



Learning moment-based fast local binary descriptor

Abdelkader Bellarbi, Nadia Zenati, Samir Otmane, Hayet Belghit

► To cite this version:

Abdelkader Bellarbi, Nadia Zenati, Samir Otmane, Hayet Belghit. Learning moment-based fast local binary descriptor. Journal of Electronic Imaging, 2017, 26 (2), (elec. proc.). 10.1117/1.JEI.26.2.023006 . hal-01491567

HAL Id: hal-01491567

<https://hal.science/hal-01491567>

Submitted on 20 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning moment-based fast local binary descriptor

Abdelkader Bellarbi,^{a,b,*} Nadia Zenati,^b Samir Otmame,^a and Hayet Belghit^b

^aUniversity of Evry, Informatique Biologie Intégrative et Systèmes Complexes (IBISC) Laboratory, Evry, France

^bCentre de Développement des Technologies Avancées (CDTA), Robotic Department, Algiers, Algeria

Abstract. Recently, binary descriptors have attracted significant attention due to their speed and low memory consumption; however, using intensity differences to calculate the binary descriptive vector is not efficient enough. We propose an approach to binary description called POLAR_MOBIL, in which we perform binary tests between geometrical and statistical information using moments in the patch instead of the classical intensity binary test. In addition, we introduce a learning technique used to select an optimized set of binary tests with low correlation and high variance. This approach offers high distinctiveness against affine transformations and appearance changes. An extensive evaluation on well-known benchmark datasets reveals the robustness and the effectiveness of the proposed descriptor, as well as its good performance in terms of low computation complexity when compared with state-of-the-art real-time local descriptors.

Keywords: binary descriptor; moments; pattern recognition; computer vision; augmented reality.

1 Introduction

Augmented reality (AR) aims to insert virtual information into the real world. This information should be accurately aligned with the real images and failure to do so can result in pose estimation and pattern recognition problems. Recent computer vision advances offer great potential to AR applications. However, one of the main drawbacks of the AR localization and tracking systems is the high complexity of computing, matching, and storing the large amount of feature-point descriptors that are used for describing and tracking real scenes. Most of the proposed AR systems use floating-point descriptors.¹ Recently, binary descriptors have provided good performance in terms of computational complexity and required memory; however, they still suffer from inaccuracy and low efficiency compared with their floating-point competitors.^{2,3}

In this paper, we propose a robust local binary descriptor for which we have performed binary tests between geometric moments, instead of traditional intensity binary tests, for each subregion in the patch. This new descriptor offers a high distinctiveness level.

The remainder of this paper is presented as follows: Sec. 2 presents the motivation and the main contributions. Section 3 provides related work. Section 4 describes our proposed approach. In Sec. 5, we evaluate the proposed detector descriptor and compare other proposed approaches. Section 6 concludes the paper.

2 Motivation and Contributions

In general, binary descriptors use a single image property, such as intensity differences or gradient directions, to build a description. The result is an insufficient description of image features; furthermore, building high-dimensional feature descriptions with binary descriptors increases

computational complexity and generates redundant and correlated descriptions.

To avoid such drawbacks, it is preferable to create multi-criteria local descriptions and find a compromise between the description dimensions and the high computational complexity.

This work is an extended version of our IEEE ISMAR 2014 publication,⁴ in which we introduced the concept of binary description based on geometric moment comparison and presented the first version of MOBIL descriptor.

In this paper, the main contributions are as follows:

- We provide a broader view of the existing literature.
- We propose an improved keypoint detection technique based on the adaptive and generic accelerated segment test (AGAST)⁵ and Shi-Tomasi⁶ detectors by taking advantage of computing the AGAST binary decision trees, then refining the detected keypoints using the Shi-Tomasi cornerness metric.
- We introduce POLAR_MOBIL, a moments-based binary descriptor, in which we concatenate the description of both Cartesian and log-polar patch images.
- We apply a bit selection strategy to pick the best binary tests that will lead to both low correlation and high distinctiveness.

3 Related Works

Several local feature descriptors that provide well matching probability have been proposed in literature. Scale-invariant feature transform (SIFT)^{7,8} is the most popular descriptor; it consists of a gradient orientation histogram obtained from difference of Gaussian points; however, there are drawbacks in terms of computation time and memory. Mikolajczyk and Schmid⁹ compared the performance of 10 recent descriptors,

and they advocate their gradient location and orientation histogram (GLOH), which is an extension of the SIFT descriptor. It outperforms SIFT and other descriptors in the sense that it increases robustness and distinctiveness. The speeded up robust features descriptor (SURF)¹⁰ was inspired by SIFT. The main contribution of SURF is that it overcame SIFT's main weakness, its computational complexity and the resulting low execution speed.

Variants of the SIFT descriptor have been proposed in order to reduce computation time, for example, principal components analysis (PCA)-SIFT,¹¹ which applies PCA to the normalized gradient patch instead of using SIFT's smoothed, weighted histograms. Mortensen et al.¹² combined a local SIFT descriptor with a global context vector similar to shape contexts to resolve ambiguities that can occur locally when an image has multiple similar regions. Some other variants of SIFT have also been proposed.^{13–18}

Tola et al.¹⁹ proposed an efficient descriptor with a daisy configuration derived from SIFT and GLOH. It has proven to be well suited for sparse matching but not for efficiency. Google's research laboratory has also invented its own descriptor called compact normalized Gabor sampling²⁰ that is based on Gabor wavelets of varying scale and orientation drawn from a variant grid around interest points. Inspired by Weber's law, Chen et al.²¹ proposed a local descriptor called the Weber local descriptor. Alcantarilla et al.²² proposed a multiscale two-dimensional (2-D) feature detection and description algorithm called KAZE that detects and describes 2-D features by means of nonlinear diffusion filtering. Yan et al.²³ recently proposed a local feature descriptor called monotonic invariant intensity descriptor (MIID), which uses spectral embedding and nonsubsampling contourlet transform. However, Yan et al. used the Laplacian matrix to construct their MIID descriptor, which does not meet the real-time requirement.

These approaches use floating-point descriptors, i.e., they are computationally expensive in terms of time and memory; therefore, binary descriptors were introduced to optimize computation resources. The first binary descriptor proposed was binary robust independent elementary features (BRIEF).²⁴ It uses simple binary tests between pixels in a smoothed image patch and random predetermined locations. BRIEF demonstrated a high recognition rate with low computational complexity during the description building and matching processes; however, rotation is variable. Oriented FAST and rotated BRIEF (ORB)²⁵ introduced rotated BRIEF (rBRIEF), which uses a learning strategy to recover from the loss of variance in steered BRIEF. The rBRIEF method demonstrates variance and correlation improvements over the steered BRIEF. Wang et al.²⁶ proposed an improved ORB descriptor that is applied to license plate localization in complex scenes.

The binary robust invariant scalable keypoints descriptor (BRISK)²⁷ uses a hand-constructed sampling pattern composed of concentric rings with more points on outer rings. The BRISK descriptor divides point pairs into two groups: long-distance pairs and short-distance pairs. It calculates the characteristic pattern direction using long-distance pairs and computes the descriptor using intensity comparisons of short-distance pairs after rotation and scale normalization. Baroffio et al.²⁸ proposed BRISKOLA, an optimized version of BRISK for low-power advanced RISC machine architectures. Inspired by the human visual system, fast retina

keypoint²⁹ uses the learning strategy of ORB with a DAISY-like sampling pattern. This sampling pattern uses overlapping concentric circles with more points on inner rings. Each circle represents a sensitive field, where the image is smoothed with its corresponding Gaussian kernel.

