IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 26, 2019, accepted December 12, 2019, date of publication December 16, 2019,
date of current version December 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2960203

An Adaptive Local Descriptor Embedding Zernike Moments for Image Matching

BIN ZHOU[1,2,3], XUE-MEI DUAN[1], WEI WEI[4], (Senior Member, IEEE), DONG-JUN YE[1], MARCIN WOŹNIAK[5], AND ROBERTAS DAMAŠEVIČIUS[5,6]

[1]School of Sciences, Southwest Petroleum University, Chengdu 610500, China
[2]Institute of Artificial Intelligence, Southwest Petroleum University, Chengdu 610500, China
[3]Research Center of Mathematical Mechanics, Southwest Petroleum University, Chengdu 610500, China
[4]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China
[5]Institute of Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland
[6]Department of Software Engineering, Kaunas University of Technology, 51386 Kaunas, Lithuania

Corresponding authors: Bin Zhou (binzhou@swpu.edu.cn) and Wei Wei (weiwei@xaut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11301414, in part by the National Key Research and Development Program of China under Grant 2018YFB0203901, and in part by the Key Research and Development Program of Shaanxi Province under Grant 2018ZDXM-GY-036.

ABSTRACT Image matching is an important problem in computer vision and many technologies based on local descriptors have been developed. In this paper, we propose a novel local features descriptor based on an adaptive neighborhood and embedding Zernike moments. Instead of a fixed-size neighborhood, a size changeable neighborhood is introduced to detect the key-points and describe the features in the frame of Gaussian scale space. The radius is determined by the scale parameter of the key-point and the dominant direction is computed based on skew distribution fitting instead of the traditional eight-direction statistics. Then a 72-dimensional features vector based on a 3×3 grid is presented. A 19-dimensional vector consists of Zernike moments is applied to achieve better rotation invariance and finally contributes to a 91-dimensional descriptor. The accuracy and efficiency of proposed descriptor for image matching are verified by several numerical experiments.

INDEX TERMS Scale invariance, Zernike moment, adaptive neighborhood, dominant direction fitting, difference of Gaussian.

I. INTRODUCTION

Image matching is one of the core tasks in computer vision and it is the basis of many subsequent applications, it often refers to get the mapping between two different images by analyzing the pixel values, structures, textures, etc. It has been widely used in many fields such as object recognition, medical diagnosis, remote sensing image analysis, motion tracking, 3D scene reconstruction and visual navigation etc. [12], [15], [21], [25], [27], [35], [38], [39], [46].

Region-based matching directly use the original pixel parts in the image pair and compute the distance [17], [55], [36], [45]. These techniques are easy to be applied, and they can be found different invariance or stability in some cases, but they are also computationally intensive and easy to be affected by the geometrical differences or appearance changes.

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

Feature-based matching methods [3], [10], [12], [24], [55] aim to achieve the matching by the simplified representation of images. Point, line, surface and some other geometric forms are usually regarded as the carriers of the features with the invariance in some transform such as scaling, rotation, illumination change, viewpoint change, and so on. Matching can be finally achieved after computing the correlation matrix or other metrics. It is generally agreed that the extraction of powerful features plays a relatively more important role, since even the best matching algorithm will fail to achieve good results if poor features were used. Local descriptor is one of the common power techniques to represent the images briefly and many approaches based on it have been presented in the past years [1], [3], [9], [16], [18], [25], [47], [50].

The local invariant feature was put forward firstly by Moravec [30] and Harris and Stephens [13]. Grayscale auto-correlation function is used in Moravec for simple corners detecting, but the results are easy to be affected by rotation or noise. Local auto-correlation matrix is introduced in Harris to find feature points adaptive to rotation

and illumination. Local pixel statistics is introduced in SUSAN for low time complexity detection [41]. However, these earlier techniques are not invariant in scale.

Scale-invariant feature transform (SIFT) was proposed by David in order to improve the invariance of previous methods [24], [25]. SIFT describes the feature by using image pyramid, dominant direction, local gradient information, etc. The 128-dimensional feature descriptor is less affected by changes in scale and brightness, and more stable to the viewing angle change, affine transformation and noise. However, the computational complexity and false matching rate are still considerable. Numerous variations on SIFT have been approached over the past years [16], [21], [29]. Principal component analysis (PCA) has been applied to replace the weighted histograms in SIFT and the dimension of the feature descriptor can be reduced [16].

By relying on integral images for image convolutions and using a Hessian matrix-based measure for the detector and a distribution-based descriptor, speed up robust features (SURF) [3] can approximate or even outperform previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. Leutenegger et.al. apply a sampling pattern consisting of points lying on appropriately scaled concentric circles at the neighborhood of each key-point to retrieve gray values and process local intensity gradients, then determine the feature characteristic direction [18]. A retina-inspired key-point descriptor is proposed to enhance the performance and a cascade of binary strings is computed by efficiently comparing image intensities over a retinal sampling patter [1]. Besides of these scale invariant matching methods, several attempts have also been made to create local image descriptors invariant to affine transformations [28], [29], [37]. There are also some other techniques such as Zernike moments, k-d tree, locally linear transforming, geometric algebra introduced to advance the traditional methods [6], [14], [21], [44], [47], [52].

Local binary patterns (LBP) is considered among the most computationally efficient high-performance texture features [23], [33], [34], [40].It is often sensitive to image noise and unable to capture macrostructure information. These techniques first compute the difference between a pixel with the neighborhood, and then apply a binary pattern to compute the represent values. Finally a histogram can be taken on each subarea of the representation version and the LBP features are achieved. However, the discriminative ability of the binary descriptors is often limited in comparison with general floating point ones. Tan and Triggs et al. extend the LBP to a local ternary patterns (LTP) for better robust results [42]. Liu et.al compare regional image medians rather than raw image intensities, then a multiscale LBP type descriptor is computed by efficiently comparing image medians over a novel sampling scheme, which can capture both microstructure and macrostructure texture information [23].

With the rapid development of machine learning and convolution neural networks (CNN), some leaning-based techniques are presented to address matching tasks [26]. Multiple local binary descriptors can be integrated in a learning frame to improve the discriminative ability of individual binary descriptors significantly [11]. After each local patch categorized into a rotational binary pattern, the orientation of each pattern and the projection matrix can be jointly learned to obtain a rotation-invariant local binary descriptor [8].

However, despite some advantages in computing speed or description quality, the previous approaches still suffer in terms of reliability and robustness as it has insufficient tolerance to mixed effects of scale changes, transformations, degradation, and so on. The inherent difficulty in extracting suitable features from an image lies in balancing two competing goals: enough quantity and accurate description.

This paper is to present a novel local descriptor aiming at more correct matching, better accuracy and robustness. Instead of a fixed neighborhood in traditional SIFT, an adaptive neighborhood is used to determine extreme points in DoG (Difference of Gaussian) and the radius is related with the scale, so the key-points will be more stable and robust to noise and some changes. The dominant direction is computed by skew distribution fitting instead of the traditional eight-direction statistics. Then Zernike moments are embedded in a local descriptor to improve the rotation invariance. Finally, Euclidean distance and inverse triangular cosine distance are applied to compute the similarity and matching result. The rest of this paper is organized as following.

In Section 2, the related fundamentals of scale space, local descriptor and Zernike moments are prepared. An improved feature descriptor is proposed in Section 3 based on the adaptive neighborhood, dominant direction fitting and embedding Zernike moments. In Section 4, several experiments are used to verify the accuracy and efficiency of proposed algorithm. Finally Section summarizes the research findings.

II. SIFT AND ZERNIKE MOMENTS

A. SCALE SPACE AND DIFFUSION EQUATION

The basic idea of scale space theory [51] is to add a scale parameter to the image processing model, and then a scale space representation sequence can be obtained with the continuous change of the parameter. Analysing and feature extraction can be implemented on the sequence for extracting features of the original image.

The representation of the image scale space can be related to the diffusion equation (1) and it means the spatial distribution of the pixel values over different observation scale. It is the classical diffusion equation and many researches have been presented based on its different forms [5], [19], [20], [48], [49].

$$\begin{cases} \dfrac{\partial u}{\partial \sigma}(x, y; t) = c\nabla^2 u(x, y; t), & (x, y; t) \in \Sigma, \\ u(x, y; 0) = f(x, y), & (x, y) \in \Omega. \end{cases} \tag{1}$$

Here $\Sigma = \Omega \times (0, +\infty)$, t means the scale parameter and c is a constant. $u(x, y; t)$ denotes the image scale space and $\Omega \subset R^2$. With some techniques such as fourier transform applied, it is

FIGURE 1. An image sequence generated by σ from 1 to 8.

easy to get the solution $u(x, y; t) = f(x, y) * g_{\sqrt{2t}}(x, y)$, where $g_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2+y^2}{2\sigma^2})$.

Then a scale space operator can be defined as $g_\sigma : f(x, y) \to F(x, y; \sigma) = f(x, y) * g_\sigma(x, y)$ and the corresponding scale space can also be denoted as the operator set $[g_\sigma]_{\sigma>0}$. It is the general Gaussian space in image processing. Figure 1 shows an image sequence generated by the scale parameter σ changed from 1 to 8.

