of European Conference on Computer Vision*, 2006, pp. 430–443.

[20] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.

[21] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *Proc. of European Conference on Computer Vision*, 2010, pp. 183–196.

[22] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. of IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571.

[23] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. of IEEE International Conference on Computer Vision*, 2011, pp. 2548–2555.

[24] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.

[25] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 553–560.

[26] M. Perdoch, J. Matas, and S. Obdrzlek, "Stable affine frames on isophotes," in *Proc. of IEEE International Conference on Computer Vision*, 2007, pp. 1–8.

[27] T. Tuytelaars and K. Mikolajczyk, *Local Invariant Feature Detectors: A Survey*. Hanover, MA, USA: Now Publishers Inc., 2008.

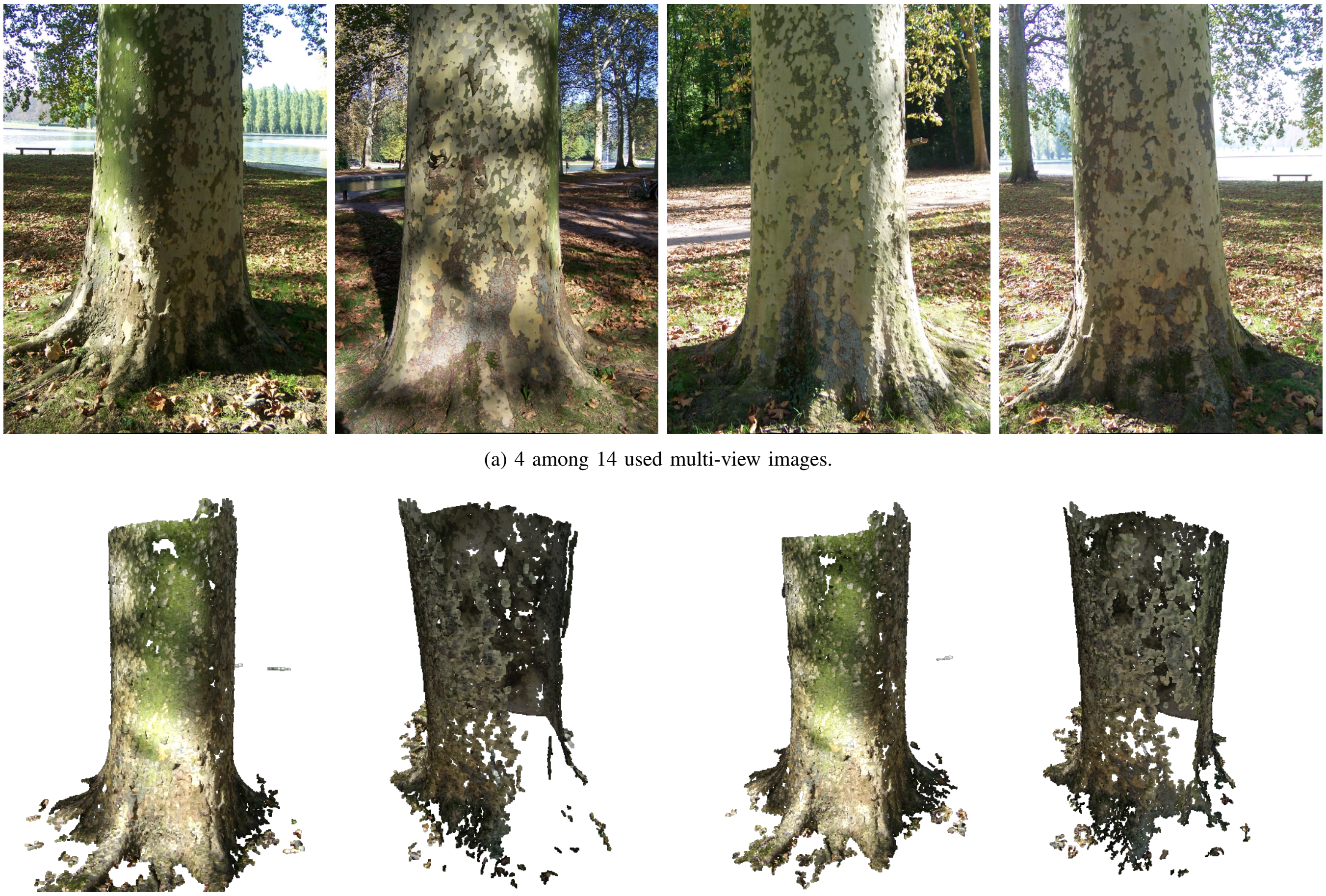
[28] S. Giannarou, M. Visentini-Scarzanella, and G.-Z. Yang, "Probabilistic tracking of affine-invariant anisotropic regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 130–143, 2013.