Ledoux et al.³⁰ proposed the color local binary pattern (LBP), an extension of the LBP that takes the vector information of color into account due to a color order. This descriptor is very sensitive to noise, which makes it adapted to texture classification. Shang et al.³¹ proposed a local derivative quantized binary pattern descriptor, in which they quantized the intensity differences between the central pixels and their neighbors for the detected local affine covariant regions. However, this proposed descriptor is not suited for real-time applications due to the applied quantization technique.

Levi and Hassner³² recently introduced the learned arrangements of three patch codes (LATCH) binary descriptor; LATCH compares the intensity of three subregions, which called the anchor and its two companions, in the patch to produce a single bit of the descriptor. The similarity comparison between the anchor subregion and the other two subregions is given by their Frobenius norm. The LATCH descriptor offers high recognition rate. However, it is slower than most of the binary descriptors with the same size. Parker et al.³³ proposed a CUDA version of the LATCH binary descriptor in order to accelerate the descriptor extraction using expensive graphics processing units.

These techniques only use the intensity to build their descriptors, which leads to loss of information. The local difference binary descriptor (LDB)³⁴ used in addition the first-order gradient to improve the description quality, but despite that, its sensitivity to viewpoint changes and noise is significant. A quantitative comparison and evaluation for feature descriptors can be found in Refs. 35–37.

Computation of geometric moments is another popular technique that is used to describe image features. Moments have been used successfully in many applications, as they can capture both appearance and shape characteristics of an image region.^{38–40} Their ability to fully describe an object also makes them a powerful tool in computer vision applications.

In this work, we propose an alternate binary descriptor that achieves a computational speed and robustness similar to state-of-the-art binary descriptors, yet offers much higher distinctiveness when compared with them. We use the geometric properties of a patch's subregions instead of the intensity when performing the binary test in order to achieve high quality from our descriptor.

4 Proposed Approach

4.1 Keypoint Detection

Obtaining stable keypoints are an essential step for every feature-based computer vision process. Recently, corner detection became the most useful technique due to its good performance in terms of repeatability and processing time. A corner is defined as the intersecting point of two connected straightedge lines. Mathematically, it refers to the point at which two dominant but different gradient orientations exist.⁴¹

Generally, corner detection methods can be further divided into three classes

- Classical gradient-based corner detection based on gradient calculation.
- Template-based corner detection based on the comparison of pixels. Recently, templates have been combined with machine-learning techniques, i.e., decision trees, for faster corner detection.
- Contour-based detection based on the results of contour and boundary detection. It relies on the prediction of edge responses to identify corners.

As a template-based corner detection technique, the fast accelerated segment test (FAST)⁴² detector represents a breakthrough in high-speed corner detectors. It is based on the accelerated segment test (AST), which is a modification of the smallest univalue segment assimilating nucleus⁴³ corner detector. The AST classifies a candidate point p (with intensity I_p) as a corner if n contiguous pixels in a circle of radius 3 around p are all brighter than $I_p + t$ or all darker than $I_p - t$, with t being a predefined threshold. Each corner is then given a score s , defined as the largest threshold for which p can be classified as a corner. A decision tree is then applied to accelerate the candidate point classification. FAST enhanced repeatability⁴⁴ increases the thickness of a circular template in order to increase the stability of detected corners; however, this makes it slower compared with FAST.

The AGAST detector⁵ is based on the AST that was developed for FAST by Rosten et al. The difference is that in the AGAST detector, the way the decision trees for the AST is built and used has been significantly improved, and being generic, the decision trees do not have to be adapted to new environments. The FAST-like detectors do not provide multiscale detection; therefore, ORB calculates a multiscale image and applies FAST to every scale. The detected points were refined by Harris corner measure;⁴⁵ however, duplicate points with different pyramid levels have been detected as nonmaximum suppression is not used between the scales.

Meanwhile, Leutenegger et al.²⁷ proposed the BRISK detector, which uses the AGAST corner detector refined by a FAST cornerness score. The BRISK detector identifies keypoints in a scale-space pyramid and performs nonmaximum suppression between the pyramid levels. Despite the good performance achieved by the BRISK detector, it still suffers from a lack of precise and effective cornerness measurements as it is very hard to use discrete cornerness measurements to satisfy the requirement of nonmaximum suppression. In addition, the learning process might cause database-dependent problems, so the general performance of machine-learning-based methods needs to be improved.

In order to achieve the invariance against the scale changes crucial for high-quality keypoints, we had to calculate a nine-level scale pyramid with the factor of $\sqrt{2}$ (square root of two) between each two levels. We then had to apply the AGAST five to eight (five decision trees for eight pixels mask) detector for each level in order to extract the maximum high-speed keypoints (Fig. 1).

Subsequently, the extracted keypoints were refined in order to select those that were more stable. For each detected keypoint, we calculated the Shi-Tomasi cornerness score $R = \min(\lambda_1, \lambda_2)$ to check if the detected keypoint was a corner within the fixed threshold. The Harris corner detector

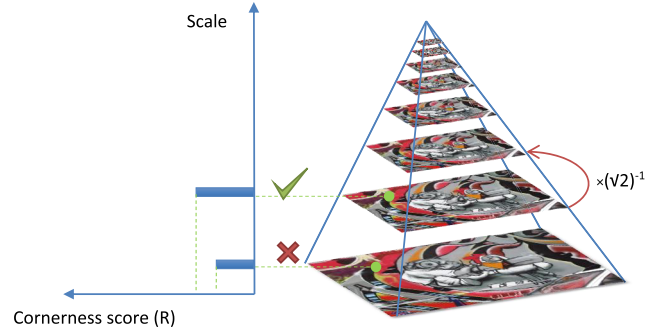


Fig. 1 Scale pyramid keypoint detection. Once the keypoints are extracted from each level by AGAST, we apply the Shi-Tomasi cornerness metric to refine the detected keypoints for each level. After that, we compare the keypoint positions for each two adjacent levels to remove the redundant points with the minimum score.

defines eigenvalues as λ_1 and λ_2 . We selected only the local extrema cornerness score in order to remove the redundant keypoints extracted from each scale level. To this end, we applied an optimization algorithm to compare the extracted keypoint positions for every two adjacent levels, and we removed the redundant points having the minimum Shi-Tomasi score. In Sec. 5.1, we describe the implementation and the evaluation of our proposed detection technique.

4.2 POLAR_MOBIL Descriptor

As most of the photometric changes, such as lighting/contrast changes, blurring, and image noises, can be removed by computing the difference between two subregions, local binary descriptors are becoming widely used in computer vision; however, the use of intensity differences may lead to a lack of discrimination.

In this paper, we propose an alternative binary descriptor in which we perform binary tests between subregion moments, instead of the classical intensity-based binary test, in order to increase descriptor distinctiveness.

Our proposed descriptor relies on two main ideas

- Performance of binary tests between the geometric moments of the patch subregions.
- Calculation of the moment differences for both Cartesian and log-polar images of the patch and concatenate the resulting descriptive vectors.

