






Article

Heterogeneous Image Matching via a Novel Feature Describing Model

Bin Zhou ^{1,2,3*} , Xuemei Duan ¹ , Dongjun Ye ¹ , Wei Wei ^{4,*}, Marcin Woźniak ⁵ 
and Robertas Damaševičius ^{5,6} 

¹ School of Sciences, Southwest Petroleum University, Chengdu 610500, China; duan_xuem@163.com (X.D.); dongjunye1996@yeah.net (D.Y.)

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; Marcin.Wozniak@polsl.pl (M.W.); robertas.damasevicius@ktu.lt (R.D.)

⁶ Department of Software Engineering, Kaunas University of Technology, 51386 Kaunas, Lithuania

* Correspondence: binzhou@swpu.edu.cn (B.Z.); weiwei@xaut.edu.cn (W.W.)

Received: 30 September 2019; Accepted: 6 November 2019; Published: 9 November 2019



Abstract: Computer vision has been developed greatly in the past several years, and many useful and interesting technologies have been presented and widely applied. Image matching is an important technology based on similarity measurement. In this paper, we propose a novel feature describing model based on scale space and local principle component analysis for heterogeneous image matching. The traditional uniform eight-direction statistics is updated by a task-related k -direction statistics based on prior information of the keypoints. In addition, the k directions are determined by an approximately solution of a Min-Max problem. The principle component analysis is introduced to compute the main directions of local patches based on the gradient field. In addition, the describing vector is formed by then implementing PCA on each sub-patch of a 4×4 mesh. Experimental results show the accuracy and efficiency of proposed method.

Keywords: main direction; principle component analysis; scale space; heterogeneous; gradient field; image matching

1. Introduction

Computer vision techniques have been widely applied in many areas such as object tracking, medical image analysis, and pattern recognition. Image matching can help to find the similar content in different images by analyzing the pixel values and their potential features [1–12]. Traditional image matching techniques are developed based on different principles and information using skills. Some of them directly use the original pixel values in a certain region, and it is easy to operate and work well in some simple scenes [13–17]. It often requires more computation costs and is sensitive to the geometrical differences. Some others process the matching computing based on features extraction instead of directly on the pixel values [2,3,6,8,14,18]. Invariance can be achieved or partly achieved in some procedures such as rotation, resizing, translation, and so on. After the features are detected, matching computing will be processed on a certain similarity.

Some local feature detection methods have been presented in the past years, and it was put forward firstly by Moravec [13] and Harris [1]. Moravec and Harris introduced auto-correlation function and auto-correlation matrix independently to find feature points (simple corners).

However, for the extracted feature points, the former is not adaptive to rotation and noise, the latter is adaptive to rotation and illumination, does not have scale invariance. Shi and Tomasi

improved the Harris, and put forward the Shi–Tomasi [19], which makes the distribution of extracted feature points more uniform and reasonable. It is computationally expensive and the feature points are not scale invariant. SUSAN (smallest univalue segment assimilating nucleus) is proposed to determine corners by the number of pixels in the preset region. The algorithm is small in computation and does not have scale invariance [4]. Then, a machine learning method is used to compute FAST (features from accelerated segment test) [20]. It can calculate faster, but when there are noise in the image, there will be more error feature points. Based on FAST, Mair put forward AGAST (adaptive and generic corner detection based on the accelerated segment test), which effectively improves the speed of feature point extraction [21]. However, none of them have scale invariance. KAZE descriptor used a nonlinear diffusion filter to construct a stable nonlinear scale space. It takes a long time to construct the nonlinear scale space, and more octaves can reduce the efficiency of KAZE [22]. With the rapid development of machine learning, some new methods, including sparse coding and convolutional neural network (CNN), are applied to feature extraction [23–30]. Sparse coding is a linear combination that reconstructs the input data into a set of hypercomplete basis vectors. The new features of input data are composed of the coefficients of base vectors. Sparse coding is very slow in practice. The eigenvalue is obtained by a convex optimization method in the test phase. From the structure perspective, the network can use the increased nonlinearity to get the approximate structure of the objective function, and at the same time get a better representation of the characteristics by increasing the depth of CNN. However, it also increases the overall complexity of the network, making it difficult to optimize the network and easy to over fit.

Scale-Invariant Feature Transform (SIFT) is proposed to improve the features invariance of previous methods [3,6]. Image pyramid and local gradient information statistics are introduced to compute the 128-dimensional features. It can work well even if there exist some changes in scale, brightness, viewing angle, etc. However, the computation and the matching error rate are still considerable. Some techniques based on SIFT have been developed over the past years. Ke and Sukthankar introduce principal component analysis to replace the weighted histograms and reduce the dimension of features to 20 [7]. Speeded up robust features (SURF) are presented based on Hessian matrix and image convolutions to faster the computing [18]. There are some other techniques that have also been introduced to achieve better performance [9,31–34].

However, heterogeneous image matching is much different from natural image matching [35–39]. Some vision characteristics may change seriously by the imaging condition. The related features describing will also change greatly. It can be found that many traditional methods do not work in heterogeneous image matching.

This paper aims to present a novel local features describing model more advantageously for heterogeneous image matching. (1) The prior information of some keypoints is used to determine k adaptive directions instead of the traditional uniform eight directions. A gradient histogram is computed on these new directions and forms the first part of the describing vector. (2) The main direction of a local area is computed by carrying out PCA on the gradient fields instead of searching the extreme of a gradient histogram. (3) PCA is independently carried out on 16 sub-patches of the local area and the first components are integrated to the second part of the describing vector. After the final $k + 32$ dimensional descriptor formed, better performances can be achieved on some matching tasks by the proposed method with respect to some traditional methods.

The rest of this paper is organized as follows: in Section 2, the related fundamentals of scale space and principle component analysis are prepared. A novel feature descriptor based on PCA is proposed in Section 3. In Section 4, several experiments are implemented to verify the accuracy and efficiency of proposed algorithm.

2. Fundamentals

2.1. Scale Space and Difference Space and Principal Component Analysis

Scale space theory [40] has been introduced in many models and it can help to distinguish the features in different levels. After a scale parameter added to the related model, a multi-scale representation sequence can be computed following the changing of the parameter. Many tasks can be achieved by analyzing or processing the sequence such as extracting features.

The classical linear scale space can be generated by a diffusion equation as shown in Equation (1), and it can be regarded as the observation of the image at different distance/scale:

$$\begin{cases} \frac{\partial I}{\partial t}(x, t) = c \nabla^2 I(x, t), & (x, t) \in \Omega \times \Sigma, \\ I(x, 0) = I_0(x), & x \in \Omega, \end{cases} \quad (1)$$

where t denotes the scale parameter and c denotes the diffusion constant. $I(x, t)$ means the a scale space and $\Omega \subset R^2$ is the image domain. This partial differential equation can be solved by some classical techniques such as Fourier transform and many related models have been presented for different image processing problems [41–43]:

$$I(x, t) = I(x) * \frac{1}{4\pi t^2} \exp\left(-\frac{|x|^2}{4t^2}\right). \quad (2)$$

It means that the scale space can be denoted as the convolution of image function $I(x)$ and a Gaussian kernel function $g_{\sqrt{2t}}$ with different t , so-called Gaussian scale space. Then, the discrete difference of Gaussian space can be denoted as

$$d I(x, t) = I(x, t + dt) - I(x, t). \quad (3)$$

Here, dt can be set according to the image data. Figure 1 shows two Gaussian images and a difference image of Lena.

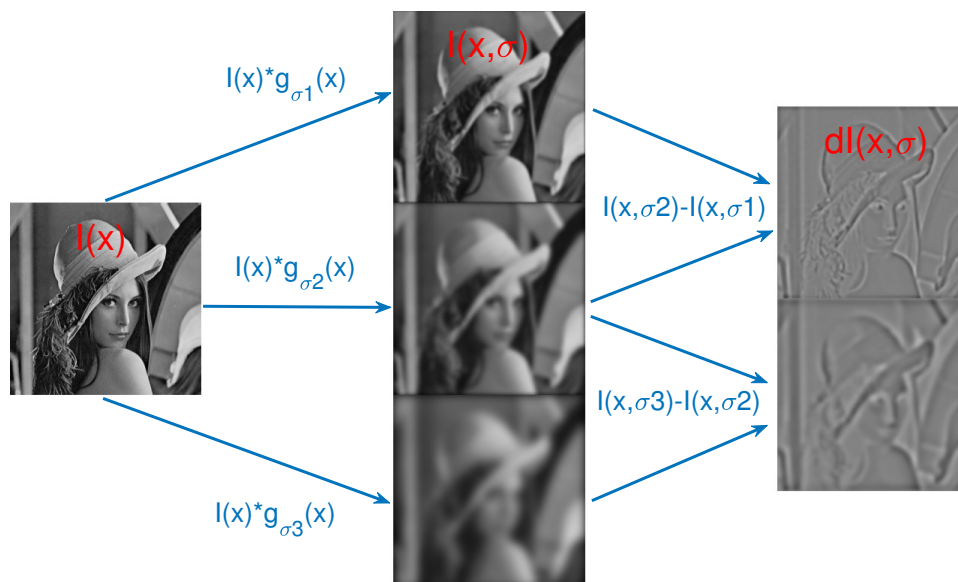


Figure 1. Gaussian scale space of Lena image.