[29] I. S. Kweon and T. Kanade, "Extracting topographic terrain features from elevation maps," *Journal of Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 59, no. 2, pp. 171–182, 1994.

[30] M. Van Kreveld, R. van Oostrum, C. Bajaj, V. Pascucci, and D. Schikore, "Contour trees and small seed sets for isosurface traversal," in *Proceedings of the Thirteenth Annual Symposium on Computational Geometry*, 1997, pp. 212–220.

[31] G. Reeb, "Sur les Points Singuliers d'une Forme de Pfaff Complètement Intégrable ou d'une Fonction Numérique," *Comptes Rendus Acad. Sciences*, vol. 222, pp. 847–849, 1946.

[32] S. Takahashi, T. Ikeda, Y. Shinagawa, T. L. Kunii,



(a) 4 among 14 used multi-view images.

(b) 3D result using DoG, 13 images are calibrated.

(c) 3D result using TBMR, 14 images are calibrated.

Fig. 18. Densified 3D reconstruction. (a) Some input images. (b) and (c) show the left and right side of the 3D structures reconstructed using respectively DoG and TBMR. The use of DoG misses the right side of the scene; the use of TBMR results in a correct 3D reconstruction; other detectors behave worse than DoG.

and M. Ueda, "Algorithms for extracting correct critical points and constructing topological graphs from discrete geographical elevation data," *Comput. Graph. Forum*, vol. 14, no. 3, pp. 181–192, 1995.

[33] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, no. 5, pp. 363–370, 1984.

[34] C. Berger, T. Géraud, R. Levillain, N. Widynski, A. Baillard, and E. Bertin, "Effective component tree computation with application to pattern recognition in astronomical imaging," in *Proc. of IEEE International Conference on Image Processing*, vol. 4, 2007, pp. 41–44.

[35] E. Carlinet and T. Géraud, "A comparison of many max-tree computation algorithms," in *Mathematical Morphology and Its Applications to Signal and Image Processing—Proc. of the International Symposium on Mathematical Morphology (ISMM)*, ser. Lecture Notes on Computer Sciences, vol. 7883. Springer, 2013, pp. 73–85.

[36] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008, http:/www.vlfeat.org/.

[37] R. Deriche, "Recursively implementating the gaussian and its derivatives," INRIA, Tech. Rep., April 1993.

[38] B. Appleton and H. Talbot, "Recursive filtering of images with symmetric extension," *Signal Processing*, vol. 85, no. 8, pp. 1546–1556, 2005.

[39] O. Miksik and K. Mikolajczyk, "Evaluation of local detectors and descriptors for fast feature matching," in *Proc. of International Conference on Pattern Recognition*, Nov 2012, pp. 2681–2684.

[40] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, J. Bach, and B. Girod, "The Stanford mobile visual search data set," in *MMSYS*, 2011, pp. 117–122.

[41] C. Strecha, W. V. Hansen, L. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[42] L. Moisan, P. Moulon, and P. Monasse, "Automatic Homographic Registration of a Pair of Images, with A Contrario Elimination of Outliers," *IPOL*, 2012, http://dx.doi.org/10.5201/ipol.2012.mmm-oh.

[43] L. Moisan and B. Stival, "A probabilistic criterion to detect rigid point matches between two images and

estimate the fundamental matrix," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 201–218, 2004.

[44] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 7–23, 2000.

[45] P.-E. Forssen and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[46] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, 2006.

[47] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proc. of IEEE Symposium on Foundations of Computer Science*, 2000, pp. 454–463.