Image moments have attracted researchers' attention for several decades, as a powerful tool to describe the content of an image. Moments have been used in many areas of computer engineering research, such as pattern recognition^{46,47} and computer vision,^{48,49} with significant results. Hu⁴⁰ was the first to introduce image moments for classification problem solving by using the geometrical, central, and normalized image moments.⁵⁰

In general, moments describe numeric quantities at some distance from a reference point or axis. Moments are commonly used in statistics to characterize the distribution of random variables. The use of moments for image analysis is straightforward if we consider a binary or gray-scale image segment as a 2-D density distribution function. Moments may be used in such cases to characterize an image segment and extract properties of the intensity distribution.

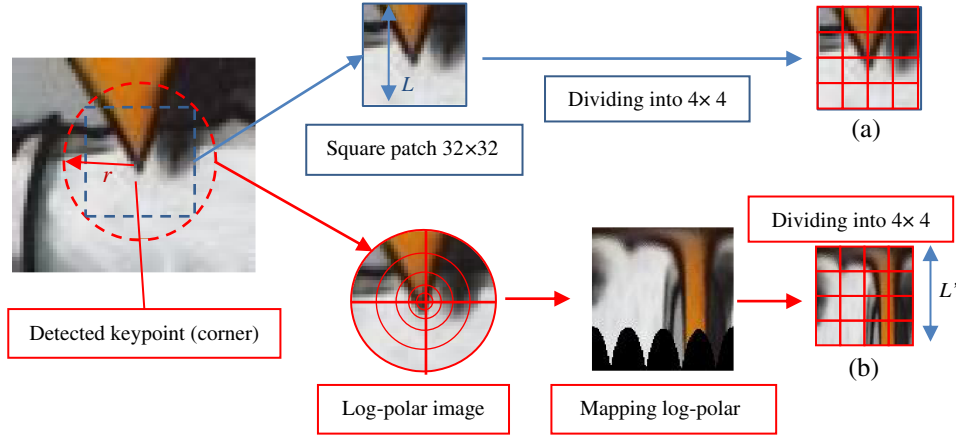


Fig. 2 Extraction and sampling both (a) the Cartesian patch and (b) log-polar patch, where $L = L' = 32$, $r = (1/2)L \times \sqrt{2} = 23$.

Hu states that the 2-D moments for an $N \times M$ image $I(x, y)$ can be defined by

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q I(x, y), \quad (1)$$

where $p, q = 0, 1, 2$.

In order to build our descriptor, we took a square image patch ρ from around the detected keypoint with length $L = 32$ pixels. Then, we divided it into 4×4 equally sized grid cells (subregions). We obtained 16 subregions with 8×8 pixels each.

On the other hand, we took a circular image patch around the same detected keypoint with radius $r = (1/2)L \times \sqrt{2} \approx 23$ pixels and we constructed its log-polar image. We subsequently cropped the exterior region of the mapped log-polar image to obtain a square image patch with length $L' = 32$ pixels. This later has been divided also into 4×4 equally sized grid cells, with 8×8 pixels each. Figure 2 illustrates this procedure. We note that several experimental tests have been carried out in order to choose the adequate size of

the patches and the number and the size of the subregions (grid cells).

Then, we calculated the first-order moments (m_{10} and m_{01}) and the second-order moments (m_{20} and m_{02}) for each grid cell in both Cartesian and log-polar patches (see Fig. 3), where:

- The first-order moments (m_{10}, m_{01}) were used to locate the centroid (center of gravity) of a region, and
- The second-order moments (c, m_{02}) were used to determine the principal axes of the pixel distribution given in the image.

We then performed a binary test τ on each pair of grid cells (x and y) in the Cartesian patch and then in the log-polar patch using the defined moments above as follows:

$$\tau[m_{pq}(x), m_{pq}(y)] = \begin{cases} 1 & \text{if } [m_{pq}(x) > m_{pq}(y)] \\ 0 & \text{otherwise} \end{cases},$$

For all $(p, q) \in \{(0,1), (1,0), (0,2), (2,0)\}$. (2)

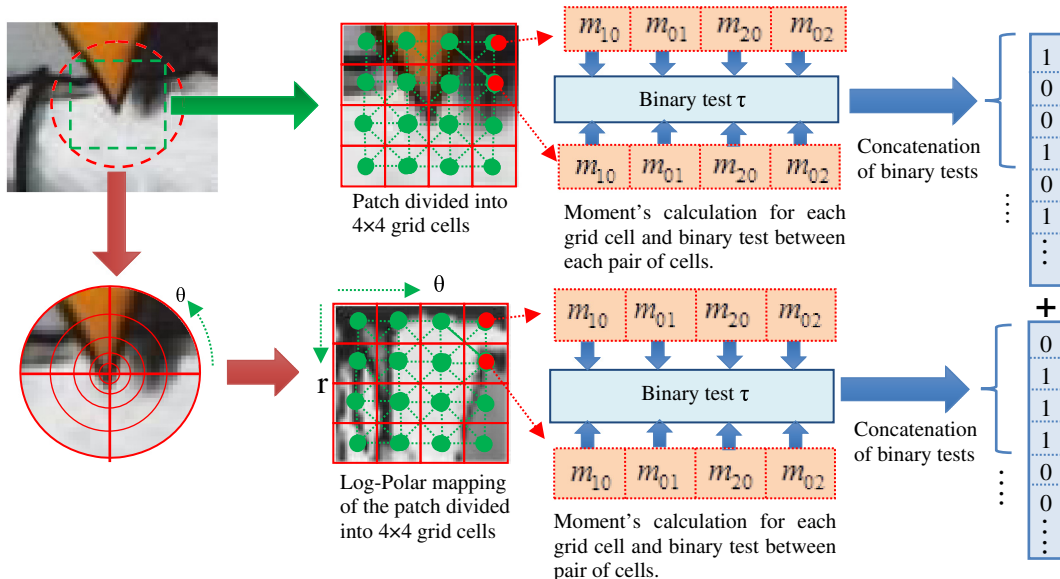


Fig. 3 The POLAR_MOBIL descriptor architecture.

At the end of this step, we obtained two binary vectors for both Cartesian and log-polar patches as shown in Fig. 3. These two vectors are finally concatenated to obtain a full description of the given patch ρ . We can generalize, therefore, the POLAR_MOBIL descriptor as

$$\text{MOBIL}(\rho) = \sum_{i=0}^{n/2} \sum_{j=0}^4 2^{j+4i} \{\tau[m_{pq}(x), m_{pq}(y)]\} + \sum_{i=n/2}^n \sum_{j=0}^4 2^{j+4i} \{\tau[m_{pq}(x), m_{pq}(y)]\}$$

For each $(p, q) \in \{(0,1), (1,0), (0,2)(2,0)\}$,

where n is the number of binary tests in both the Cartesian and log-polar images of the patch ρ . The resulting binary vector then has $n \times 4$ bits.

In order to make our descriptor invariant to an in-plane rotation, we computed a dominant orientation for every image patch, and then aligned the patch to its dominant orientation before computing its descriptor. Many dominant orientation estimation methods can be used. Here, we implemented the intensity moments-based method⁵¹ for its efficiency and robustness.