Principal Component Analysis (PCA) [44] is a common technique for dimensionality reduction and has been widely applied to many computer vision problems such as feature selection, object recognition, and so on. The main ideas of PCA can be summarized as follows. Assume $n \times d$ matrix Y

is the normalized observation data of random variables Y_1, Y_2, \dots, Y_d . Then, the d -dimensional data can be mapped into an orthogonal space for a maximal variation in each new dimension. The orthogonal space can be computed by applying matrix eigen value theory on the covariance of Y :

$$Y^T Y v_i = \lambda_i v_i, \quad i = 1, 2, \dots, d. \tag{4}$$

The components can be denoted as $V_i = Y v_i$ with the assumption $\lambda_1 > \lambda_2 > \dots > \lambda_d$ and v_i means the direction with maximal projection variation of original data. Then, the first k components can be used for dimension reduction of the original data, and k can be determined by specific needs.

Though PCA suffers from some restriction and shortcomings, it is still popular due to its simplicity. Fergus et al. introduce PCA to unsupervised scale-invariant learning for object recognition [45]. This technique is also applied to represent keypoint patches and improves SIFT's matching performance [7], named PCA-SIFT.

2.2. SIFT and PCA-SIFT

Based on the scale space and the difference space, the original Scale-Invariant Feature Transform (SIFT) can be described as four main steps.

1. Scale-space peaks selection.

The extreme points in the DoG (difference of Gaussian) space are detected by comparing the pixels in a $3 \times 3 \times 3$ neighborhood, named non-maximum suppression. Let $D(x)$ and $U_r(x)$ denote the DoG function and r -neighborhood of x ; then,

$$EPTs = \{x, D(x) \geq D(y) \forall y \in U_1(x)\}. \tag{5}$$

2. Keypoint localization.

Taylor expansion and principle curvature are both applied to check the detected extreme points again and then keypoints can be located. Let H denote the Hessian of DoG function $D(x)$; then,

$$KPTs = \{x \in EPTs, | -(\frac{\partial^2 D}{\partial x^2})^{-1} \frac{\partial D}{\partial x} | < 0.5 \wedge |D(x)| < 0.03 \wedge \frac{Tr(H)^2}{|H|} < \frac{(r+1)^2}{r}\}. \tag{6}$$

Generally, set $r = 10$.

3. Orientation assignment.

To achieve rotation invariance, eight-direction statistics is applied on the gradient field of $A(x)$, the local patch centered at x . Then, the orientation can be computed by searching the maximal direction of the histogram.

$$\theta(A) = \operatorname{argmin}_{\text{hist8}(\theta_k, \nabla A)}, \quad k = 1, 2, \dots, 8. \tag{7}$$

Here, $\text{hist8}(\theta_k, \nabla A)$ means the value in $[\theta_k - \frac{\pi}{8}, \theta_k + \frac{\pi}{8})$ when the histogram was carried out on a local gradient patch $\nabla A(x)$. θ_k is one of the eight uniform directions.

4. Keypoint descriptor.

Then, eight direction statistics is applied on each subarea after the local gradient field has been rotated by θ_0 and divided into 4×4 sub-areas. Let $B(i, j)$ denote the subareas of rotated $A(x)$; then, a 128-dimensional descriptor can be obtained by arranging them as follows:

$$F_{SIFT} = [\text{hist8}(\theta_k, \nabla B_{i,j})]_{4 \times 4 \times 8}. \tag{8}$$

The PCA-SIFT is similar to the SIFT in many details and they only differ in Step 4. It can be summarized as follows.

1. An eigenspace of the image gradient fields is pre-computed.

The gradient fields on a 41×41 patch centered at the keypoint can be rearranged to a $2 \times 39 \times 39 = 3042$ -dimensional vector. In addition, it is normalized to unit magnitude for reducing the unexpected impact of variations in illumination. Such about 21,000 patches will be collected and then principle component analysis is processed. The first 20 components will be selected to generate the projection matrix for any other observation. Let C denote the normalized version of collected patches; then, the projection matrix for any path can be computed as

$$M_{proj} = [v_1, v_2, \dots, v_{20}]. \tag{9}$$

Here, $C^T C v_i = \lambda_i v_i$, $i = 1, 2, \dots, d$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$.

2. A simplified descriptor is used on the eigenspace.

A 3042-dimensional vector z on a given image patch centered at x will be computed and normalized, and then projected into the pre-computed feature space as denoted by Equation (10). A 20-dimensional vector can be used to describe the feature point. Figure 2 shows some keypoints and patches of the Lena image. Figure 2b shows the gradient filed on the yellow patch. Figure 2c shows the pre-computed feature space to be dimension reduced:

$$F_{PCA-SIFT} = z M_{proj}. \tag{10}$$

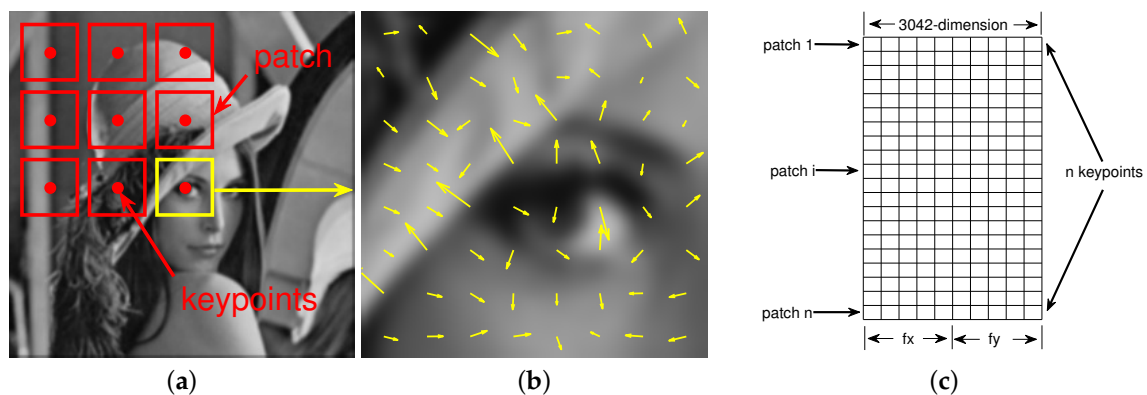


Figure 2. Keypoints and the patches: (a) some keypoints and the patches centering them; (b) the gradient field on a patch; (c) a 3042-dimensional feature space for dimension reduction.

After then, the Euclidean distance can be used to determine whether the two vectors mean the same keypoint in different images. In more detail, a proper threshold ρ_0 is presented for the ratio between the best and second-best match as in the SIFT. Let I_a and I_b be the two images, $x \in I_a$ and $y \in I_b$ are keypoints, and then the matching between x and y can be determined based on PCA-SIFT features $F_{PCA-SIFT}(x)$ and $F_{PCA-SIFT}(y)$

$$IsMatch(x, y) = \begin{cases} 1, & \rho > \rho_0, \\ 0, & \rho \leq \rho_0. \end{cases} \tag{11}$$

Here,

$$\rho = \max \left\{ \frac{dist(F_{PCA-SIFT}(x), F_{PCA-SIFT}(t))}{dist(F_{PCA-SIFT}(x), F_{PCA-SIFT}(y))}, t \in I_b \wedge t \neq y \right\}. \tag{12}$$

It is obvious that a smaller ρ_0 means fewer false positives and more false negatives. On the contrary, a bigger ρ_0 leads to more false positives and fewer false negatives. A proper threshold ρ_0 is advantageous to an appropriate trade-off between false positives and false negatives.

Different from the features represented by histogram of local gradient fields in traditional SIFT, PCA-SIFT represents the features in the eigenspace. The descriptor can be computed more efficiently after the eigenspace pre-computed based on sufficient prior keypoints. Speeded up robust features (SURF) detect the extremes in the difference of Hessian (DoH) instead of DoG. Then, the downsampling in the Gaussian pyramid is replaced by a box-filter with some fixed size, and the eight-direction statistics is replaced by computing a Harr wavelet response. However, the strong dependence on local gradient and discrete scale in DoH should still be addressed in some cases. KAZA constructs a nonlinear scale space via a nonlinear diffusion filter for better matching performance in complex cases. In addition, the additive operator splitting (AOS) scheme requires more computation.