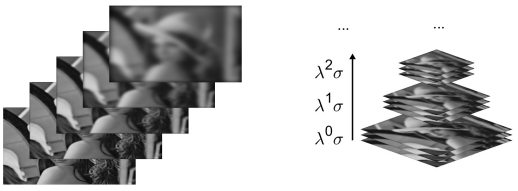
B. SCALE-INVARIANT FEATURE TRANSFORM

Scale-Invariant Feature Transformation(SIFT) [24], [25] is a popular technique describing image local features. The basic idea of SIFT is to approximate the Gaussian Laplace space with an easy-to-calculate Gaussian difference space, and to achieve fast scale crossing by down-sampling. After some stable key-points being selected, the corresponding feature representation vectors called descriptors can be constructed through neighborhood gradient statistics.

The SIFT mainly includes Gaussian pyramid construction, key-points detection, and feature descriptors generation.

(1)Gaussian pyramid construction. As shown in Figure 2(a), the traditional Gaussian space is generated by convoluting the input image with different Gaussian kernel function as $F(x, y; \sigma) = f(x, y) * g_\sigma(x, y)$.

In order to reduce the actual calculation, the downsampling is usually introduced to achieve the scale crossing and form a Gaussian pyramid as shown in Figure 2(b).



(a) Traditional Gaussion space (b) Gaussian Pyramid

FIGURE 2. Traditional Gaussian space and Gaussian pyramid.

Lindeberg [22] believes that the scale normalized Laplacian of Gaussion (LoG) operator $\nabla^2 g$ holds the true scale invariance and the stable extreme points in LoG scale space can be related to the image feature points. Based on a diffusion equation $\frac{\partial g}{\partial \sigma} = \sigma\nabla^2 g$, the LoG operator in the computation can be approximated by the following expression

$$\nabla^2 g \approx \frac{g_{k\sigma}(x, y) - g_\sigma(x, y)}{\sigma^2(k-1)}. \quad (2)$$

It means the product of a constant factor $(k-1)\sigma^2$ and the Difference of Gaussion operator (DoG) $D_\sigma(x, y) = g_{k\sigma}(x, y) - g_\sigma(x, y)$.

Suppose the image Gaussian pyramid has O groups and S layers in each group, the number O and S have been proper set for the image. To ensure the scale continuity in the difference of Gaussian space, the layers number in each group is extended to $S + 3$. In the image scale space, the bottom image is generally the λ times size of the original image. The scale ratio of the adjacent two layers in the same group is $k = \lambda^{1/S}$ and the first layer image in the next group is obtained by downsampling the $S+1$ layer image in the previous group. Difference of Gaussion space can be generated by the subtraction of each two adjacent layer images in the same group. A specific structure with $S = 2$ is shown in Figure 3.

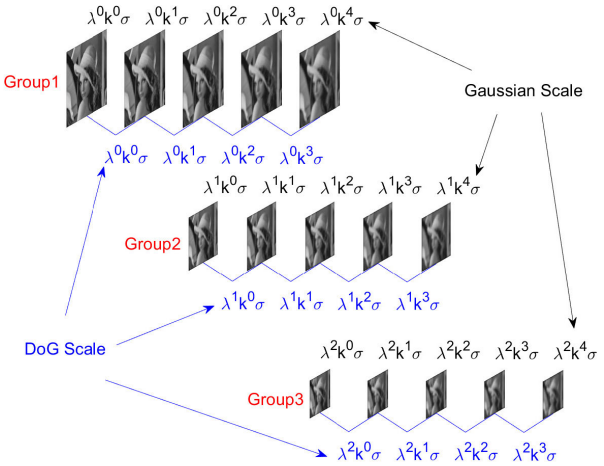


FIGURE 3. A specific scale space structure with $S = 2$.

(2)Key-points detection. The difference of Gaussian space can be used to efficiently extracted the key-points. A preliminary screening is performed by comparing the candidate key-points (extreme points) with a total of 26 pixels in the $3 \times 3 \times 3$ neighborhood around it. Let $D(\mathrm{x})$ and $U_\delta(\mathrm{x})$ denote the DoG function and δ-neighborhood of x, the extreme points set can be denoted by

$$KPT_0 = \{\mathrm{x}, D(\mathrm{x}) \geq D(\mathrm{y}) \forall \mathrm{y} \in U_1(\mathrm{x})\}. \quad (3)$$

Here $\mathrm{x} = [x, y, \sigma]^T$ denotes the image coordinate of a candidate key-point.

Then the Taylor expansion (4) is used to correct the location.

$$D(\mathrm{x} + \epsilon) = D(\mathrm{x}) + \frac{\partial D}{\partial \mathrm{x}}(\mathrm{x})^T \epsilon + \frac{1}{2}\epsilon^T \frac{\partial^2 D}{\partial \mathrm{x}^2}\epsilon + o(|\epsilon|^2) \quad (4)$$

while $\epsilon = [\Delta x, \Delta y, \Delta\sigma]^T$ means a small change. It can be solved by

$$\epsilon = -(\frac{\partial^2 D}{\partial \mathrm{x}^2})^{-1}\frac{\partial D}{\partial \mathrm{x}}. \tag{5}$$

However, the DoG operator will produce a strong response near the edge. In order to ensure the stability and reliability of key-points, it is necessary to further eliminate some bad candidate points such as low-contrast feature points and edge response points. Let the Hessian matrix at x denoted by

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}. \tag{6}$$

The trace and determination of H can be computed as $Tr(H) = D_{xx} + D_{yy}$ and $Det(H) = D_{xx}D_{yy} - D_{xy}^2$. Then bad key-points can be eliminated by the condition $\frac{Tr(H)^2}{Det(H)} > \frac{(\gamma+1)^2}{\gamma}$ with a threshold γ set properly. The final key-point set can be denoted by

$$\begin{aligned} KPT = \{\mathrm{x} \in KPT_0,\ |-(\frac{\partial^2 D}{\partial \mathrm{x}^2})^{-1}\frac{\partial D}{\partial x}| < 0.5 \\ \wedge\ |D(x)| < 0.03 \wedge \frac{Tr(H)^2}{|H|} < \frac{(r+1)^2}{r})\}. \end{aligned} \tag{7}$$

(3)Feature descriptors generation. In order to obtain the descriptors of key-points, the dominant direction will be determined by the neighborhood gradient statistics. More exactly, in the 3σ-radius circular neighborhood of a key-point, the inner points' gradient can be counted in eight directions which are uniformly distributed on $[0, 2\pi)$. The dominant directions of the key-points can be obtained from some peak directions of the histogram. Then the neighborhood will be rotated the same angle as dominant directions, and divided into 4×4 sub-regions. Gradients in each sub-region are all counted in the eight directions to generate a local descriptor. All the 16 local descriptors can be combined to the final 128-dimension descriptor of the key-point. The scale invariance and rotation invariance can be preserved by the descriptor.

C. ZERNIKE MOMENTS

Zernike proposed a set of complex polynomials defined on unit circles in polar coordinates as follows [31]

$$V_{n,m}(x, y) = V_{n,m}(\rho, \theta) = R_{n,m}(\rho)\exp(jm\theta) \tag{8}$$

where (ρ, θ) is the polar coordinate representation of the point (x, y). Nonnegative integer n and integer m satisfy

$$\begin{cases} mod(n - |m|, 2) = 0 or m = 0, \\ n \geq |m|. \end{cases}$$

$R_{n,m}(\rho)$ is a set of orthogonal radial polynomials and satisfies $R_{n,-m}(\rho) = R_{n,m}(\rho)$

$$R_{n,m}(\rho) = \sum_{k=0}^{(n-|m|)/2} \frac{(-1)^k (n-k)!\rho^{n-2k}}{k!(\frac{n+|m|}{2}-k)!(\frac{n-|m|}{2}-k)!}. \tag{9}$$

Based on the Zernike orthogonal polynomials, Teague defined the Zernike moments of a two-dimensional function $f(x, y)$ [43] as

$$Z_{n,m} = \frac{n+1}{\pi}\iint_{x^2+y^2\leq 1} V_{n,m}^*(x, y) f(x, y)\, dxdy \tag{10}$$

where $V_{n,m}^*(x, y) = R_{n,m}(\rho)\exp(-jm\theta)$ denotes the conjugation of $V_{n,m}(x, y)$. For a discrete digital image, its order n and repetition m Zernike moments can be denoted by

$$A_{n,m} = \frac{n+1}{\pi}\sum_{x^2+y^2\leq 1} f(x, y)V_{n,m}^*(\rho, \theta). \tag{11}$$

It is easy to derived that $|Z_{n,m}|$ satisfies rotation invariance. Assume that $f(\rho, \theta)$ is the polar version of image function $f(x, y)$, then the new Zernike moment of f rotated ϕ can be computed as follows

$$\begin{aligned} Z_{n,m}^{new} &= \frac{n+1}{\pi}\iint_{\rho\leq 1} V_{n,m}^*(\rho, \theta) f(\rho, \theta + \phi)\rho\, d\rho d\theta \\ &= \frac{n+1}{\pi}\iint_{\rho\leq 1} V_{n,m}^*(\rho, t - \phi) f(\rho, t)\rho\, d\rho dt \\ &= Z_{n,m}\cdot\exp(jm\phi) \end{aligned} \tag{12}$$

It means that $|Z_{n,m}^{new}| = |Z_{n,m}|$.

Before computing the Zernike moments of a given image, all the pixels should be mapped into a unit circle centered the centroid of the image. Zernike moments and Zernike polynomials have been widely applied in the fields of image processing and pattern recognition [4], [7], [32]. Some Zernike moments will be selected and applied to improve the feature descriptors in next section.