We calculated the moments for the entire patch intensity using Eq. (1). The centroid intensity of the patch can be defined as

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right). \quad (4)$$

In general, the intensity centroid of a patch is offset from its geometric center. Because of this, we constructed a vector from the geometric center O to the intensity centroid C . The orientation of the patch can be given as: $\theta = \tan^{-1}(m_{01}, m_{10})$.

4.3 Picking the Best of POLAR_MOBIL

Our descriptor generates a high-dimension bit string [(all possible tests in each patch: 120) \times (moments: 4) \times (Cartesian and log-polar: 2) = 960 bits], which increases resource consumption in terms of processing time and storage. Moreover, high-dimension descriptive vectors may contain highly correlated bits that reduce the discriminative ability of the descriptor. For this end, we applied a binary test-selection strategy, which is a modified version of the learning technique used by ORB. In order to both reduce the high

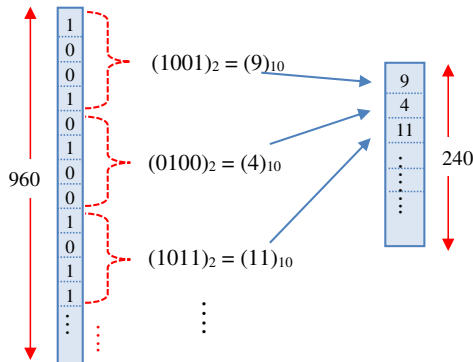


Fig. 4 Conversion of the binary vectors to decimal vectors.

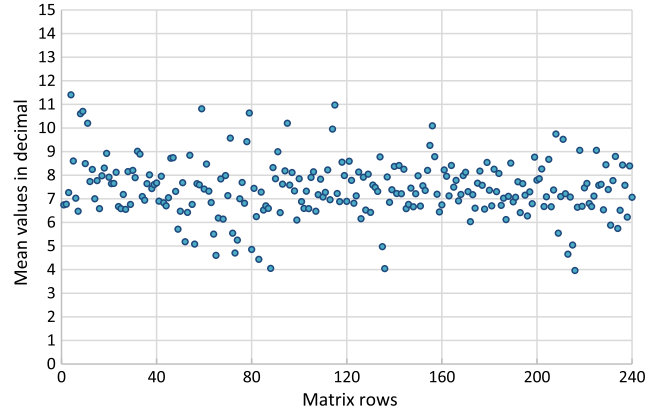


Fig. 5 The mean values of the 240 matrix rows. We can see that the correlated binary test is far from the center (7.5).

dimensionality of this descriptor and to enable us to select the moment's binary tests that would lead to high variance and low correlation.

To accomplish this, we took a large set of images. We used the Caltech256 dataset⁵² that consists of 30,607 images, and we extracted an average of 100 patches (with size of

Algorithm 1 Moments binary test selection.

Input:

- The $k \times n$ matrix of learned patches, where its values are in $[0, 15]$
- Thresholds $\varepsilon_1, \varepsilon_2$ for selecting the mean and the variance, respectively
- $k' = 32 * 2 = 64$: the number of rows to be selected

Output:

- I : a set of selected binary test indices.

Begin:

1. For each row i in the matrix, calculate the mean m_i and the variance v_i .
2. Sort the rows in increasing order of the absolute value of their means -7.5 .
3. $i = 1$; Initialize I ;
4. while $|m_i - 7.5| < \varepsilon_1$, do:
 - a. If $(|v_i - 21.25| < \varepsilon_2)$, then put i in I
 - b. If the size of I achieves k' , then go to step 5
 - c. If $(i = k)$, then readjust ε_1 and ε_2 , go to step 3
 - d. $i++$;
5. End while.
6. Terminate the learning function and output I .

End.

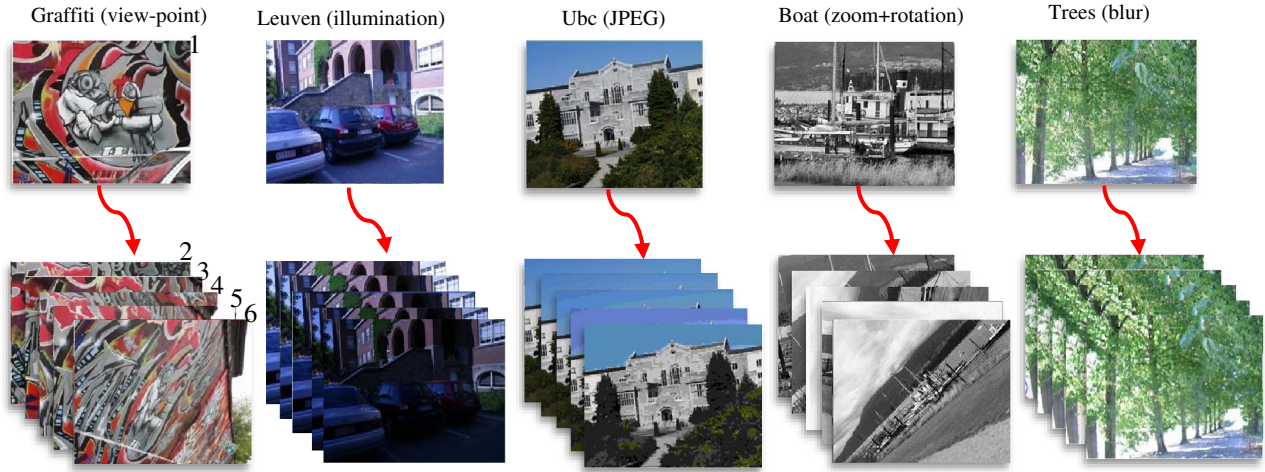


Fig. 6 Example of image sets from the Mikolajczyk benchmark.

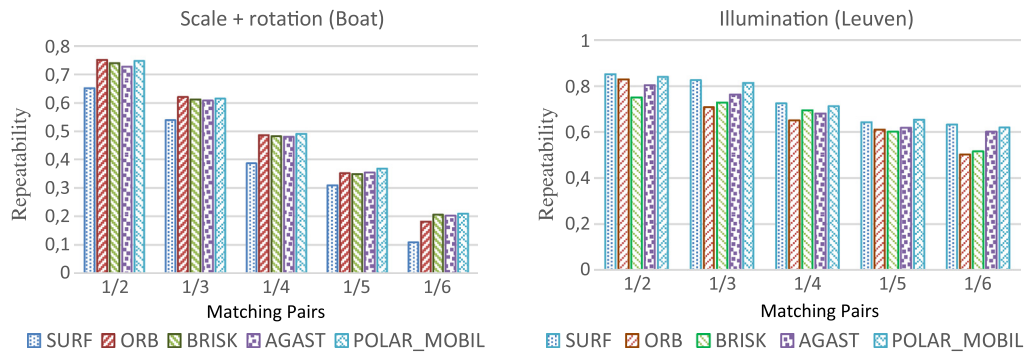


Fig. 7 Comparison of the repeatability between the of POLAR_MOBIL, SURF, ORB, BRISK, and AGAST detectors, using Mikolajczyk benchmark. Each group (1/x) in the horizontal axis indicates the matching result of image 1 to image x, where $x = 2, 3, 4, 5, 6$, in the related image set.

Table 1 Mean detection time for 500 keypoints of POLAR_MOBIL detector and detectors and SURF, ORB, BRISK, and AGAST detectors.