3. Proposed Method

The original PCA-SIFT describes the feature via the first 20 components of collected patches centered known keypoints. It is advantageous for the linear structure, but some information of nonlinear structure will be ignored. In addition, pre-computing should be carried out on considerable classical images/patches for good representativeness. Furthermore, 20 components may be not sufficient in some scenes for the information lost.

In our opinion, PCA is more advantageous to achieve local representativeness than the global, and it is robust to some image changes such as noise, rotation, and so on. Thus, we introduce it to describe local information and then form a novel feature describing model.

Moreover, the traditional uniform eight directions in $[0, 2\pi)$ are easy to be used to describe the local gradient distribution centered at a keypoint. However, various distributions can be found in different tasks indeed. Task-related directions will be more beneficial to achieve accurate features and correct matches. The prior information of keypoints can be used to determine the directions/intervals which were used to implement gradient statistics.

Heterogeneous images are often captured by difference imaging devices/methods on a same scene/object. Figure 3 shows two types of heterogeneous images and their local gradient fields. The traditional image changes such as noise, rotation, and scaling can be regarded as specific heterogeneity. Heterogeneous features or keypoints can be explained in a similar way. It can be found that the local gradient fields of heterogeneous images are very different from the original. Traditional methods based on gradient statistics often fail to achieve matching tasks on such heterogeneous images. This paper aims to present a novel feature describing model that is advantageous to address some heterogeneous images matching tasks.

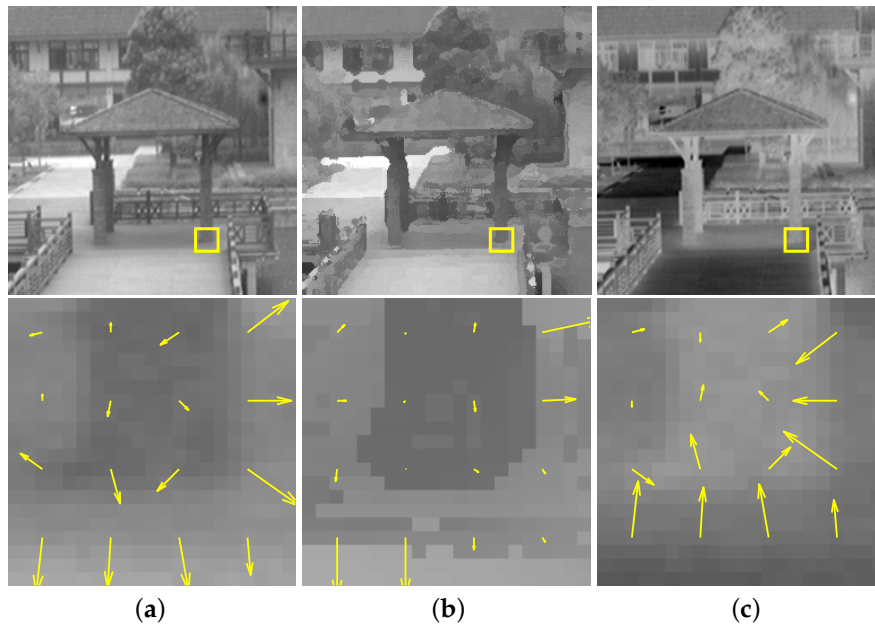


Figure 3. Two types of heterogeneous images: (a) original; (b) heterogeneous type 1; (c) heterogeneous type 2.

3.1. Main Steps

Based on previous sections, a novel frame for heterogeneous image matching can be formulated as follows.

1. Gradient histogram and adaptive intervals.

According to the matching task, some prior information of keypoints can be helpful to determine proper candidate intervals for computing gradient histogram. Such adaptive intervals are more advantageous to describe the keypoints accurately and distinguish different keypoints than traditional uniform eight-intervals.

The local gradient information of some known keypoints can be integrated as $\nabla I(x), x \in \Omega_{key}$. Ω_{key} means the set of keypoints and $\nabla I(x)$ is the gradient at a keypoint x in the rotated local patch. Then, refined interval statistics is carried out on it, and a weighted histogram can be denoted as

$$h_i = \sum_{|\theta(x)-c_i| \leq \frac{\pi}{m}} |\nabla I(x)|, i = 1, 2, \dots, m. \tag{13}$$

Here, $\{c_i\}_{i=1}^m$ means the centers set of candidate intervals. $\theta(x)$ denotes the direction angle of $\nabla I(x)$. Figure 4 shows the traditional histogram and the weighted version of Lena image as $m = 36$.

Then, an optimization model (14) is introduced to determine a set of adaptive intervals for local gradient statistics:

$$\begin{aligned}
 \min_{k, C_j} \quad & \frac{1}{k} \sum_{j=1}^k \left(\sum_{c_i \in E_j} h_i - \frac{\sum_i h_i}{k} \right)^2, \quad \max_{k, C_j} \quad \frac{1}{k} \sum_{j=1}^k \frac{\sum_{c_j \in E_j^h} h_i}{\sum_{c_i \in E_j} h_i}, \\
 \text{s.t.} \quad & E_1 = \begin{cases} [C_1 - \frac{\epsilon_0}{2}, C_1 + \frac{\epsilon_1}{2}], & C_1 + C_k > \pi, \\ [C_k + \frac{\epsilon_0}{2}, \frac{3\pi}{2}] \cup [-\frac{\pi}{2}, C_1], & C_1 + C_k \leq \pi, \end{cases} \\
 & E_k = \begin{cases} [C_k, \frac{3\pi}{2}] \cup [-\frac{\pi}{2}, C_1 - \frac{\epsilon_0}{2}], & C_1 + C_k > \pi, \\ [C_k - \frac{\epsilon_{k-1}}{2}, C_k + \frac{\epsilon_0}{2}], & C_1 + C_k \leq \pi, \end{cases} \\
 & E_j = [C_j - \frac{\epsilon_{j-1}}{2}, C_j + \frac{\epsilon_j}{2}], E_j^h = [C_j - \frac{\epsilon_{j-1}}{4}, C_j + \frac{\epsilon_j}{4}], j = 2, 3, \dots, k-1, \\
 & \epsilon_0 = \frac{2\pi + C_1 - C_k}{2}, \epsilon_j = C_{j+1} - C_j, j = 1, 2, \dots, k-1, \\
 & -\frac{\pi}{2} \leq C_j < C_{j+1} \leq \frac{3\pi}{2}, j = 1, 2, \dots, k-1.
 \end{aligned} \tag{14}$$

Here, $E_j(j = 1, 2, \dots, k)$ means the adaptive intervals and $C_j(j = 1, 2, \dots, k)$ denotes the centers, k is the number of selected intervals. It can be approximately solved by a dynamic programming algorithm after k preset.

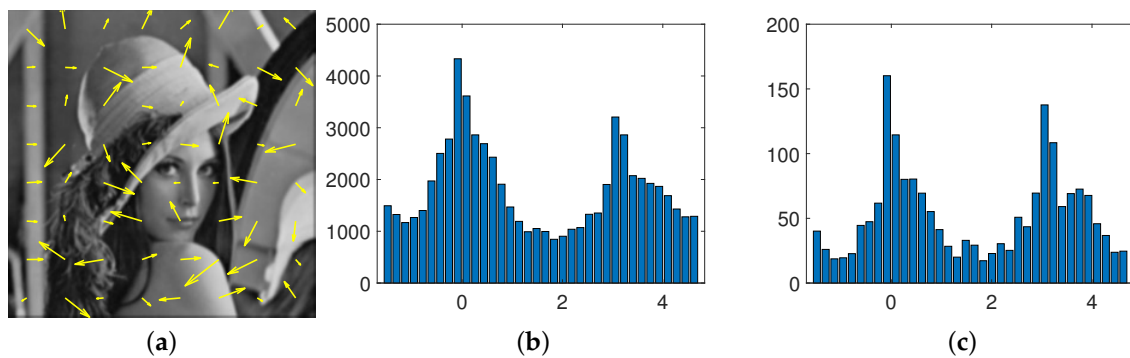


Figure 4. Gradient field statistics: (a) a gradient field on Lena image; (b) a traditional histogram; (c) the weighted histogram.

2. SIFT keypoints localization.

In this step, the keypoints are detected as the SIFT. The extreme points in the DoG space can be detected by non-maximum suppression. Let $D(x)$ and $U_r(x)$ denote the DoG function and r -neighborhood of x ; then, the keypoints can be located as Equations (5) and (6).