III. A NOVEL LOCAL DESCRIPTOR

It is the most important idea of SIFT that LoG can be approximated by DoG. After searching extreme points in DoG space, dominant direction computing and neighborhood gradient statistics can be achieved for stable key-points and efficient feature descriptors. It provides effective support for feature matching, image retrieval and some other researches and has been widely applied in the field of computer vision [2], [53], [54].

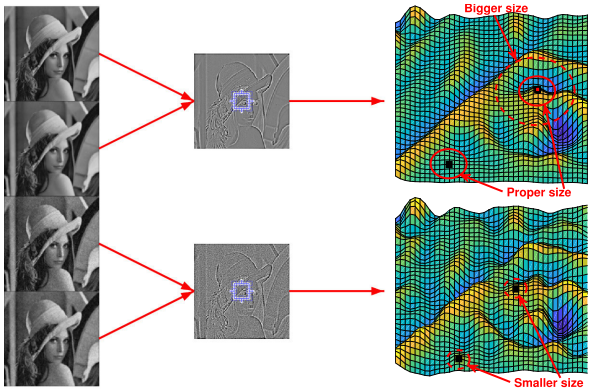
However, key-points detection of traditional SIFT relies on the fixed $3 \times 3 \times 3$ neighborhood, the computing is easy to be interfered by the noise or error, so that the position of the key-points may be seriously biased or even lost. When a dominant direction calculated, eight-direction statistics is implemented in the 3σ circular neighborhood and it may cause an error which will be $\pi/8$ in the worst case.

In this section, an adaptive area including correct local information is introduced to determine the key-points for better accuracy and stability of the key-points. Skew distribution fitting is applied to compute the dominant direction for better describing the main changes in the local area. Then some

Zernike moments are applied to enhance the rotation invariance. A novel adaptive local descriptor embedded Zernike moments can be finally contributed.

A. ADAPTIVE NEIGHBORHOOD FOR KEY-POINTS

A fixed-size neighborhood can be used to detect local extremum points in the traditional SIFT and it is advantageous for simplifying the algorithm and reducing the computation. However, it might affect the accuracy of key-points when the size is not proper or noise exist. If the size was set to be bigger, some true key-points might be cover by some more remarkable extreme points nearby and more computation is required. If it was set to be smaller, some key-points affected by the noise nearby might be mistaken while some false key-points detected.



(a) A DoG example (b) Local

FIGURE 4. Adverse effects of fixed-size neighborhood.

Though the neighborhood used to check key-points in SIFT is a cubic, we will illustrate the improved idea in 2-dimension. A local DoG of original Lena image and the same local of the polluted version are marked in Figure 4(a). They are both magnified and shown in Figure 4(b). There are both two candidate key-points in each local image. The middle-right candidate of original Lena DoG is a true one while the bottom-left is a false. The first one will be accepted if the neighborhood is set properly while be rejected if a oversize neighborhood is set. And the latter will be recognized correctly in a wide range of neighborhood size because of no noise effects. In the local DoG of polluted Lena image, two false candidates will be both mistaken as true ones if smaller neighborhood set because of the noise effects.

It is natural to introduce a dynamic pattern for neighborhood size to reduce the adverse effects above mentioned. The neighborhood size of a candidate key-point will be determined adaptively by the corresponding scale. Such a scale-dependent neighborhood is more advantageous to recognize the key-points correctly. The candidate key-points set can be formulated as

$$E_\sigma = \{\mathrm{x}, D(\mathrm{x}) \geq D(\mathrm{y}) \forall \mathrm{y} \in U_{\alpha\sigma}(\mathrm{x})\}. \tag{13}$$

Then the final key-points set can be given as

$$K_\sigma = \{\mathrm{x} \in E_\sigma, \ |-(\frac{\partial^2 D}{\partial \mathrm{x}^2})^{-1}\frac{\partial D}{\partial x}| < 0.5 \wedge |D(x)| < 0.03 \wedge \frac{Tr(H)^2}{|H|} < \frac{(r+1)^2}{r})\}. \tag{14}$$

Here $\mathrm{x} = [x, y, \sigma]^T$ denotes the image coordinate of a candidate key-point and $U_{\alpha\sigma}(\mathrm{x}) = \{\mathrm{y}, |\mathrm{y} - \mathrm{x}| \leq \alpha\sigma\}$. α means a factor link the neighborhood size and the scale σ. In this paper, it is set to be 3. D and H denote the DoG operator and Hessian matrix.

B. FITTING DOMINANT DIRECTIONS

In the traditional SIFT algorithm, the dominant direction is determined by the neighborhood gradient statistic. Eight candidate directions are uniformly distributed in $[0, 2\pi)$ and the gradient on each inner pixel of the neighborhood will be classified to one of them. Gradients in each class will be calculated the total amount by Gaussian weight on spatial distance. Then a histogram can be obtained and the peak direction(s) are taken to be dominant direction(s).

Suppose that the computed dominant direction θ is one of the preset eight directions while the idea value might be any one in the range of $(\theta - \pi/8, \theta + \pi/8)$ as shown in Figure 5(a). Therefore, the deviation of the dominant direction might be $\pi/8$ in the worst case.

In this paper, skew distribution fitting will be introduced to improve the accuracy of dominant directions.

$$f_{skew}(t) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(t-\mu)^2}{2\sigma_n^2}}, & t \leq \mu \\ \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{(t-\mu)^2}{2\sigma_p^2}}, & t > \mu \end{cases} \tag{15}$$

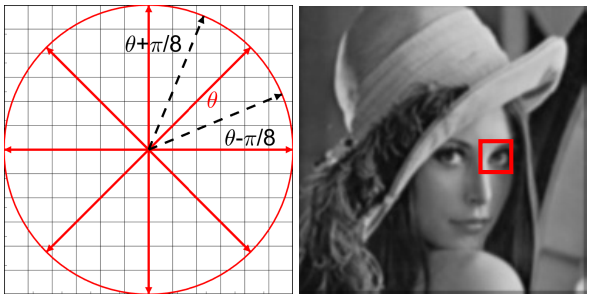
A local area including a key-point of Lena image is shown in Figure 5(b) and the gradient field shown in Figure 5(c) can be used to compute the dominant direction. Figure 5(d) shows the traditional histogram on eight directions. Figure 5(e) shows our histogram and the skew distribution fitting curve based on equation (15). Our fitting result and the traditional result are both shown in Figure 5(f). It can be found that the new result better describing the main change of the local pixels. The dominant direction at x can be formulated as

$$\theta_\sigma(\mathrm{x}) = \underset{\mu}{argmin} \sum_t |f_{skew}(t, \mu) - h(t, \sigma)|^2. \tag{16}$$

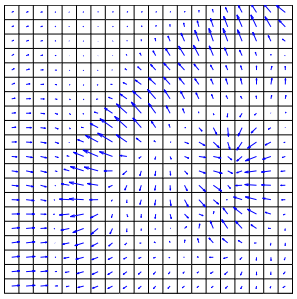
Here $h(t, \sigma)$ means the histogram of the adaptive neighborhood $U_{\alpha\sigma}(\mathrm{x})$.

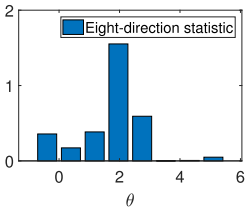
C. EMBEDDING ZERNIKE MOMENTS

Because of the adaptive neighborhood and histogram fitting, key-points and dominant directions can be solved more accurately. The traditional 128-dimensional SIFT descriptor (eight-direction statistics on each grid on a 4×4 mesh) can be constructed more efficiently.

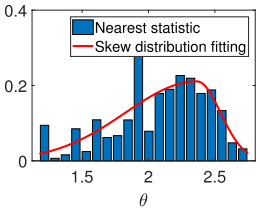


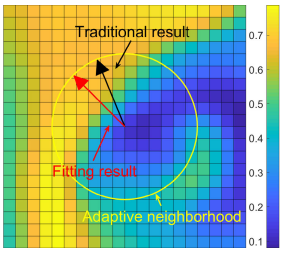
(a) Range (b) Local area





(c) Local gradient (d) Original





(e) New histogram (f) Result

FIGURE 5. Fitting result of a dominant direction.

Firstly, the neighborhood $U_{\alpha\sigma}(\mathrm{x})$ of a key-point x is rotated by the more accurate dominant direction $\theta_\sigma(\mathrm{x})$, and then evenly divided into $b \times b$ subregions. Eight-direction statistics can be implemented in each grid and the size of the resulting feature vector is $8b^2$. As the complexity of the descriptor grows, it will be able to better in a large database, but it will also be more sensitive to some small changes in the local area.