Detectors	Detection time (ms)
SURF	72.63
ORB	8.37
BRISK	4.12
AGAST	1.87
POLAR_MOBIL	3.76

32×32 pixels) from each image ($\sim 3M$ patches at all). Then, we calculated their descriptors by POLAR_MOBIL.

Then, for each binary vector, we grouped together each four (4) successive bits (four bits given from the four moments at each binary test). Moreover, we converted the 4 bits to the decimal equivalent [i.e., $(1001)_2 = (9)_{10}$].

Table 2 Mean description time for the state-of-the-art descriptors and the POLAR_MOBIL descriptor.

Descriptors	Time per description (ms)
SIFT	3.121
SURF	1.488
BRISK	0.072
A-KAZE	0.094
ORB	0.146
LDB	0.139
LATCH	0.437
MOBIL	0.127
POLAR_MOBIL	0.123

Therefore, we obtained decimal vectors with dimension of $960/4 = 240$, and their values are between 0 and 15 as shown in Fig. 4.

Subsequently, those decimal vectors formed $k \times n$ matrix, where $k = 2 * 120 = 240$ is the total number possible binary tests for both Cartesian and log-polar patches, and n is the total number of trained patches ($\sim 3M$). We then calculated the mean and the variance for each row. Figure 5 shows the plotted mean values.

Selecting uncorrelated tests refers to selecting rows that are of uniform distribution, i.e., their means are close to 7.5 (Fig. 5), and their variances should be near to 21.25, according to the equation of discrete uniform distribution variance: $V = (n^2 - 1)/12$ (for proof, see Ref. 53).

The following algorithm (Algorithm 1) details this approach.

Once the learning technique was applied, we noticed that, for the Cartesian image patch, most of the horizontal long-distance tests were removed. One explanation is that as the patch was rotated to its dominant orientation, it appeared horizontally homogenous. For the same reason, long-distance vertical tests in the log-polar patch were also removed.

5 Implementation and Test

We have implemented our proposed detector-descriptor using the Visual Studio 2013 environment, with OpenCV 2.4.4, running on an Intel® Core™ i3 3.20 GHz processor.

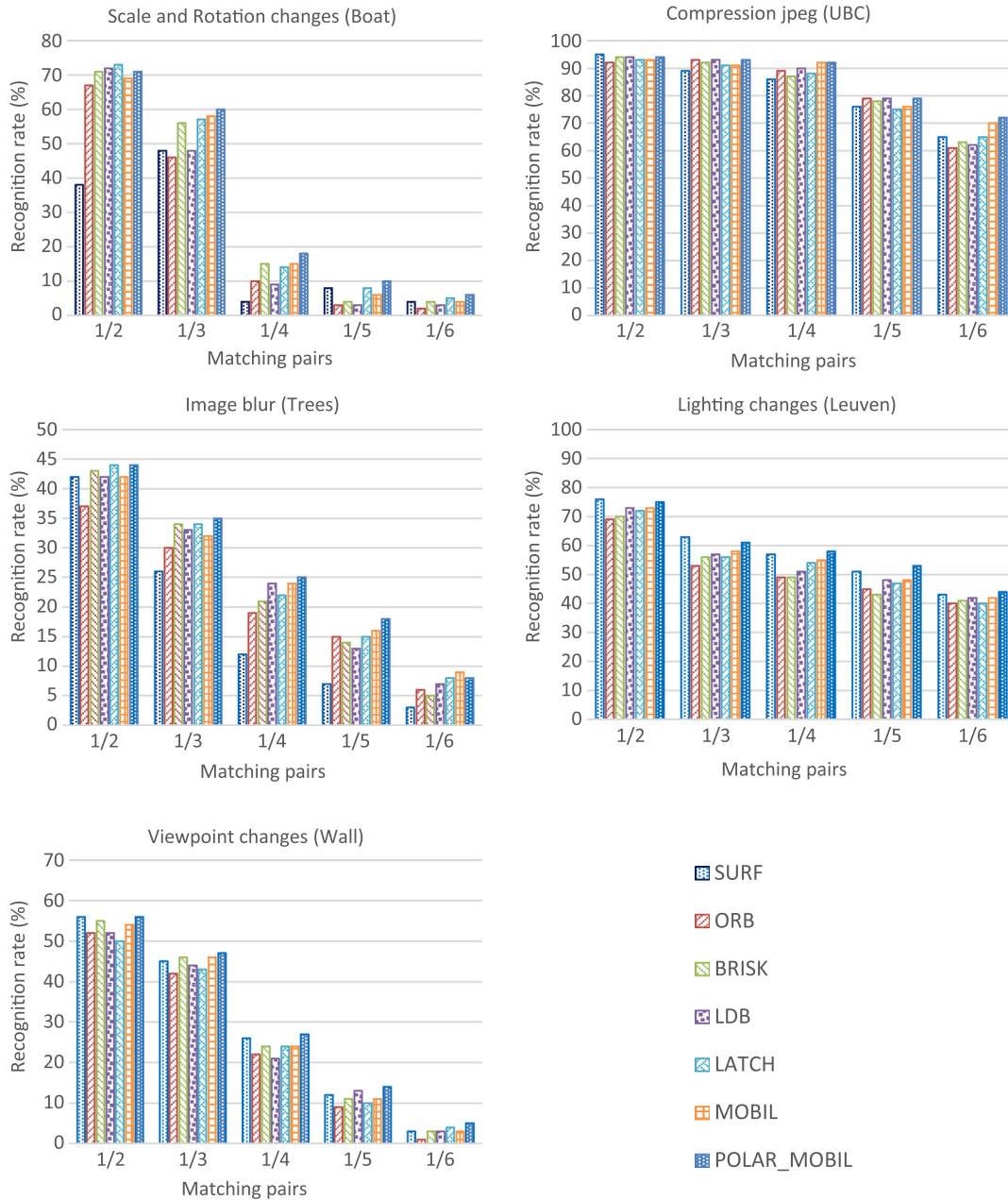


Fig. 8 Comparison of the recognition rates for the POLAR_MOBIL, MOBIL, LDB, ORB, BRISK, SURF, and LATCH descriptors, using Mikolajczyk benchmark. Each group (1/x) in the horizontal axis indicates the matching result of image 1 to image x, where $x = 2, 3, 4, 5, 6$, in the related image set.

We also performed different types of tests in order to evaluate the POLAR_MOBIL performance.

For the evaluation step, we conducted extensive experiments on different public image datasets including: Caltech256,⁵² Fountain-P11,⁵⁴ and Mikolajczyk benchmark,⁹ this later became a standard benchmark for evaluating keypoints detectors and descriptors. It offers five different transformation types represented by eight image sets. Each one contains six images that depict an increasing degree of a specified image transformation. Figure 6 shows the five transformations with one image set each.

5.1 Detector Evaluation

In order to evaluate the POLAR_MOBIL detector performance and compare it with the other proposed detectors, we used the repeatability metric⁵⁵ that represents the ability to detect the same point in the scene under viewpoint and lighting changes and subject to noise. It can be calculated using

$$\text{repeatability} = \frac{\text{number of correspondences}}{\text{total number of detected keypoints}}. \quad (5)$$

We calculated the repeatability of our proposed detector under viewpoint and lighting changes, and we compared it with recent well-known detectors. Figure 7 shows that the POLAR_MOBIL detector performs much better than the ORB and BRISK detectors when changing illumination, and it is slightly better for scale and rotation sequences.