3. Main direction computing based on PCA of local patch.

The main direction will be computed on the local gradient field of patch $A(x)$ centered the keypoint x . To recognize heterogeneous image features, we normalize the keypoint a maximal point in the difference space. The local gradient field extracted from the pre-computed global gradient field will be treated in the same way. Then, PCA will be carried out on the normalized rearranged $2 \times n$ matrix L , and the first component can be used to determine the main direction:

$$\theta(A) = \arctan \frac{v_1[2]}{v_1[1]}. \tag{15}$$

Here, $v_1 = \text{PCA}(1, \nabla A)$ means the first component of PCA carried out on ∇A ; it can be denoted in more detail as

$$\lambda_1 = \max_{\lambda_i} \left\{ |\lambda_i|, LL^T v_i = \lambda_i v_i, i = 1, 2 \right\}, L = \frac{\nabla I_{loc} - \text{mean}(\nabla I_{loc})}{\text{std}(\nabla I_{loc})}.$$

Then, the patch will be rotated by the main angle before computing the feature vector.

4. Feature vector computing based on PCA and the adaptive intervals.

To enhance the ability recognizing heterogeneous image features, we define the feature vector by two parts. Gradient statistics of the rotated local patch B will be taken on the k selected intervals, and the first part of the feature vector is generated. Then, the local patch B will be divided into 4×4 sub-patches, denoted by $B_{s,t}$ ($s, t = 1, 2, 3, 4$). Different amounts of sub-patches can be set for different performances on accuracy or computation. A bigger amount means more computation and higher sensitivity. PCA will be carried out independently on each small patch and determine the main gradient. Figure 5 shows the 16 divided sub-patches and the main gradient computed by PCA on a local gradient field. These 16 main gradients generate the second part of the feature vector. The final $32 + k$ dimensional descriptor can be obtained by arranging them as

$$F = [F_1, F_2]_{k+32}, \quad (16)$$

where $F_1 = [\text{hist}(E_j, \nabla A)]_k$, $F_2 = [\text{PCA}(1, \nabla B_{s,t})]_{4 \times 4 \times 2}$, $\text{hist}(E_j, \nabla A)$ means the value in interval E_j when the histogram was implemented on local patch $A(x)$. E_j is one of the selected intervals computed in Step 1.

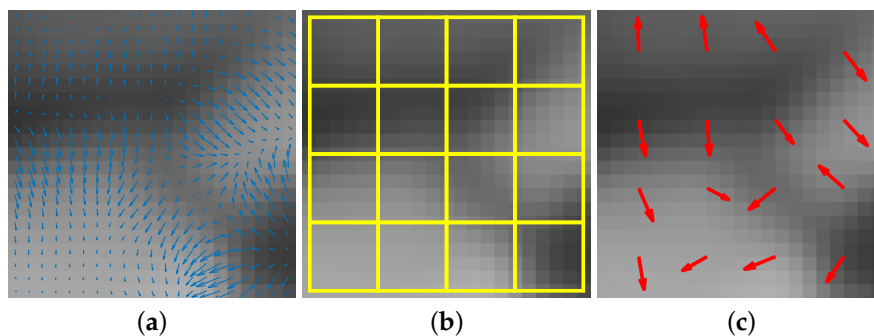


Figure 5. Divided sub-patches and the main gradients: (a) a local gradient field; (b) divided sub-patches; (c) main gradients computed by PCA.

3.2. Parameters Setting

The adaptive intervals are determined by the optimization model (14) based on m candidate intervals. Too big of an m is not necessary for the limited discrete information of the keypoints, and we set $m = 360$ in the experiments.

k adaptive intervals are used for gradients statistics and embedded in the feature vector. It can be found that a smaller k is advantageous to save computation but not beneficial to precision. A smaller k means less information from the keypoint collected in F_1 , the first part of the feature vector. It will reduce the accuracy in some sense. A bigger k required more computation, and it is beneficial to precision. More information for the keypoint can be collected in the F_1 . However, it also leads to higher sensitiveness. For a trade-off between computation and precision, it is advantageous to set $k = 36$ in our experiments.

Then, the matching between two keypoints x, y can be determined based on features $F(x)$ and $F(y)$ by Equation (11) and Equation (12). The threshold ρ_0 is set in $[1, 3]$. A smaller ρ_0 prefers to find more matches, but the precision may be low. A bigger ρ_0 is advantageous to capture accurate matches, but the amount may be more limited. In our experiments, $\rho_0 = 1.6$ is set for an expected precision.

4. Experiments

In real scenes, images are often affected by changes of viewpoint, scale, illumination, and so on. We ran two main types of experiments to explore the difference between the traditional methods and the proposed method. The first types of experiments explore the robustness to effects caused by rotation and the addition of noise. The second types of experiments verify the efficiency of the proposed method to heterogeneous image matching. We collected several classical natural images and heterogeneous image pairs. Some transforms are applied to them: (1) additional Gaussian noise (SNR = 20 db); (2) rotation of $7\pi/36$ followed by additional Gaussian noise; and (3) scaled 50% and rotated $\pi/6$.

The experiments are completed under Windows 7 with Matlab R2017b. To evaluate the performance, the proposed algorithm is compared with Harris, SIFT, and PCA-SIFT. For more credible and stable results, the first two experiments are repeated for 10 times, and the results are computed on average. In addition, according to the widely known work for features matching performance evaluation [5,46], correct number, cost time per match (ms), precision, and recall are applied to evaluate the performance. The related definitions are stated as follows. The correspondence between two images can be computed as illustrated in [47]:

$$\begin{aligned} \text{Precision} &= \frac{\text{Num}_{\text{correctmatches}}}{\text{Num}_{\text{correctmatches}} + \text{Num}_{\text{falsematches}}}, \\ \text{Recall} &= \frac{\text{Num}_{\text{correctmatches}}}{\text{Num}_{\text{correspondences}}}. \end{aligned} \quad (17)$$

Noised images matching test. The four images (Pavilion, Monkey, Fruits, Elaine) are all added Gaussian noise (SNR=20db) for their polluted versions. Then, the proposed method and the compared methods are all carried out on each pair (original and polluted).

Matching results of one implementation are shown in Figure 6. The rows from top to bottom mean different image pairs, Pavilion, Monkey, Fruits and Elaine, corresponding to the horizontal axis in Figure 7. The columns from left to right mean different methods: Harris, SIFT, PCA-SIFT, and the proposed method. Harris finds the least matching number while SIFT and PCA-SIFT achieve better performance. More matching can be found by the proposed method than others in this test. It can be found that 91 matches in the Pavilion image pair while 100 matches in the Monkey image pair by the proposed method. In addition, 52 matches and 98 matches can be found independently in the Fruits image pair and Elaine image pair. More details of an accurate matching number can be found in Figure 7a.

The cost time per match is shown in Figure 7b, and the proposed method achieves better performance than SIFT and PCA-SIFT. The bar value means the total cost time and the green part means the localization time. Precision and recall of each image pair are shown in Figure 7c,d. It can be found PCA-SIFT loses on precision but gains on recall compared to SIFT. The proposed method can achieve higher precision and recall than SIFT and PCA-SIFT, but roughly equivalent to Harris.

Rotated and noised images matching test. The four images (Pavilion, Monkey, Fruits, Elaine) are all added Gaussian noise (SNR = 20 db) after rotation by $\frac{\pi}{6}$. Then, the proposed method and the compared methods are all carried out on each pair. Matching results are shown in Figures 8 and 9. The columns mean different method results. Harris achieves accuracy of about 95% on the Pavilion image pair, and it finds only one match in the Fruits image pair. Our method has an advantage on the correct matching number. The details of an accurate matching number can be found in Figure 9a.

As shown in Figure 9a, Harris finds the least matching number in each image pair. The proposed method finds the most matches in three of four image pairs. It can be found that there are 42 matches in the Pavilion image pair while 60 matches in the Monkey image pair. In addition, 23 matches and 41 matches can be found independently in the Fruits image pair and Elaine image pair. More details of accurate matching numbers can be found in Figure 9.