Figure 6 shows experimental results with different mesh width b and image changes. The graph was generated for different cases of a rotation transformation ($2\pi/9$) and additional Gaussian noise (SNR $= 20db$). The graph shows that the results continue to improve up to a 3×3 mesh. After that, increasing the mesh width can actually hurt matching by making the descriptor more sensitive. These results were similar for other image changes and noise, although in some simpler cases continued to improve (from already high levels) with a bigger mesh width. Thus we use a 3×3 mesh as shown in Figure 8(a), resulting in a 72-dimensional feature vector denoted as

$$F_{72} = [h_1, h_2, \cdots, h_9]^T. \tag{17}$$

Here $h_i (i = 1, 2, \cdots, 9)$ means the eight-direction histogram on i-th subregion as shown in Figure 8(a). Though the

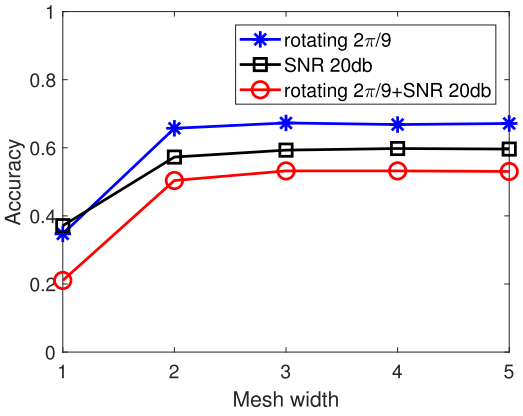
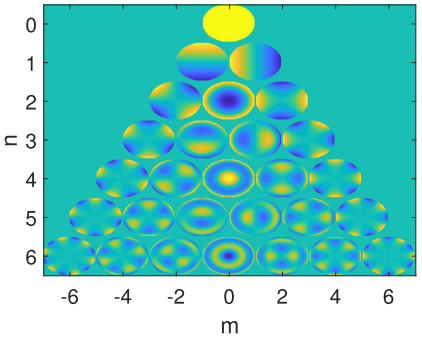
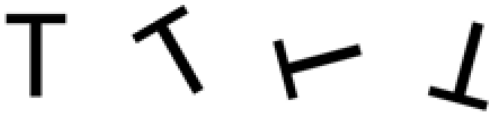


FIGURE 6. Mesh width test.

key-points and the dominant directions have been improved by adaptive neighborhood and skew distribution fitting, the deviation still cannot be eliminated completely. For more robust results, some Zernike moments are introduced to construct a new descriptor with better rotation invariance.



(a) Zernike polynomials



(b) Test images

FIGURE 7. Zernike polynomials and test images.

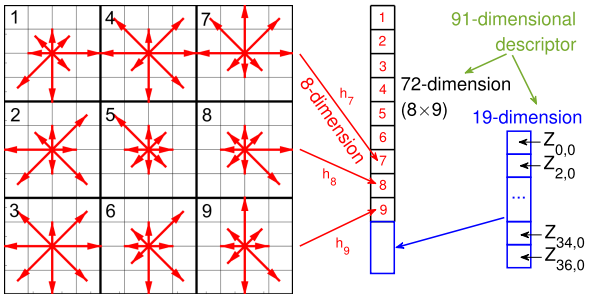
Zernike moments are ideal image feature descriptors with rotation invariance and they have been widely used in target recognition, template matching and some other fields. However, complex form is not advantageous to the computing and the real moments ($m = 0$ and n is even) are selected to contribute the proposed descriptor. For better understanding, Figure 7(a) shows some Zernike polynomials at rank 0 to 6. More detailed, because $Z_{n,m} = \overline{Z_{n,-m}}$, Figure 6(a) shows the real part of Zernike polynomial $V(n, m)$ as $m > 0$ while the image part of $V(n, m)$ as $m < 0$. Specifically, some Zernike moments of a test image are shown in Figure 7(b) and the rotated version are listed in Table 1.

It can be found that there almost no effect on these Zernike moments no matter different rotation angles. It is

TABLE 1. Zernike moments of the test image.

angle	$Z_{0,0}$	$Z_{2,0}$	$Z_{4,0}$	$Z_{6,0}$
0	0.08496234	-0.14817995	0.06997987	-0.03397009
$\pi/6$	0.08496436	-0.14817442	0.06997765	-0.03398700
$13\pi/12$	0.08496474	-0.14817540	0.06997725	-0.03398150
$23\pi/12$	0.08496492	-0.14817542	0.06997683	-0.03398146

natural to embed these Zernike moments in the previous 72-dimensional local feature vector. In most cases, Zernike moments with rank from 0 up to 36 are sufficient to describe the valuable information. In this paper, a set of Zernike moments $\{Z_{2k,0}\}_{k=0}^{18}$ is used to contribute to a 19-dimensional feature vector denoted by F_Z, and then a final 91-dimensional local feature descriptor $F_{ALZ} = [F_{72}, F_Z]$ as shown in Figure 8(b).



(a) 9 subregions (b) Embedding Zernike moments

FIGURE 8. A 91-dimensional descriptor.

D. PROPOSED ALGORITHM

Based on previous sections, the novel descriptor can be computed by Algorithm 1.

Algorithm 1 Adaptive Local Descriptor

Initial I_0, σ, λ, S
for $s = 0$ to S **do**
  for $t = 0$ to $S + 2$ **do**
    Gaussian image $I_{\lambda^s k^t \sigma} = I_s * g_{k^t \sigma}$
    if $t > 0$ **then**
      DoG image $d\ I_{\lambda^s k^{t-1} \sigma} = I_{\lambda^s k^t \sigma} - I_{\lambda^s k^{t-1} \sigma}$
    end if
  end for
  Scale I_s to $\frac{1}{\lambda}$ as I_{s+1}
  for $t = 1$ to S **do**
    (1)Locate each key-point p with adaptive radius $\sigma\lambda^s$
    (2)Skew distribution fitting the orientation $\theta(p)$
    (3)Compute 72-dimensional feature vector $F_{72}(p)$ on a rotated 3×3 mesh
    (4)Compute 19-dimensional Zernike feature descriptor $F_Z(p)$
    (5)Embed Zernike feature $F_{ALZ}(p) = [F_{72}, F_Z]$
  end for
end for

In this paper, $\sigma = 1, S = 4, \lambda = 2$ are set for all the experiments if without specific explanation.

IV. EXPERIMENTS AND RESULTS

In different scenes, real images are often affected by some different geometric and photometric transformations include viewpoint changes, scale changes, image blur, JPEG compression, illumination, and so on. There are ten images (five are natural 256×256 images and five are remote sensing images with different size) for features detecting and matching test on rotation, shift, noising and some other degradation. Then ten image pairs (five are natural image pairs and five are remote sensing image pairs) are prepared for comprehensive matching test. The experiments had been completed under Window 7 and with Matlab R2017b. For well evaluating the performance, some metrics have been introduced in previous literature [24], [25]. In this paper, the proposed algorithm is compared with seven commonly used and well recognized algorithms, namely, Harris, SIFT, BRISK, SURF, FREAK, LBP and LTP on three measurements, correct matching number *num.*, accuracy *acc.*, cost time per feature τ (it means the sum of the average detecting time and average describing time for one feature).

A. EX.1 MATCHING TEST FOR FIXED ROTATING AND SCALING

The five images (Lena, Boat, Couple, Peppers, Hill) are all rotated by $\frac{4}{15}\pi$ after they had been scaled to 70%. Then different algorithms are all implemented on each pair. Detected feature points and matching results are shown in Figure 9. Table 2 shows the correct matching number *num.*, accuracy *acc.* (%) and cost time per feature τ (millisecond, ms). The best scores in each column are set bold. The line charts of correct matching number and accuracy are shown in Figure 15(a).

With the well working of the adaptive neighborhood and dominant direction fitting in proposed method, more stable feature points can be detected and described more accurately. Then it is advantageous to found more correct matches than other seven methods in each image pair. The proposed method achieves best accuracy (about 97.18% in average) and maximum correct matching number (about 48 in average). SIFT, BRISK and SURF catch correct matches more than other fours.

SIFT, BRISK and SURF achieve moderate scores on accuracy (higher than 70%). LBP and LTP only achieve less than 40% in average because some information have been discarded by the simplified binary pattern or ternary Pattern. These two methods are more beneficial to achieve some large scale matching task with a large number of texture features such as face recognition.

FREAK achieves a best performance (less than 0.2) on cost time per feature. The proposed method achieves a moderate performance about 1.16, more detailed, it costs less than SIFT, LBP and LTP but more than other four.

TABLE 2. Matching results of Ex.1 (rotating $\frac{4\pi}{15}$ and scaling 70%).

Method	Lena			Boat			Couple			Peppers			Hill		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	11	55.00	0.21	4	80.00	0.20	5	71.43	0.18	13	54.17	0.28	1	33.33	0.15
SIFT	31	91.18	2.14	23	95.83	1.67	21	87.50	1.64	28	82.35	1.80	29	93.55	1.56
BRISK	30	93.75	0.72	34	94.44	0.36	26	89.66	0.37	16	84.21	1.09	33	94.29	0.56
SURF	17	77.27	0.33	20	83.33	0.22	18	66.67	0.22	20	83.33	0.28	24	85.71	0.23
FREAK	9	40.91	**0.19**	6	75.00	**0.15**	5	62.50	**0.15**	22	52.38	**0.22**	2	50.00	**0.13**
LBP	18	60.00	3.69	3	17.65	3.31	9	34.62	3.15	9	40.91	3.37	7	29.17	2.95
LTP	18	47.37	3.44	2	10.53	2.93	7	29.17	2.90	11	34.38	3.04	8	28.57	2.73
proposed	**48**	**100**	1.28	**52**	**98.11**	1.11	**48**	**96.00**	1.10	**48**	**94.12**	1.24	**45**	**97.83**	1.11

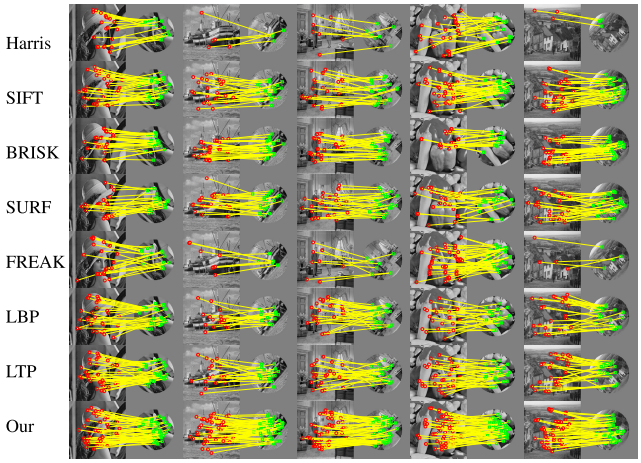


FIGURE 9. Matching test for rotating $\frac{4\pi}{15}$ and scaling 70% (Ex.1).