On the other hand, we measured the detection time for detectors tested above and POLAR_MOBIL detector (Table 1). The mean detection time of the tested detectors is measured for extracting 500 keypoints from different image sets. All the descriptors are executed on the same images, under the same configuration as described in Sec. 5.

As illustrated in Table 1, POLAR_MOBIL performs slightly better than BRISK detector, and better ORB and SURF. However, AGAST detector is faster because our

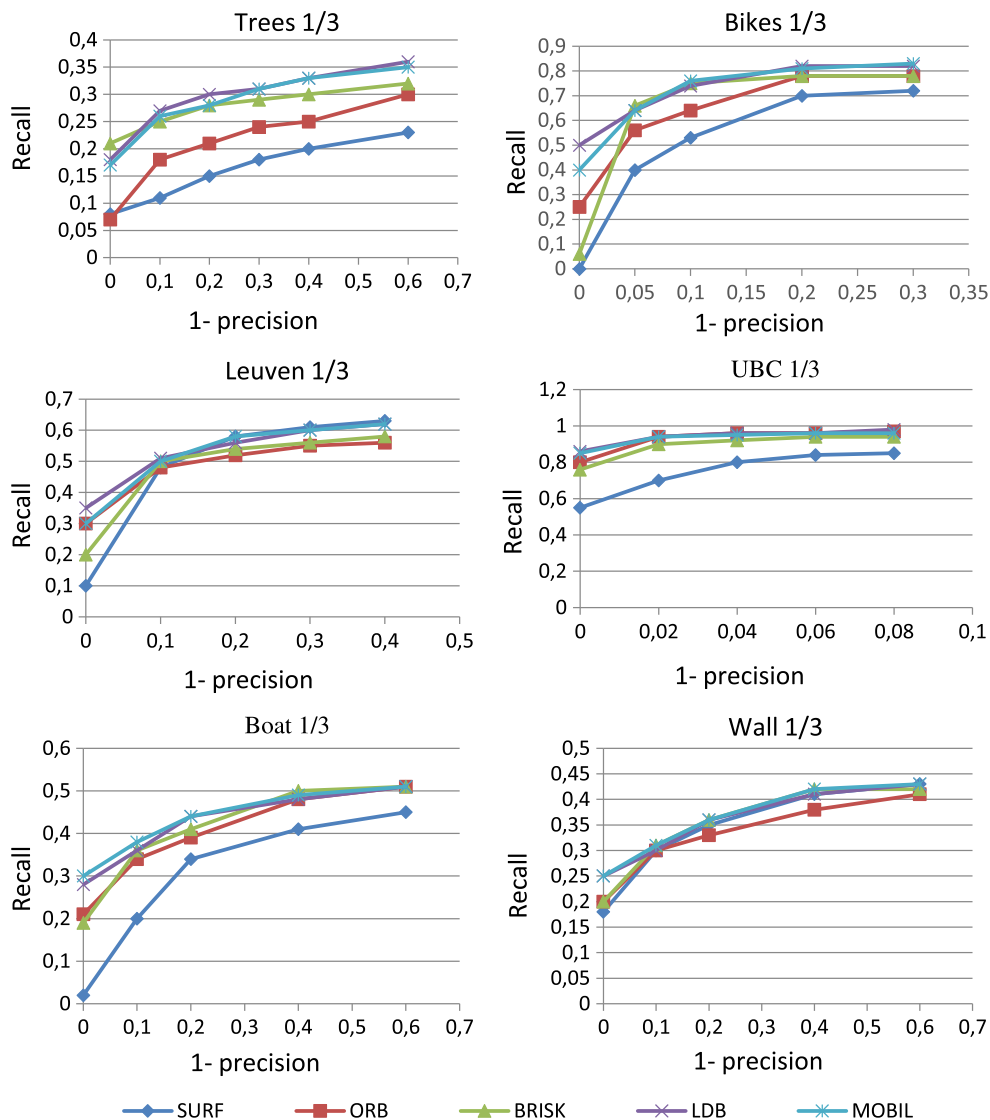


Fig. 9 Recall versus 1-precision curves for POLAR_MOBIL, LDB, ORB, BRISK, and SURF, given by the matching result of the pairs 1/3 (matching images 1 to 3) of six different image sets of the Mikolajczyk benchmark.

detector uses in addition Shi-Tomasi cornerness measure for refinement.

5.2 POLAR_MOBIL Descriptor Performance Test

In this section, we present a set of test results for the POLAR_MOBIL descriptor, and we compare it to other state-of-the-art descriptors.

First, we calculated the average description time for single feature, and we compared it to other feature descriptors. Table 2 shows that the POLAR_MOBIL descriptor has a lower construction time than the ORB and LDB descriptors and a significantly better time than SURF. In addition, POLAR_MOBIL is faster than our first version of MOBIL descriptor.⁴

The average description time is calculated more than 5000 image features from different datasets. All the descriptors compared are executed on the same images and with the same configuration as described in the beginning of Sec. 5.

The first results that we achieved (Fig. 8) demonstrate that, compared with state-of-the-art descriptors, the POLAR_MOBIL descriptor results in a higher recognition rate than the other descriptors for both scale and rotation and view-point changes. (The recognition rate is the number of correct matches divided by the total number of matches.)

We also calculated the recall versus 1-precision curves in order to compare the matching results of our POLAR_MOBIL descriptor with other well-known descriptors. Therefore, we took the matching pair 1/3 (i.e., matching images 1 to 3) from six different image sets in the Mikolajczyk dataset. Then, we tuned from 0 to 255, the matching threshold applied on the Hamming distance in order to discard bad matching results. We measured each time the recall and the

precision of the evaluated descriptors. A robust descriptor should offer high precision at any given recall.

Note that the recall is defined as the number of correctly returned matches divided by the total number of correct matches. The precision is the number of correctly returned matches over the total number of returned matches. A distinctive description should offer high precision at any given recall. The results shown in Fig. 9 demonstrate that POLAR_MOBIL performs significantly better than SURF and ORB, especially for illumination and affine transformation changes.

On the other hand, in order to evaluate the discriminative power of the POLAR_MOBIL descriptor, we applied our descriptor on pairs of images and analyzed the distance between the matching pairs (Hamming distances). Therefore, we took the pairs 1/4 (i.e., matching images 1 to 4) for each of wall, leuven, trees, and boat image sets from the Mikolajczyk dataset. Then, we calculated the matching distance and plotted the distribution of Hamming distances between the different image pairs. Figure 11 shows the obtained Hamming distances distribution (the histograms are normalized). The correct matches are in green, and the mismatches are in red. We note that the maximum possible Hamming distance for POLAR_MOBIL is 256 bits.

As illustrated by the graphs in Fig. 10, most of the curves of the matched pairs and the nonmatched pairs are distinctly separated, which means that the correct matches are easily distinguished from the mismatches. This can validate the high recognition rate of the POLAR_MOBIL descriptor shown in Fig. 8. Figure 11 shows some obtained matching results, for POLAR_MOBIL descriptor with different image datasets.