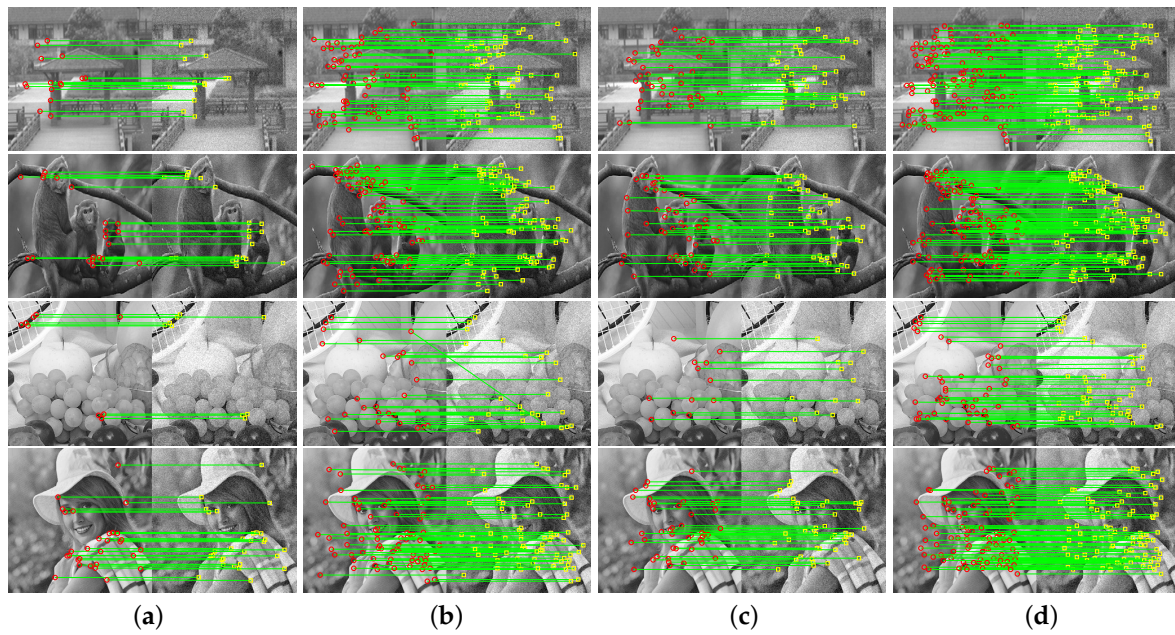


Figure 6. Matching results of noised images: (a) Harris results; (b) SIFT results; (c) PCA-SIFT results; (d) our results.

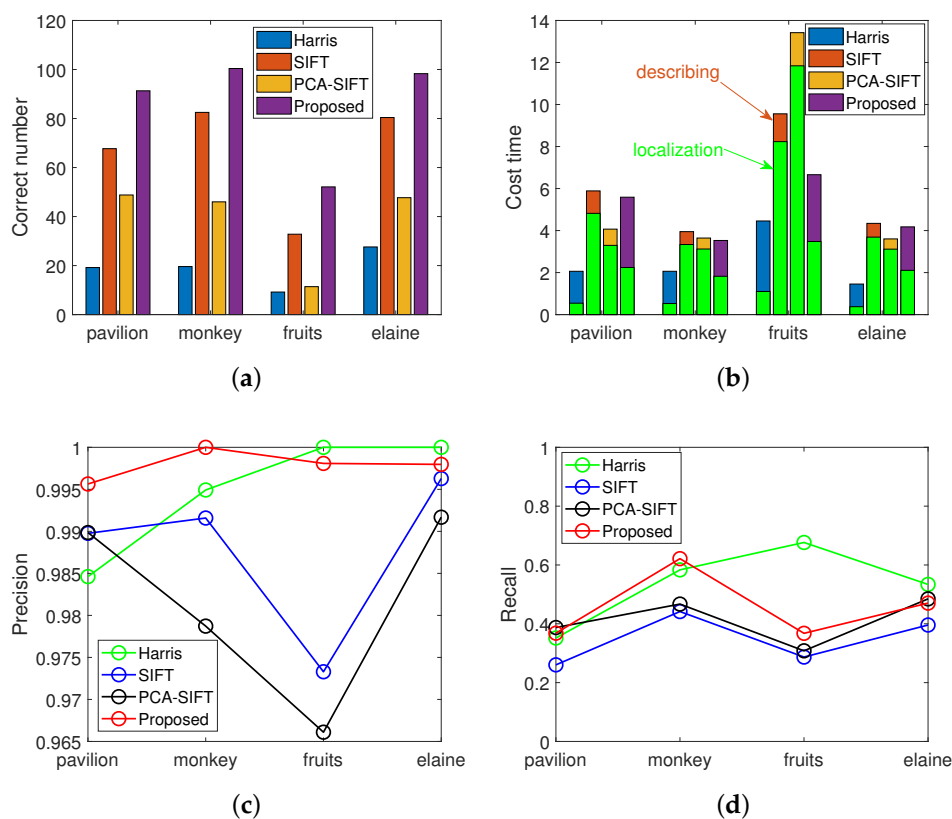


Figure 7. Results analysis of noised images matching test: (a) correct matching number; (b) cost time per match; (c) precision of the matches; (d) recall of the matches.

The cost time per match is shown in Figure 9b, and the proposed method achieves more stable performance. Precision and recall are shown in Figure 9c,d. It can be found that the proposed method achieves better precision and recall.

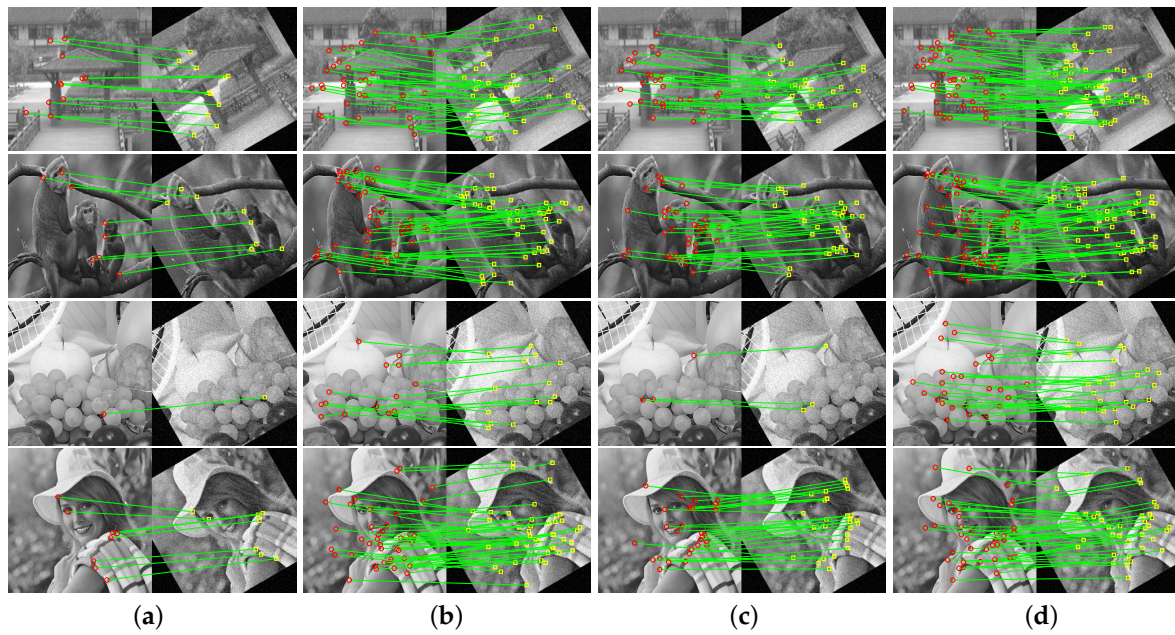


Figure 8. Matching results of rotated and noisy images: (a) Harris results; (b) SIFT results; (c) PCA-SIFT results; (d) our results.

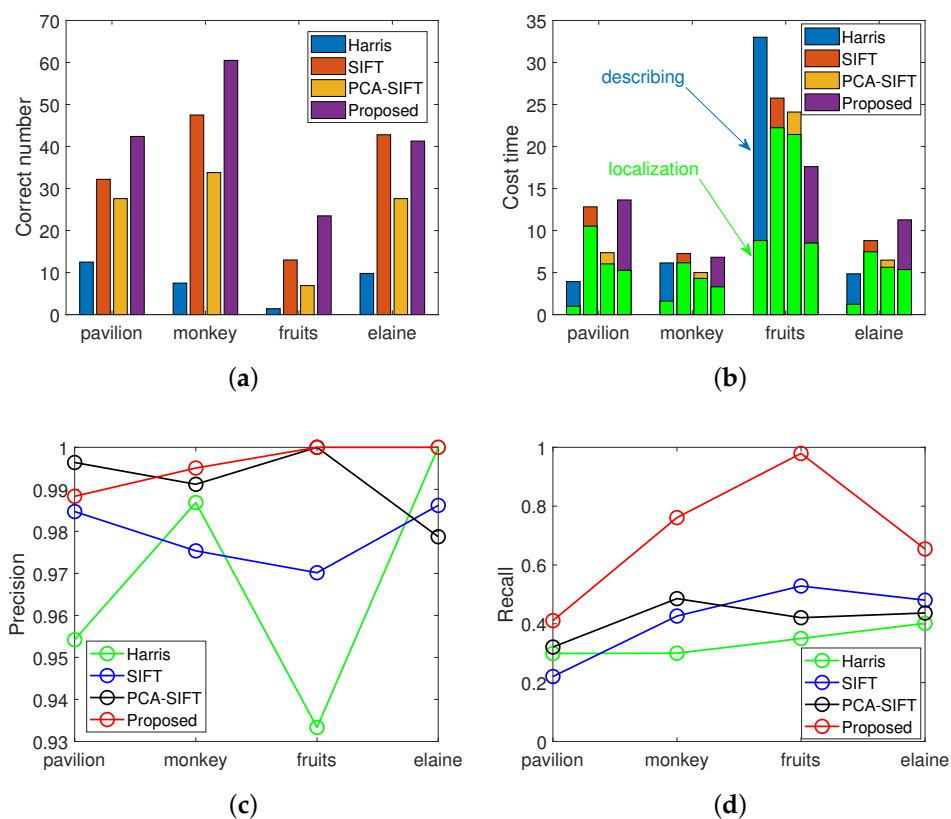


Figure 9. Results analysis of rotated and noisy images matching test: (a) correct matching number; (b) cost time per match; (c) precision of the matches; (d) recall of the matches.