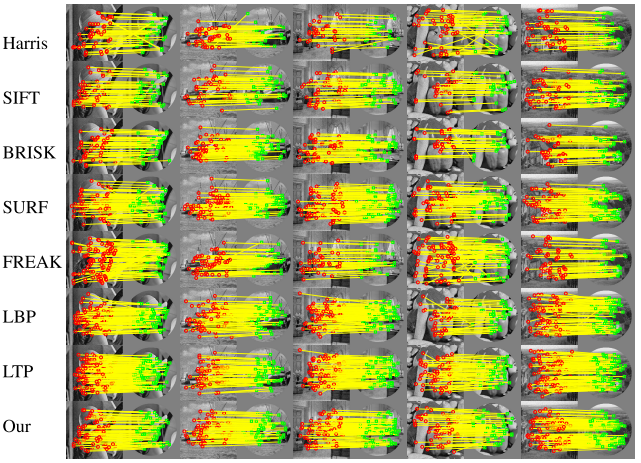


FIGURE 10. Matching test for noise 20*db* and scaling 90% (Ex.2).

B. EX.2 MATCHING TEST FOR NOISE AND FIXED SCALING

The five images (Lena, Boat, Couple, Peppers, Hill) are added Gaussian noise (SNR $=$ 20*db*) for the first images of the pairs. The second image of each pair is generated by scaling the original image to 90%. To achieve a stable and convinced performance measurement, the proposed algorithm and the compared algorithms are all implemented 10 times on each image pair. Figure 10 shows the matching results in one test and Table 3 gives the detailed average results. Figure 15(b) and Figure 16(b) show the line char of correct matching number and accuracy.

It can be found the proposed method helped to capture more correct matches (about 68 in average) with high accuracy for the effective working mechanism. The robust to noised images of proposed method is verified in some sense. Comparing to Ex.1, though the noise is added while the scaling effect is much weakened (the scale rate is changed from 70% to 90%), and finally contributes to a better performance of each method. In general, the proposed method achieves an average accuracy 98.58% increasing from 97.18% in Ex.1. SIFT, SURF and BRISK achieve moderate scores on accuracy (higher than 70%). LBP and LTP achieve lower than 60% in average and the others achieve higher than 70%.

Harris, SURF and FREAK achieve better performance on the cost time per feature τ. SIFT, BRISK and the proposed method achieve a moderate performance.

C. EX.3 MATCHING TEST FOR FIXED SCALING AND DIFFERENT ROTATING

The original Lena image is rotated by different angles $(\frac{\pi}{9}, \frac{17\pi}{36}, \frac{7\pi}{9}, \frac{7\pi}{6}, \frac{14\pi}{9})$ after being scaled to 85%. Matching results are shown in Figure 11 and details are given by Table 4. It can be found that the proposed method catches more correct matches with better accuracy than other methods no matter different rotating angles. Figure 15(c) and Figure 16(c) show the correct matching number and accuracy.

The proposed method achieves accuracy 95.50% and correct matching number 59 in average, it has significant advantages over other methods. It is partly verified that better rotation invariance of the novel descriptor embedding Zernike moments.

Harris, SURF and FREAK cost less time on each feature than the other methods. Our method and BRISK achieve moderate performance.

D. EX.4 MATCHING TEST FOR NOISE, SCALING AND DIFFERENT ROTATIONS

The original peppers image is rotated different angles $(\frac{\pi}{9}, \frac{17\pi}{36}, \frac{7\pi}{9}, \frac{7\pi}{6}, \frac{14\pi}{9})$ after being scaled to 85%. Then Gaussian noise (SNR$=$ 20*db*) is added for the final version. To achieve a stable and convinced performance measurement, the proposed algorithm and the compared algorithms are all

TABLE 3. Matching results of Ex.2 (noise 20*db* and scaling 90%).

Method	Lena			Boat			Couple			Peppers			Hill		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	38	77.8	0.08	28	82.44	0.08	33	97.63	0.08	33	72.28	0.08	35	90.63	0.07
SIFT	35	84.45	0.83	43	89.63	0.74	49	89.91	0.69	29	83.43	0.86	53	89.93	0.68
BRISK	44	88.26	0.54	59	94.69	0.28	51	95.35	0.28	37	92.91	0.64	49	94.97	0.37
SURF	53	76.91	0.08	65	81.85	0.07	64	85.48	0.06	**49**	73.83	0.07	56	85.24	0.07
FREAK	47	62.95	**0.06**	37	82.14	**0.05**	42	97.42	**0.05**	41	61.55	**0.06**	45	91.53	**0.05**
LBP	40	48.24	1.56	28	47.88	1.55	41	58.45	1.47	20	43.64	1.61	36	51.07	1.39
LTP	60	58.89	1.55	42	59.08	1.47	54	63.38	1.41	36	56.21	1.59	49	59.78	1.37
proposed	**65**	**96.17**	0.75	**69**	**99.56**	0.67	**82**	**99.64**	0.67	**49**	**98.20**	0.72	**75**	**98.94**	0.64

TABLE 4. Matching results of Ex.3 (scaling 85% and rotating different angles).

Method	$\frac{\pi}{9}$			$\frac{17\pi}{36}$			$\frac{7\pi}{9}$			$\frac{7\pi}{6}$			$\frac{14\pi}{9}$		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	48	82.76	0.21	56	81.16	0.21	33	66.00	0.19	36	72.00	0.20	60	90.91	0.21
SIFT	40	93.02	2.78	33	86.84	2.60	20	64.52	2.23	29	70.73	2.14	26	78.79	2.29
BRISK	37	94.87	0.75	41	83.67	0.67	27	75.00	0.63	35	83.33	0.64	46	95.83	0.62
SURF	31	67.39	0.31	52	85.25	0.27	13	56.52	0.48	23	63.89	0.30	29	59.18	0.45
FREAK	52	91.23	**0.20**	58	85.29	**0.16**	28	54.90	**0.18**	42	80.77	**0.17**	59	84.29	**0.17**
LBP	22	55.00	3.22	29	56.86	3.02	25	64.10	2.99	24	63.16	2.97	35	74.47	3.04
LTP	35	70.00	2.80	37	56.06	2.55	20	66.67	2.51	27	64.29	2.47	40	68.97	2.51
proposed	**60**	**100**	1.25	**61**	**96.83**	1.12	**50**	**90.91**	1.13	**60**	**92.31**	1.13	**66**	**97.06**	1.08

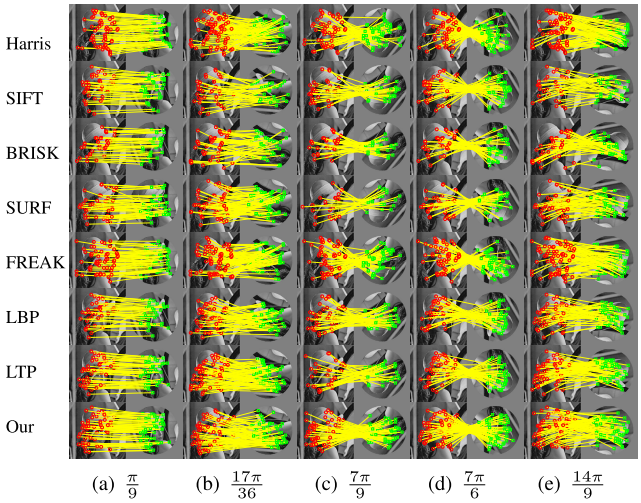


FIGURE 11. Matching test for scaling 85% and rotating different angles (Ex.3).

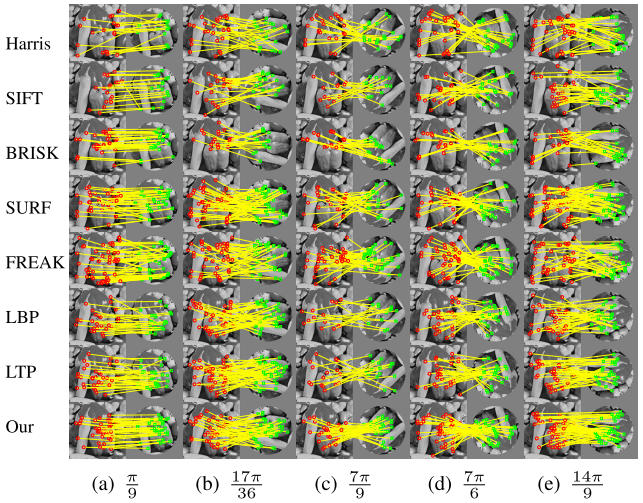


FIGURE 12. Matching test for noise, scaling and rotating (Ex.4).

implemented 10 times on each image pair. Figure 12 shows the detected feature points and matching results in one test. Table 5 gives detailed results. Correct matching numbers and accuracy are shown in Figure 15(d) and Figure 16(d). It can be found that the proposed method is advantageous to accurately catch more right matches even if exist mixed effects consists of noise, scaling and rotating.