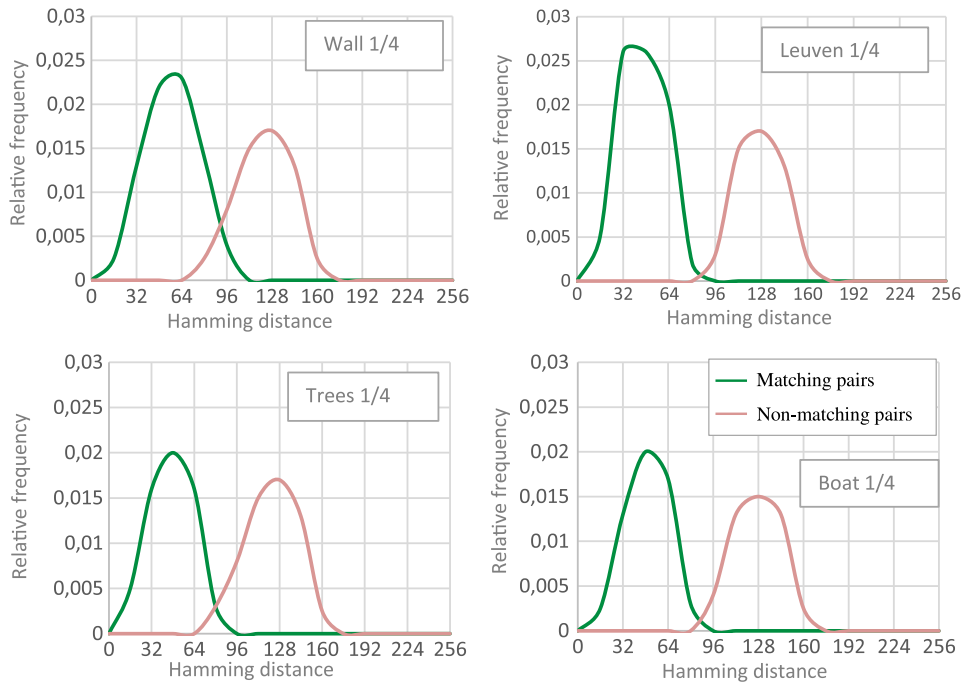


Fig. 10 Distributions of Hamming distances, matching pairs of points are in green, and nonmatching pairs are in red. We see that for most of the pairs, the curves are distinctly separated except in wall pairs, where the curves are partially overlapped.

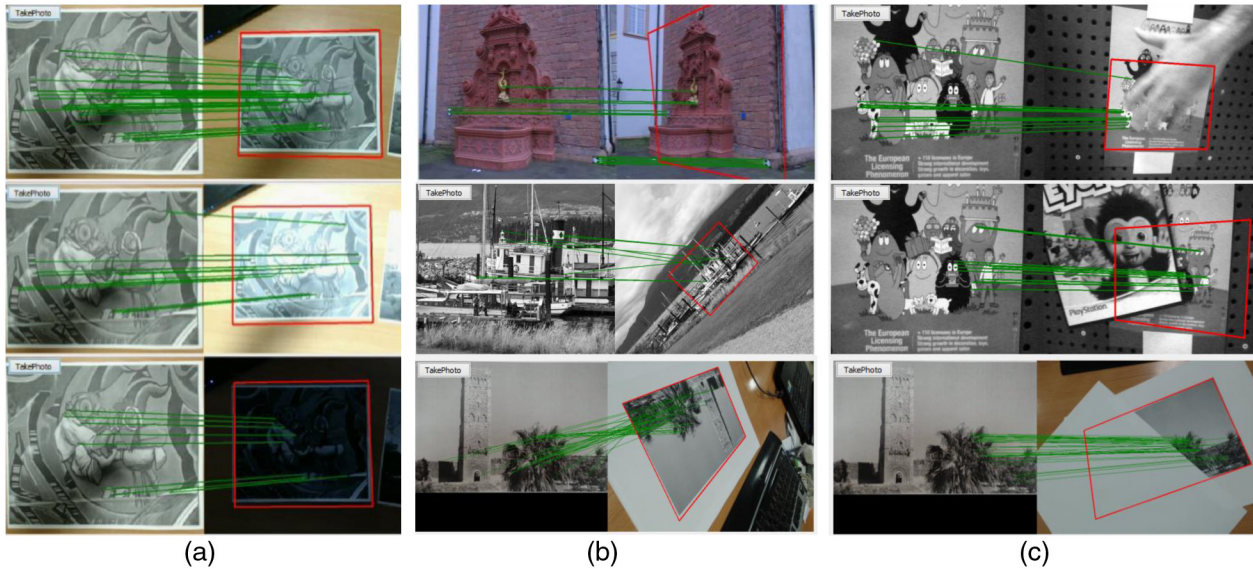


Fig. 11 Some examples of POLAR_MOBIL descriptor with different types of transformations. (a) illumination changes, (b) viewpoint changes, and (c) occlusions.

6 Conclusion

In this paper, we have presented a new way to describe image patches using binary description. In our descriptor, we have introduced geometric information as binary tests to enhance robustness and distinctiveness. We have applied this descriptor to both Cartesian and log-polar image patches in order to create a more robust binary descriptor for affine transformations.

The initial results achieved were very positive, showing that its performance and efficiency are comparable to other popular features descriptors. Our tests demonstrate that this description technique provides more robustness and distinctiveness, especially for affine transformations and viewpoint changes. This new approach has given us new ideas for improving the performance and efficiency of the proposed POLAR_MOBIL descriptor that we hope to implement in the near future.