Scaled and rotated images matching test. The four images (Pavilion, Monkey, Fruits, Elaine) are all scaled 50% after rotation by $\pi/6$. After applying the proposed method and the compared methods on each image pair, the matching results are shown in Figure 10. The proposed method finds more

matches than other methods, and Harris finds no match. PCA-SIFT finds less matches than SIFT, and it means that the components do not work so well with the mixed changes.

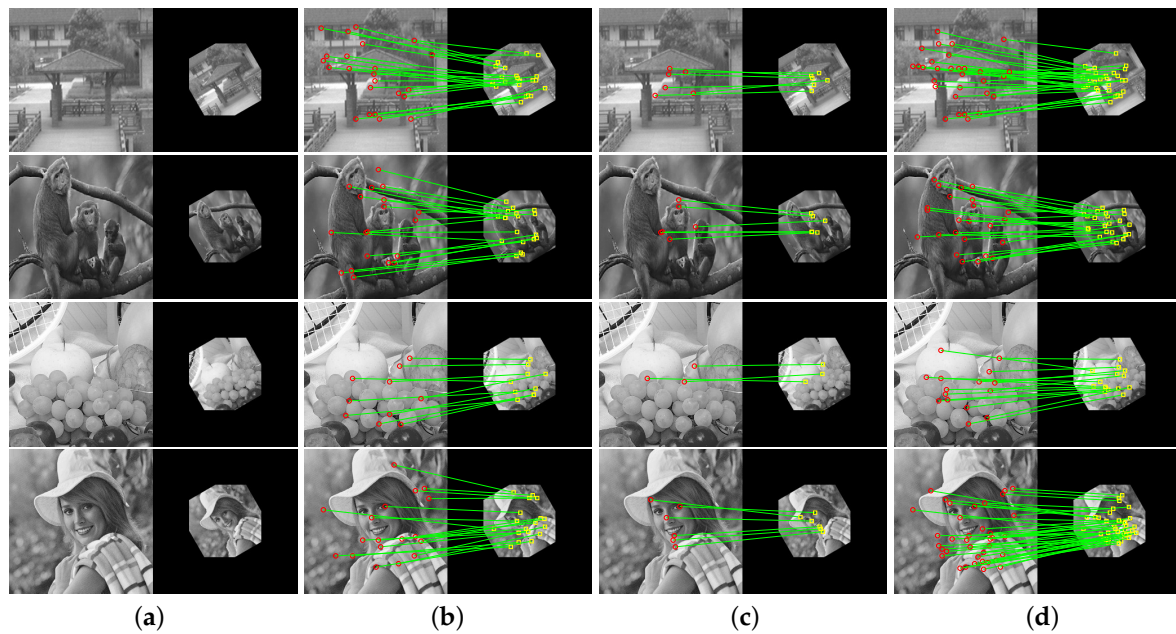


Figure 10. Matching results of scaled and rotated and images: (a) Harris results; (b) SIFT results; (c) PCA-SIFT results; (d) our results.

Furthermore, the proposed method finds 30 matches in the Pavilion image pair while 23 matches in the Monkey image pair by the proposed method. In addition, 15 matches and 37 matches can be found independently in the Fruits image pair and Elaine image pair. PCA-SIFT only catches about five matches on average, while Harris catches no matches. Other results can be found in Figure 11a.

As shown in Figure 11b, the proposed method achieves the best performance on cost time than SIFT and PCA-SIFT (Harris fails to find any match, and the cost time can't be computed). Precision and recall are shown in Figure 11c,d. PCA-SIFT achieves four 100% on Precision and the proposed method achieves three. PCA-SIFT achieves the best Recall, and our method performs better than SIFT and Harris.

Heterogeneous image matching test. The original four images are all brushed to simulate the heterogeneous version. After applying the proposed method and the compared methods on each image pair, the matching results are shown in Figure 12. Harris only can find very few matches in the second pair and the fourth pair but could find no matches in other pairs. SIFT and PCA-SIFT can find roughly the same matches in each pair. Our method can find more matches than the other three methods in each pair.

To be more exact, the proposed method finds 46 matches in the Pavilion image pair while 40 matches in the Monkey image pair by the proposed method. Thirty matches and 54 matches can be found independently in the Fruits image pair and Elaine image pair. SIFT and PCA-SIFT only catch about half of the proposed method. Harris fails to find any match in Pavilion and Fruits. Other results can be found in Figure 13a.

As shown in Figure 13b, the proposed method achieves the best performance on cost time compared to SIFT and PCA-SIFT (Harris fails to find any matches in two pairs, and the cost time can't be computed). Precision and recall are shown in Figure 13c,d. PCA-SIFT achieves a little worse performance than SIFT on Precision. SIFT achieves two 100%, and our method achieves one, but our method achieves obviously better performance in Monkey and Fruits. The proposed method achieves the best performance on recall in three pairs and only a little lower than PCA-SIFT in the Elaine image pair.

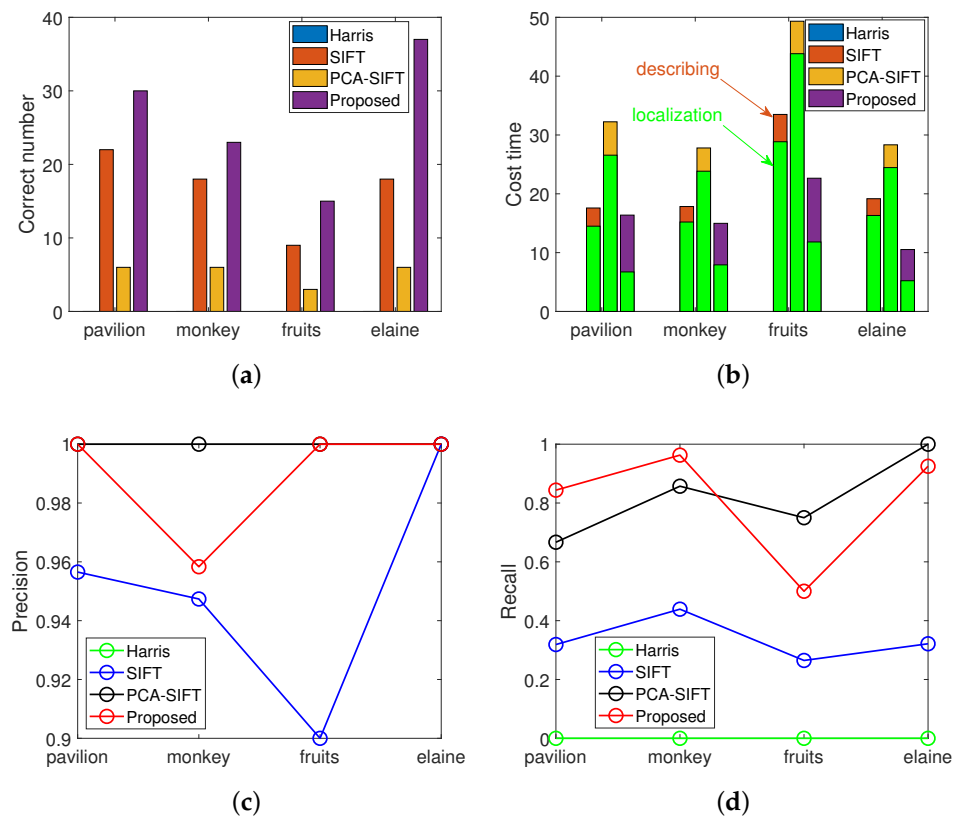


Figure 11. Results analysis of the scaled and rotated images matching test: (a) correct matching number; (b) cost time per match; (c) precision of the matches; (d) recall of the matches.