Better performance on accuracy (96.03% in average) and correct matching number (39 in average) show that the proposed method has advantages in accurately describing the features over other methods. The well working of the adaptive neighborhood and dominant direction fitting is verified in some sense. Better performance in each angle shows that the rotation invariance is effectively enhanced by embedding Zernike moments.

Comparing to Ex.3, a significant decrease can be found on the correct matching number *num*. (reduced from 59 to 39 in average) because of the additional noise effects. However, only a slight change can be found on the average accuracy (95.50% in Ex.3 and 96.03% here).

Harris, SURF and FREAK still achieve better performance on the cost time per feature τ. BRISK and the proposed method achieve moderate performance while SIFT, LBP and LTP cost more time on each feature for detecting and describing.

E. EX.5 MATCHING TEST FOR SCALED AND ROTATED REMOTE SENSING IMAGES

The original five remote sensing images are all resized to the same size 200 × 200 for more comprehensive test. The polluted version is prepared for the first image of each pair by adding a Gaussian noise with $SNR = 20db$. Different rotating angles ($\frac{\pi}{9}$, $\frac{17\pi}{36}$, $\frac{7\pi}{9}$, $\frac{7\pi}{6}$, $\frac{14\pi}{9}$) are independently taken on the unpolluted images after scaling 75% for the second image.

TABLE 5. Matching results of Ex.4 (noise 20*db*, scaling 85% and rotating different angles).

Method	$\frac{\pi}{9}$			$\frac{17\pi}{36}$			$\frac{7\pi}{9}$			$\frac{7\pi}{6}$			$\frac{14\pi}{9}$		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	18	65.11	0.19	20	71.64	0.20	16	65.34	0.25	17	64.48	0.20	18	66.19	0.22
SIFT	16	71.36	1.32	12	58.67	1.32	10	52.60	1.29	12	55.40	1.31	15	55.34	1.33
BRISK	27	88.78	0.74	23	85.98	0.73	23	92.09	0.76	22	88.66	0.75	25	90.15	0.74
SURF	25	62.47	0.33	33	62.36	0.32	16	64.88	0.38	20	65.00	0.34	32	68.66	0.30
FREAK	24	60.00	**0.14**	24	60.88	**0.13**	23	57.43	**0.15**	21	59.27	**0.13**	21	57.07	**0.15**
LBP	11	37.67	2.19	14	38.25	2.22	8	31.17	2.17	8	34.02	2.19	15	40.80	2.22
LTP	15	48.06	2.02	19	48.14	2.05	9	36.07	2.00	12	40.67	2.04	17	43.03	2.01
proposed	**40**	**98.03**	0.86	**36**	**96.00**	0.89	**36**	**95.45**	0.87	**42**	**96.74**	0.87	**43**	**94.05**	0.86

Matching results in one test are shown in Figure 13 after different methods applied on each pair for 10 times. Detailed results are given in Table 6. The proposed method still achieves better performance on accuracy and correct matching number no matter the varying effects of noise, scaling and different rotating angles. Figure 15(e) and Figure 16(e) show the line charts of correct matching number and accuracy.

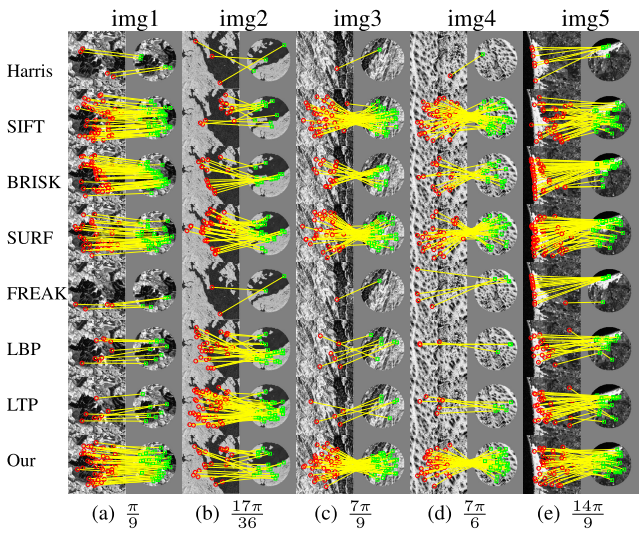


FIGURE 13. Matching test for rotated remote sensing images (Ex.5).

Our method achieves accuracy 97.10% and correct matching number 37 in average. SIFT, BRISK and SURF find more than 20 matches in average while the other four find less than 5 in average. SIFT and BRISK achieve moderate accuracy no less than 80%. LBP and LTP only achieve a poor accuracy less than 30%.

Harris, BRISK, SURF and FREAK achieve better performance on the cost time per feature τ. SIFT and the proposed method achieve moderate performance while LBP and LTP cost more time on each feature for detecting and describing.

F. EX.6 MATCHING TEST FOR DEGRADED IMAGES

The original five images are degraded by different types (scattering on Lena, brushing on Boat and Hill, marking on Couple, penciling on Peppers) and then scaled to 90%. Then the eight methods are independently carried on these image pairs. Detected feature points and matching results are

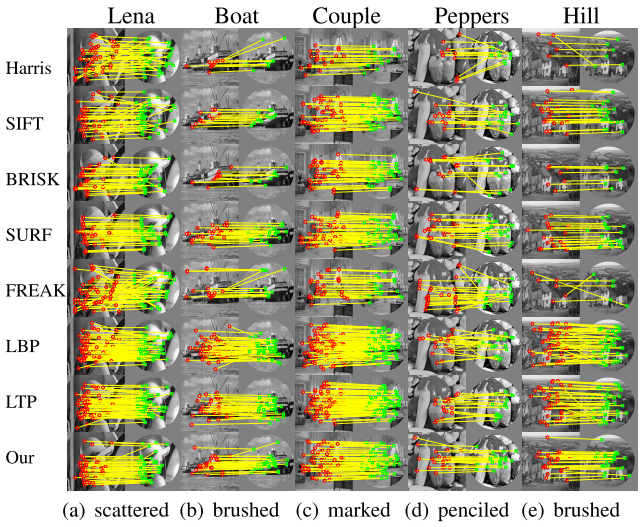


FIGURE 14. Matching test for degraded images (Ex.6).

shown in Figure 14 and Table 7 after the test is implemented 10 times. Figure 15(f) and Figure 16(f) show the line charts of correct matching number and accuracy.

It seems that the penciling make the matching more difficult and the accuracy is often lower than other degraded image pair. The correct matches are also reduced. In general, our method achieves the best performance on accuracy (94.62% in average) and correct matching number (35 in average). SIFT, BRISK and SURF achieves moderate accuracy about 85% in average. Harris and FREAK find fewer correct matches in average while LBP and LTP achieve lower accuracy (less than 50%). Our method has some advantageous to accurately detect more matches in these degraded image pairs.

G. EX.7 MATCHING TEST FOR NATURAL IMAGE PAIRS

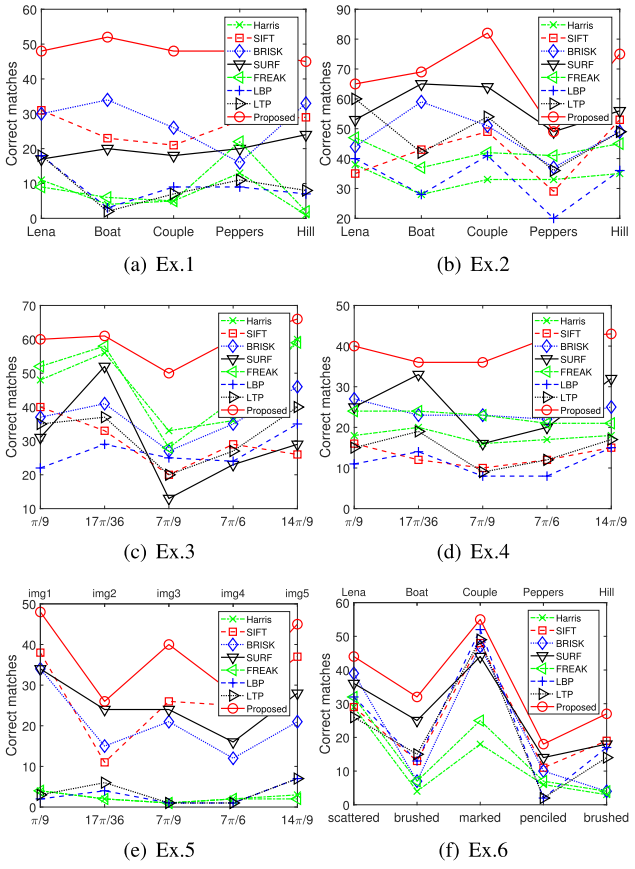
Five pairs of remote sensing images (named a1 to a5) are shown in Figure 17(a) and five pairs of natural images (named b1 to b5) are shown in Figure 17(b). Each pair is affected by one or more of viewpoint changes, scale changes, image blurring, illumination changes. To well distinguish the features from noise, we also set $S = 4$ in this experiment, it means four effective levels in an octave. After the proposed method and the compared methods are implemented on these image

TABLE 6. Matching results of Ex.5 (scaled and rotated remote sensing images).