References

1. E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: a hands-on survey," *IEEE Trans. Visualization Comput. Graph.* **22**(12), 2633–2651 (2016).
2. A. Bursuc, T. Zaharia, and F. Prêteux, "Retrieval of multiple instances of objects in videos," in *MMM'12 Proc. 18th Int. Conf. Advances in Multimedia Modeling*, Springer Science & Business Media, Klagensfurt, Austria, Vol. **7131**, p. 358 (2012).
3. A. Canclini et al., "Evaluation of low-complexity visual feature detectors and descriptors," in *18th Int. Conf. on Digital Signal Processing (DSP)*, pp. 1–7 (2013).
4. A. Bellarbi et al., "MOBIL: a moments based local binary descriptor," in *Int. Symp. on Mixed and Augmented Reality (ISMAR 2014)*, pp. 251–252, IEEE, Munich, Germany (2014).
5. E. Mair et al., "Adaptive and generic corner detection based on the accelerated segment test," in *Computer Vision (ECCV 2010)*, pp. 183–196, Springer, Berlin, Heidelberg (2010).
6. J. Shi and C. Tomasi, "Good features to track," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Proc. (CVPR 1994)*, pp. 593–600, IEEE (1994).
7. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. on Computer Vision* (1999).
8. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
9. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), 1615–1630 (2005).
10. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," in *Computer Vision (ECCV 2006)*, pp. 404–417, Springer, Berlin, Heidelberg (2006).
11. Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2004)*, Vol. 2, pp. II-506, IEEE (2004).
12. E. N. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, Vol. 1, pp. 184–190, IEEE (2005).
13. G. Kordelas and P. Daras, "Robust SIFT-based feature matching using Kendall's rank correlation measure," in *16th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 325–328, IEEE (2009).
14. K. Liao, G. Liu, and Y. Hui, "An improvement to the SIFT descriptor for image representation and matching," *Pattern Recognit. Lett.* **34**(11), 1211–1220 (2013).
15. Y. Bastanlar, A. Temizel, and Y. Yardımcı, "Improved SIFT matching for image pairs with scale difference," *Electron. Lett.* **46**(5), 346–348 (2010).
16. P. Moreno, A. Bernardino, and J. Santos-Victor, "Improving the SIFT descriptor with smooth derivative filters," *Pattern Recognit. Lett.* **30**(1), 18 (2009).
17. G. Zhao et al., "KPB-SIFT: a compact local feature descriptor," in *Proc. of the Int. Conf. on Multimedia*, pp. 1175–1178, ACM (2010).
18. R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, "Linear dimensionality reduction applied to scale invariant feature transformation and speeded up robust feature descriptors," *J. Electron. Imaging* **23**(3), 033017 (2014).
19. E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, pp. 1–8 (2008).
20. Y. T. Zheng et al., "Tour the world: building a web-scale landmark recognition engine," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 1085–1092, IEEE (2009).
21. J. Chen et al., "WLD: a robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1705–1720 (2010).
22. P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Computer Vision (ECCV 2012)*, pp. 214–227, Springer, Berlin, Heidelberg (2012).
23. P. Yan et al., "Local feature descriptor invariant to monotonic illumination changes," *J. Electron. Imaging* **25**(1), 013023 (2016).
24. M. Calonder et al., "Brief: binary robust independent elementary features," in *Computer Vision (ECCV 2010)*, pp. 778–792, Springer, Berlin, Heidelberg (2010).
25. E. Rublee et al., "ORB: an efficient alternative to SIFT or SURF," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 2564–2571 (2011).
26. R. Wang et al., "License plate localization in complex scenes based on oriented FAST and rotated BRIEF feature," *J. Electron. Imaging* **24**(5), 053011 (2015).
27. S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: binary robust invariant scalable keypoints," in *IEEE Int. Conf. on Computer Vision (ICCV 2011)*, pp. 2548–2555 (2011).

28. L. Baroffio et al., "Briskola: brisk optimized for low-power ARM architectures," in *IEEE Int. Conf. on Image Processing (ICIP)*, pp. 5691–5695 (2014).
29. A. Alahi, R. Ortiz, and P. Vanderghenst, "FREAK: fast retina key-point," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 510–517 (2012).
30. A. Ledoux, O. Losson, and L. Macaire, "Color local binary patterns: compact descriptors for texture classification," *J. Electron. Imaging* **25**(6), 061404 (2016).
31. J. Shang et al., "Robust image region descriptor using local derivative ordinal binary pattern," *J. Electron. Imaging* **24**(3), 033009 (2015).
32. G. Levi and T. Hassner, "LATCH: learned arrangements of three patch codes," in *IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Lake Placid, New York (2016).
33. C. Parker et al., "The CUDA LATCH binary descriptor: because sometimes faster means better," in *European Conf. on Computer Vision (ECCV)*, Amsterdam, The Netherlands (2016).
34. X. Yang and K. T. Cheng, "LDB: an ultra-fast feature for scalable augmented reality on mobile devices," in *Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp. 49–57 (2012).
35. A. Choksuriwong et al., "Comparative study of global invariant descriptors for object recognition," *J. Electron. Imaging* **17**(2), 023015 (2008).
36. S. Bianco et al., "Local detectors and compact descriptors for visual search: a quantitative comparison," *Digital Signal Process.* **44**, 1–13 (2015).
37. L. Eikvil and M. Holden, "Evaluation of binary descriptors for fast and fully automatic identification," in *22nd Int. Conf. on Pattern Recognition (ICPR 2014)*, pp. 154–159, IEEE (2014).
38. G. A. Papakostas, E. G. Karakasis, and D. E. Koulouriotis, "Accurate and speedy computation of image legendre moments for computer vision applications," *Image Vision Comput.* **28**(3), 414–423 (2010).
39. G. A. Papakostas et al., "Moment-based local binary patterns: a novel descriptor for invariant pattern recognition applications," *Neurocomputing* **99**, 358–371 (2013).
40. M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory* **IT**(8), 179–187 (1962).
41. Y. Li et al., "A survey of recent advances in visual feature detection," *Neurocomputing* **149**, 736–751 (2015).
42. E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conf. on Computer Vision (ECCV)*, Vol. 1 (2006).
43. S. M. Smith and J. M. Brady, "SUSAN—a new approach to low level image processing," *Int. J. Comput. Vision* **23**(1), 45–78 (1997).
44. E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1), 105–119 (2010).
45. C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conf.*, pp. 147–151 (1988).
46. K. Demaagd et al., *Practical Computer Vision with Simple, CV-Making Computers See in Python*, O'Reilly Media Inc., California (2012).
47. J. Matas, O. M. U. Chum, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conf.*, pp. 384–393 (2002).
48. F. Jurie and C. Schmid, "Scale-invariant shape features for recognition of object categories," in *Computer Vision and Pattern Recognition*, Vol. II, pp. 90–96 (2004).
49. L. M. J. Florack et al., "General intensity transformations and differential invariants," *JMIV* **4**(2), 171–187 (1994).
50. J. Koenderink, "The structure of images," *Biolog. Cybern.* **50**, 363–370 (1984).
51. P. L. Rosin, "Measuring corner properties," *Comput. Vision Image Understand.* **73**(2), 291–307 (1999).
52. G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, Mimeo, Caltech, California (2007).
53. N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, Vol. **444**, John Wiley & Sons, New York City (2005).
54. C. Strecha et al., "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, Anchorage, AK, pp. 1–8 (2008).
55. C. Bauckhage, R. Mohr, and C. Schmid, "Evaluation of interest point detectors," *Int. J. Comput. Vision* **37**(2), 151–172 (2000).

Abdelkader Bellarbi received his BS and MS degrees in engineering from Blida University and Laghouat University, Algeria, in 2009 and 2012, respectively. He is a PhD student at the Université d'Evry Val d'Essone, France. He works as a research engineer at the Centre de Développement des Technologies Avancées (CDTA), Algiers. His research interests include computer vision, augmented reality (AR), and three-dimensional (3-D) interaction.

Nadia Zenati received her PhD in electronic from the University of Franche-Comté, France, in 2008. She is a researcher at the CDTA, Algeria. She has been the head of the IRVA team since 2014 and head of the Robotics Department since 2016. Her areas of interest are image processing, computer vision, collaborative work, mixed reality, and 3-D interaction. She has led several research projects in computer vision and AR.

Samir Otmame is a full professor of computer science at the Université d'Evry Val d'Essone, France, and a member of the Informatique, Biologie Intégrative et Systèmes Complexes Lab. He received his PhD and Habilitation à Diriger des Recherches in computer science from Evry University in 2000 and 2010, respectively. From 2012 to 2014, he served as a project manager for research development under Evry University presidents. In 2015, he was elected as a vice chairman of the board of directors of the university. His research interests include augmented and virtual reality, computer vision, and 3-D interaction.

Hayet Belghit received her MS and engineering degrees from Tizi Ouzou University and Blida University, Algeria, in 2015 and 2010, respectively. She works as a research engineer at the CDTA, Algiers. Her research interests include computer vision, AR, and 3-D interaction.