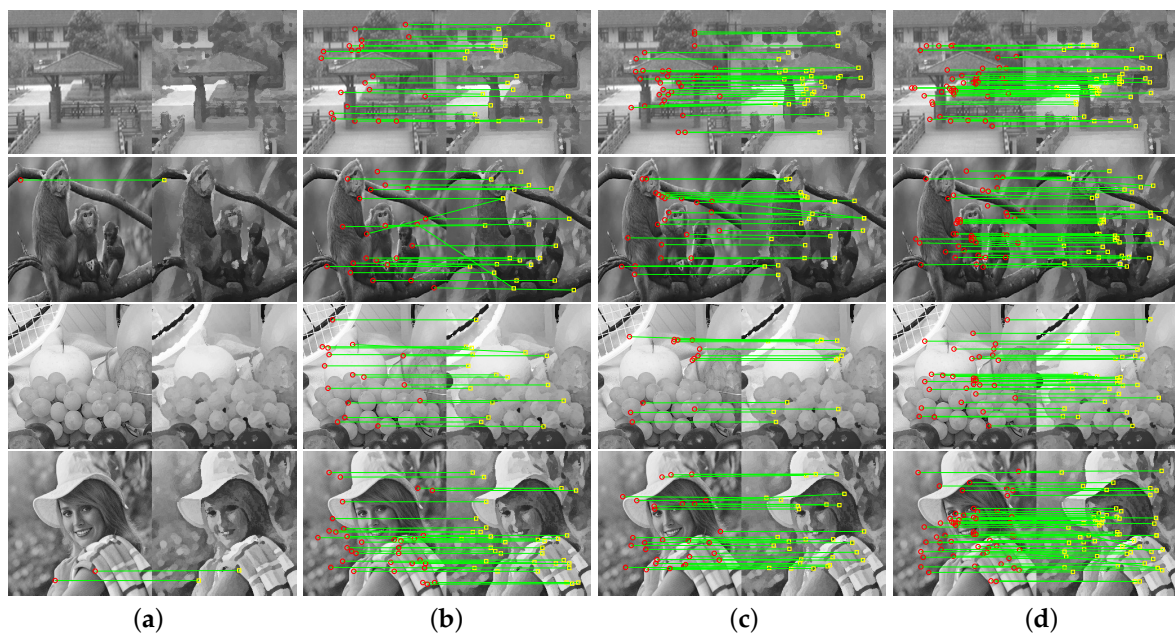


Figure 12. Matching results of heterogeneous images: (a) Harris results; (b) SIFT results; (c) PCA-SIFT results; (d) our results.

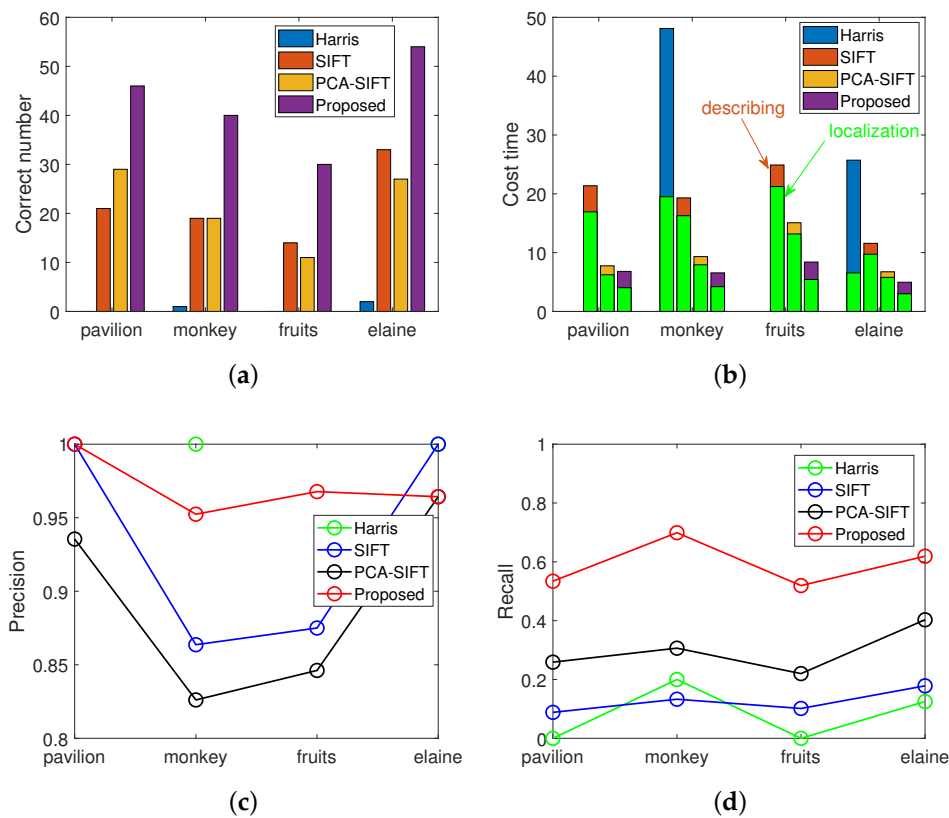


Figure 13. Results analysis of the heterogeneous and noised images matching test: (a) correct matching number; (b) cost time per match; (c) precision of the matches; (d) recall of the matches.

Based on the four experiments above, we use three performance levels (Low, Moderate, High) to comprehensively evaluate the methods as shown in Table 1. It can be found the proposed method is advantageous in mixed changed image matching and heterogeneous image matching. For more comprehensive results from the proposed method, we carry out these methods on the Oxford database in the last experiment.

Table 1. Performance levels of correct number.

Changes	Noise	Rotation + Noise	Rotation + Scale	Heterogeneous
Harris	Low	Low	Low	Low
SIFT	High	Moderate	Moderate	Moderate
PCA-SIFT	Moderate	High	Moderate	Moderate
Proposed	High	High	High	High

Matching test on Oxford database. Eight image pairs are selected from the well-known feature matching test database Oxford provided by Mikolajczyk et al. [11]. All the images are scaled to one-third of the original size, and the right one of each pair is transformed into a heterogeneous version as shown in Figure 14. After applying the proposed method and the compared methods on each image pair, the matching results can be obtained as shown in Figure 14. Harris finds few matches and the proposed method finds the most matches in each pair. SIFT and PCA-SIFT can find moderate matches. Our method can find more matches than the other three methods in each pair as shown in Figure 15a and Table 2.

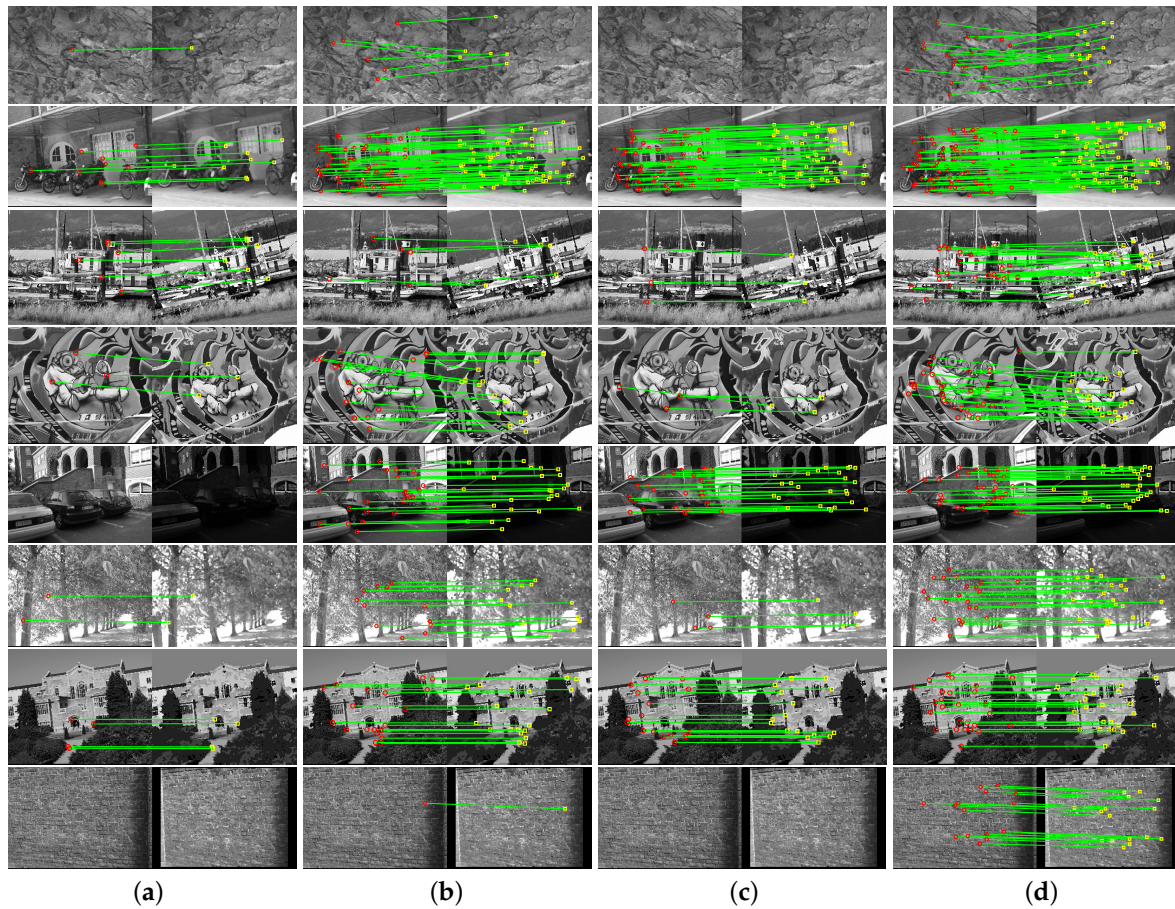


Figure 14. Matching results of Oxford database: (a) Harris results; (b) SIFT results; (c) PCA-SIFT results; (d) our results.