Method	img1, $\frac{\pi}{9}$			img2, $\frac{17\pi}{36}$			img3, $\frac{7\pi}{9}$			img4, $\frac{7\pi}{6}$			img5, $\frac{14\pi}{9}$		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	4	94.87	0.25	2	87.50	0.35	1	83.33	0.20	2	88.24	0.19	3	32.91	0.26
SIFT	38	91.28	0.98	11	72.48	1.27	26	77.31	0.95	25	70.86	0.96	37	89.13	1.01
BRISK	34	95.22	0.31	15	86.90	0.74	21	88.03	0.20	12	80.26	0.22	21	58.40	0.41
SURF	34	88.63	**0.15**	24	75.86	0.23	24	60.55	0.14	16	59.92	**0.12**	28	56.22	0.20
FREAK	4	86.96	0.16	2	76.67	**0.18**	1	90.91	**0.12**	2	62.50	0.14	2	19.00	**0.16**
LBP	2	21.52	2.09	4	14.98	2.48	1	29.27	1.95	1	14.08	2.10	7	32.60	2.04
LTP	3	29.91	2.10	6	13.25	2.22	1	16.67	1.98	1	12.12	2.05	7	24.91	1.89
proposed	**48**	**100**	0.84	**26**	**92.55**	0.92	**40**	**97.58**	0.85	**28**	**92.38**	0.81	**45**	**99.56**	0.81

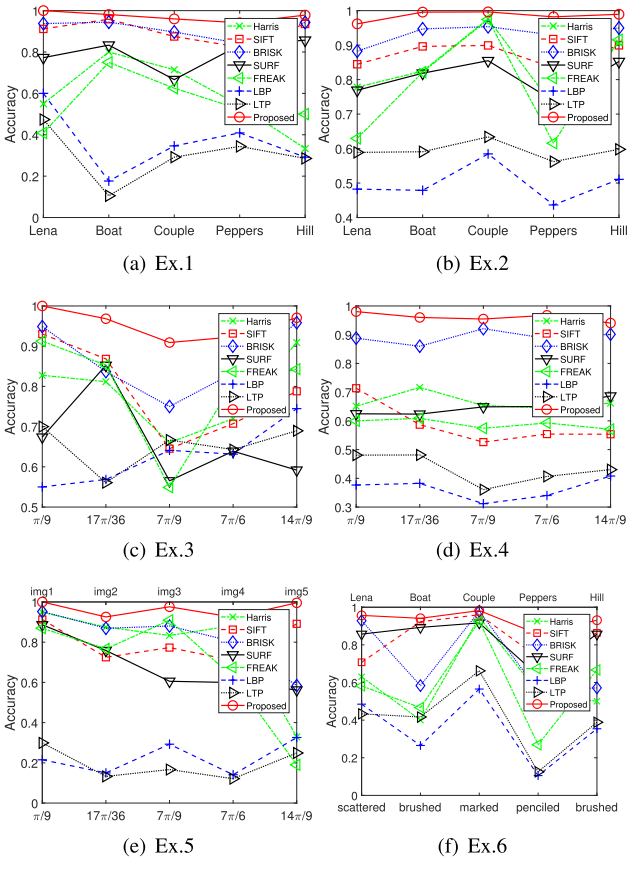
TABLE 7. Matching results of Ex.6 (degraded images).

Method	Lena,scattering			Boat,brushing			Couple, marking			Peppers,penciling			Hill,brushing		
	num.	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ	*num.*	*acc.*	τ
Harris	29	63.04	0.1	4	40	0.13	18	94.74	0.13	6	54.55	0.16	3	50	0.12
SIFT	29	70.73	1	12	92.31	1.37	48	96	0.78	11	73.33	1.18	19	86.36	1.31
BRISK	39	92.86	0.55	7	58.33	0.34	47	97.92	0.29	10	55.56	0.4	4	57.14	0.47
SURF	36	85.71	0.16	25	89.29	0.13	44	91.67	0.13	14	63.64	**0.11**	18	85.71	0.14
FREAK	32	58.18	**0.08**	7	46.67	**0.09**	25	92.59	**0.08**	7	26.92	**0.11**	4	66.67	**0.08**
LBP	32	48.48	1.78	13	26.53	2.38	52	56.52	1.55	2	10.53	2.42	17	35.42	2.51
LTP	26	43.33	1.76	15	41.67	2.73	49	66.22	1.46	2	12.5	2.59	14	38.89	2.85
proposed	**44**	**95.65**	0.74	**32**	**94.12**	1.11	**55**	**98.21**	0.7	**18**	**85.71**	1.01	**27**	**93.10**	1.16



(a) Ex.1 (b) Ex.2 (c) Ex.3 (d) Ex.4 (e) Ex.5 (f) Ex.6

FIGURE 15. Correct matching number in the experiments.



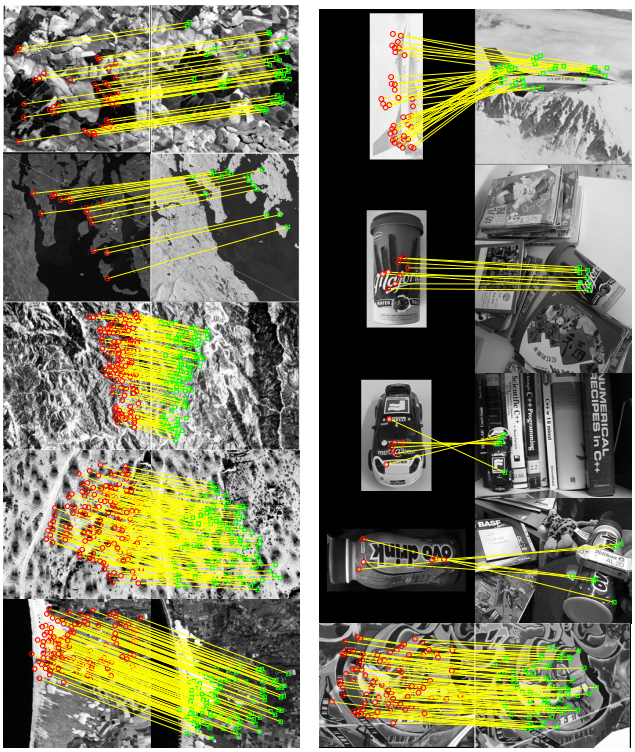
(a) Ex.1 (b) Ex.2 (c) Ex.3 (d) Ex.4 (e) Ex.5 (f) Ex.6

FIGURE 16. Matching accuracy in the experiments.

pairs, correct matching numbers are shown in Figure 18. The matching results of proposed method are shown in Figure 17.

For better illustrating the results, we draw the matching number curve separately in two parts. The left of Figure 18 shows that more matching can be found in image pairs a4, a3, a5, b5 and b1. It is obviously because there are a large number of similar local features with slight change in each of these pairs as shown in Figure 17. a4, a3, a5 are remote sensing image pairs with slight viewpoint change and shift. b5 is a natural image pair with slight viewpoint change and rotation. b1 is a natural image pair with slight illumination change and rotation.

The right of Figure 18 shows that few matches can be caught in other five image pairs a1, a2, b4, b3 and b2. It can be found there are few similar local features or multiple changes in each of these pairs as shown in Figure 17. Remote sensing image pair a1 has few similar features and suffers sight viewpoint change. Remote sensing image pair a2 suffers slight viewpoint change and serious illumination change. Image pairs b4, b3 and b2 are all natural image pairs with serious viewpoint change and rotation.



(a) Remote sensing images (b) Natural images

FIGURE 17. Features matching of ten pairs of images.

In short, with the advantageous effects of adaptive neighborhood, dominant direction fitting and embedding Zernike moments, the proposed method finds more correct matching than others in each pair.

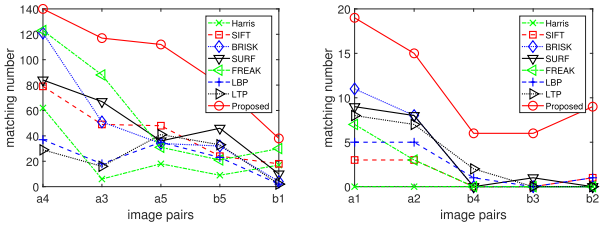


FIGURE 18. Correct matching number in Ex.7.

The seven experiments show that proposed method can achieve better results. The matching number and accuracy can be both improved by the novel method for the introduce of adaptive neighborhood, dominant direction fitting and embedding Zernike moments.

As we found in other experiments, more key-points will be found if a smaller neighborhood set. However, many of them cannot be denoted correctly by the descriptor because of the local information lost, so the matching accuracy will decrease. Such a key-point with insufficient information is often not stable enough and easy to be affected by noise or other factors. The computation of each point will be reduced and the total running time will decrease in some sense.

On the contrary, if a bigger neighborhood set, less key-points can be found while many local extreme points lost. However, such a key-point maybe not accurate described because of the excessive information, so the matching number and accuracy will both decrease. The computation on each key-point will increase and the total running time may increase in some sense.

In a word, the proposed method can achieve better performances on matching number and accuracy. However, there still some should be addressed. Compared to some traditional methods such as SIFT and BRISK, the cost time per feature can be reduced in some sense because of the well works of adaptive neighborhood, dominant direction fitting and embedding Zernike moments. But it still more than some state-of-the-art methods such as SURF and FREAK though the proposed method has advantageous on matching number and accuracy in these cases. We will continue to optimize the computation and Matlab code for more efficient computing in future work.

V. CONCLUSION

In this paper, a novel local features descriptor is proposed based on adaptive neighborhood and embedding Zernike moments. Adaptive neighborhood is introduced to detect key-points more correctly and skew distribution fitting is applied to compute the dominant direction more accurately. Zernike moments are embedded to improve the rotation invariance. Then these contribute to a novel 91-dimensional local adaptive feature descriptor. Experimental results show that the proposed method is advantageous to find more efficient features points and catch more accurate matching than some traditional methods.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive remarks and suggestions which helped them to improve the quality of the manuscript to a great extent.

REFERENCES

[1] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.

[2] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A sift-based forensic method for copy–move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1099–1110, Sep. 2011.

[3] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.

[4] A. Bouziane, Y. Chahir, M. Molina, and F. Jouen, "Unified framework for human behaviour recognition: An approach using 3D zernike moments," *Neurocomputing*, vol. 100, pp. 107–116, Jan. 2013.

[5] H. Chen, J. Feng, B. Zhou, Y. Hu, and K. Guo, "An anisotropic diffusion-based dynamic combined energy model for seismic denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1061–1065, Jul. 2017.

[6] Z. Chen and S.-K. Sun, "A Zernike moment phase-based descriptor for local image representation and matching," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 205–219, Jan. 2010.

[7] A.-W. Deng, C.-H. Wei, and C.-Y. Gwo, "Stable, fast computation of high-order zernike moments using a recursive method," *Pattern Recognit.*, vol. 56, pp. 16–25, Aug. 2016.

[8] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Learning rotation-invariant local binary descriptor," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3636–3651, Aug. 2017.

[9] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: A new feature descriptor for image retrieval in medical CT databases," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5892–5903, Dec. 2015.

[10] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Proc. ISPRS Intercommission Conf. Fast Process. Photogrammetric Data*, interlaken, Switzerland, 1987, pp. 281–305.

[11] Y. Gao, W. Huang, and Q. Yu, "Learning multiple local binary descriptors for image matching," *Neurocomputing*, vol. 266, pp. 239–246, Nov. 2017.

[12] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A comprehensive performance evaluation of 3D local feature descriptors," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 66–89, 2016.

[13] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, vol. 15, no. 50, 1988, p. 10–5244.

[14] Y. He, G. Deng, Y. Wang, L. Wei, J. Yang, X. Li, and Y. Zhang, "Optimization of SIFT algorithm for fast-image feature extraction in line-scanning ophthalmoscope," *Optik*, vol. 152, pp. 21–28, Jan. 2018.

[15] C. Huang, Z. He, G. Cao, and W. Cao, "Task-driven progressive part localization for fine-grained object recognition," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2372–2383, Dec. 2016.

[16] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun./Jul. 2004, p. 2.

[17] S. Korman, D. Reichman, G. Tsur, and S. Avidan, "Fast-match: Fast affine template matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2331–2338.

[18] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2548–2555.

[19] C. Li, J. C. Gore, and C. Davatzikos, "Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation," *Magn. Reson. Imag.*, vol. 32, no. 7, pp. 913–923, Sep. 2014.

[20] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2007–2016, Jul. 2011.

[21] Y. Li, Q. Li, Y. Liu, and W. Xie, "A spatial-spectral sift for hyperspectral image matching and classification," *Pattern Recognit. Lett.*, vol. 127, pp. 18–26, Nov. 2018.

[22] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.

[23] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1368–1381, Mar. 2016.

[24] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Sep. 1999, pp. 1150–1157.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[26] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2041–2056, Oct. 2015.

[27] B. Ma, J. Shen, Y. Liu, H. Hu, L. Shao, and X. Li, "Visual tracking using strong classifier and structural local sparse descriptors," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1818–1828, Oct. 2015.

[28] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[29] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.

[30] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. Int. Conf. Artif. Intell.*, 1977, p. 584.

[31] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Amer.*, vol. 66, no. 3, pp. 207–211, 1976.

[32] M. Novotni and R. Klein, "Shape retrieval using 3D zernike descriptors," *Comput.-Aided Des.*, vol. 36, no. 11, pp. 1047–1062, 2004.

[33] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proc. Int. Conf. Adv. Pattern Recognit.* Berlin, Germany: Springer, 2001, pp. 399–408.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[35] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: A review," *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.

[36] S. Oron, T. Dekel, T. Xue, W. T. Freeman, and S. Avidan, "Best-buddies similarity—Robust template matching using mutual nearest neighbors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1799–1813, Aug. 2017.

[37] M. Rodríguez, J. Delon, and J.-M. Morel, "Fast affine invariant image matching," *Image Process. Line*, vol. 8, pp. 251–281, Sep. 2018.

[38] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.

[39] L. Shi, W. Lou, A. Wong, F. Zhang, J. Abrigo, W. C. Chu, T. C. Kwok, K. K. Wong, D. Abbott, and D. Wang, "Neural evidence for long-term marriage shaping the functional brain network organization between couples," *NeuroImage*, vol. 199, pp. 87–92, Oct. 2019.

[40] C. Singh, E. Walia, and K. P. Kaur, "Color texture description with novel local binary patterns for effective image retrieval," *Pattern Recognit.*, vol. 76, pp. 50–68, 2018.

[41] S. M. Smith and J. M. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, 1997, doi: 10.1023/A:1007963824710.

[42] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[43] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Amer.*, vol. 70, no. 8, pp. 920–930, 1980.

[44] Y. Tian, B. Fan, and F. Wu, "L2-net: Deep learning of discriminative patch descriptor in Euclidean space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 661–669.

[45] S. Todorovic and N. Ahuja, "Region-based hierarchical image matching," *Int. J. Comput. Vis.*, vol. 78, no. 1, pp. 47–66, Jun. 2008.

[46] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013.

[47] R. Wang, Y. Shi, and W. Cao, "GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra," *Pattern Recognit. Lett.*, vol. 127, pp. 11–17, Nov. 2018.

[48] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Inf. Sci.*, vol. 408, pp. 100–114, Oct. 2017.

[49] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognit.*, vol. 92, pp. 64–81, Aug. 2019.

[50] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou, "Selective convolutional descriptor aggregation for fine-grained image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2868–2881, Jun. 2017.

[51] A. P. Witkin, "Scale-space filtering," in *Readings Computer Vision*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 329–332.

[52] X. Zhen, M. Yu, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Descriptor learning via supervised manifold regularization for multioutput regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 2035–2047, Sep. 2017.

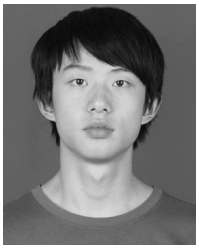
[53] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[54] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comput. Vis. Image Understand.*, vol. 113, no. 3, pp. 345–352, Mar. 2009.
[55] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.

BIN ZHOU received the B.S. degree in mathematics from Shaanxi Normal University, Xi'an, China, the M.S. degree in software engineering from Xi'an Jiaotong University, Xi'an, and the Ph.D. degree in applied mathematics from Sichuan University, Chengdu, China, in 2010. He is currently an Associate Professor with Southwest Petroleum University, China. His research interests focus on applied mathematics, image processing, statistical learning, and intelligent systems.

XUE-MEI DUAN received the B.S. degree in information and computing science from Leshan Normal University, Leshan, China, in 2017. She is currently pursuing the master's degree in mathematics with Southwest Petroleum University, Chengdu, China. Her research interests include mathematical modeling and image processing.

WEI WEI received the M.S. and Ph.D. degrees from Xi'an Jiaotong University, in 2005 and 2011, respectively. He is currently an Associate Professor with the Xi'an University of Technology. His research interests include wireless networks and wireless sensor networks application, mobile computing, distributed computing, and pervasive computing.



DONG-JUN YE received the B.S. degree in mathematics from Southwest Petroleum University, Chengdu, China, in 2017. He is currently pursuing the master's degree with Southwest Petroleum University. His research interests include data mining and image processing.

MARCIN WOŹNIAK received the Diploma degrees in applied mathematics and computational intelligence. He is currently an Associate Professor with the Institute of Mathematics, Silesian University of Technology, Gliwice, Poland. In his scientific career, he visited the University of Würzburg, Germany, the University of Lund, Sweden, and the University of Catania, Italy. His main scientific interest is neural networks with their applications together with various aspects of applied computational intelligence. He is a Scientific Supervisor in the editions of "The Diamond Grant" and "The Best of the Best" programs for highly gifted students from the Polish Ministry of Science and Higher Education. He served as an editor for various special issues, and as an organizer or the session chair at various international conferences and symposiums, including the IEEE SSCI, IEEE FedCSIS, APCASE, ICIST, ICAISC, and WorldCIST.

ROBERTAS DAMAŠEVIČIUS received the B.Sc. degree in informatics and the M.Sc. degree *(cum laude)* from the Faculty of Informatics, Kaunas University of Technology (KTU), Kaunas, Lithuania, in 1999 and 2001, respectively. He also defended his Ph.D. thesis at KTU, in 2005. He is currently a Professor with the Software Engineering Department, KTU, where he lectures robot programming and software maintenance courses. He is the author or coauthor of over 100 articles as well as a monograph published by Springer. His research interests include brain–computer interface, bioinformatics, data mining, and machine learning.