Table 2. Correct numbers in the matching test on the Oxford database.

Image	Bark	Bikes	Boat	Graf	Leuven	Trees	Ubc	Wall
Harris	1	9	6	3	0	2	4	0
SIFT	6	74	5	15	25	12	19	1
PCA-SIFT	0	73	3	3	26	4	17	0
Proposed	19	98	25	28	35	24	26	16

As shown in Figure 15b, the proposed method achieves better performance on cost time than Harris and SIFT but worse than PCA-SIFT. Though Harris finds the least matches, it still achieves the best performance on Precision excluding Leuven and Wall. The proposed method achieves the best performance as shown in Figure 15c. PCA-SIFT achieves four 100%, and SIFT achieves seven 100%. The recall performances are shown in Figure 15d, and it can be found that the proposed method achieves better performance than the other compared methods.

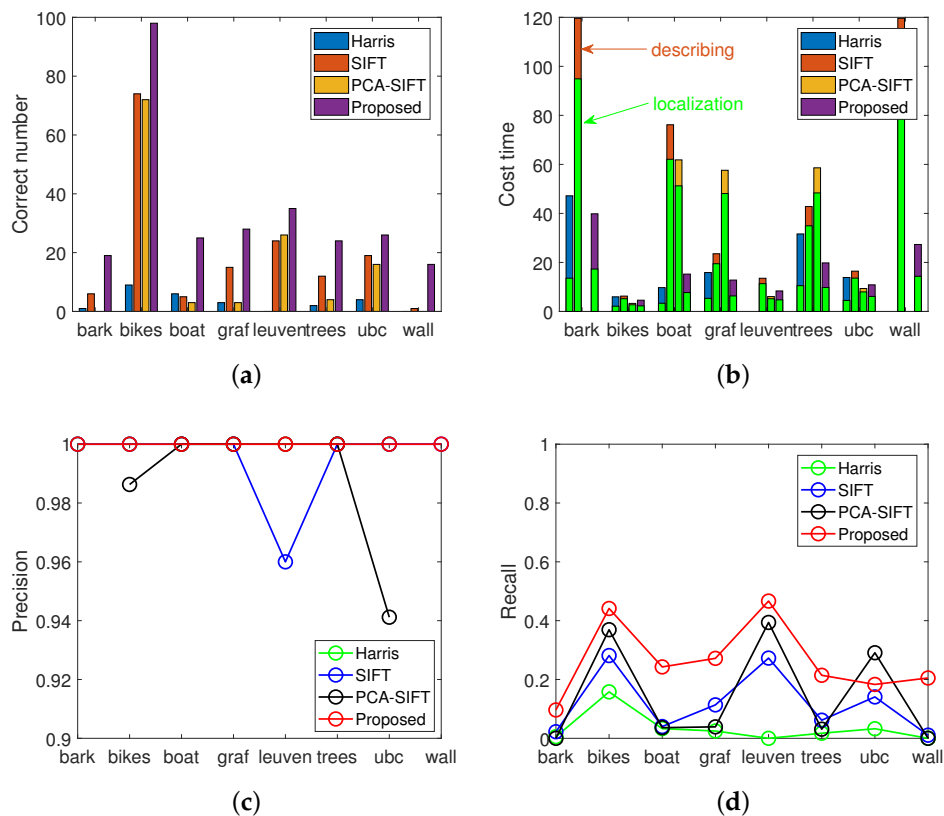


Figure 15. Results analysis of matching test on Oxford database: (a) correct matching number; (b) cost time per match; (c) precision of the matches; (d) recall of the matches.

5. Conclusions

In this paper, a novel features describing method based on scale space and local principle component analysis is presented for heterogeneous image matching. The keypoints are located in the DoG space and similar to traditional SIFT. Then, k directions are optimized for gradient statistics on some prior local patches centered keypoints and the first part of the feature describing vector is formed. Then, PCA is introduced to compute the main direction of each local patch instead of using the traditional eight-direction statistics. In addition, the second part of the feature describing vector is then achieved by applying PCA on each sub-patch of the rotated local patch. Several experiments' results verified the efficiency and accuracy of the proposed method.

Author Contributions: Conceptualization, B.Z. and X.D.; resources, W.W.; data curation, D.Y.; writing—original draft preparation, B.Z. and X.D.; writing—review and editing, M.W. and R.D.; visualization, D.Y.; supervision, B.Z.

Funding: This work is supported in part by the National Nature Science Foundation of China (11301414).

Acknowledgments: The authors would like to thank the anonymous reviewers for their insightful comments that helped to improve the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Harris, C.; Stephens, M. A combined corner and edge detector. *Alvey Vis. Conf.* **1988**, *15*, 10–5244.
- Förstner, W.; Gülch, E. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, 2–4 June 1987; pp. 281–305.

3. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
4. Smith, S.M.; Brady, J.M. SUSAN—A new approach to low level image processing. *Int. J. Comput. Vis.* **1997**, *23*, 45–78. [[CrossRef](#)]
5. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [[CrossRef](#)] [[PubMed](#)]
6. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
7. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2, pp. 506–513.
8. Guo, Y.; Bennamoun, M.; Soheli, F.; Lu, M.; Wan, J.; Kwok, N.M. A comprehensive performance evaluation of 3D local feature descriptors. *Int. J. Comput. Vis.* **2016**, *116*, 66–89. [[CrossRef](#)]
9. Li, Y.; Li, Q.; Liu, Y.; Xie, W. A spatial-spectral SIFT for hyperspectral image matching and classification. *Pattern Recognit. Lett.* **2018**, *127*, 18–26. [[CrossRef](#)]
10. Matas, J.; Chum, O.; Urban, M.; Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput.* **2004**, *22*, 761–767. [[CrossRef](#)]
11. Mikolajczyk, K.; Tuytelaars, T.; Schmid, C.; Zisserman, A.; Matas, J.; Schaffalitzky, F.; Kadir, T.; Van Gool, L. A comparison of affine region detectors. *Int. J. Comput. Vis.* **2005**, *65*, 43–72. [[CrossRef](#)]
12. Rodríguez, M.; Delon, J.; Morel, J.M. Fast Affine Invariant Image Matching. *Image Process. Online* **2018**, *8*, 251–281, doi:10.5201/ipol.2018.225. [[CrossRef](#)]
13. Moravec, H.P. Towards automatic visual obstacle avoidance. In Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, CA, USA, 22–25 August 1977.
14. Zitova, B.; Flusser, J. Image registration methods: a survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [[CrossRef](#)]
15. Liu, C.; Yuen, J.; Torralba, A. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 978–994. [[CrossRef](#)] [[PubMed](#)]
16. Tau, M.; Hassner, T. Dense correspondences across scenes and scales. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 875–888. [[CrossRef](#)] [[PubMed](#)]
17. Todorovic, S.; Ahuja, N. Region-based hierarchical image matching. *Int. J. Comput. Vis.* **2008**, *78*, 47–66. [[CrossRef](#)]
18. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision, Graz, Austria, 7–13 May 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
19. Shi, J.; Tomasi. Good features to track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
20. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In *European Conference on Computer Vision, Graz, Austria, 7–13 May 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 430–443.
21. Mair, E.; Hager, G.D.; Burschka, D.; Suppa, M.; Hirzinger, G. Adaptive and generic corner detection based on the accelerated segment test. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 183–196.
22. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 214–227.
23. Olshausen, B.A.; Field, D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **1996**, *381*, 607. [[CrossRef](#)] [[PubMed](#)]
24. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1794–1801.
25. Zhang, C.; Liu, J.; Liang, C.; Xue, Z.; Pang, J.; Huang, Q. Image classification by non-negative sparse coding, correlation constrained low-rank and sparse decomposition. *Comput. Vis. Image Underst.* **2014**, *123*, 14–22. [[CrossRef](#)]
26. Liu, Y.; Canu, S.; Honeine, P.; Ruan, S. Mixed Integer Programming for Sparse Coding: Application to Image Denoising. *IEEE Trans. Comput. Imaging* **2019**, *5*, 354–365. [[CrossRef](#)]

27. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote. Sens.* **2016**, *54*, 6232–6251. [[CrossRef](#)]
28. Yang, A.; Yang, X.; Wu, W.; Liu, H.; Zhuansun, Y. Research on feature extraction of tumor image based on convolutional neural network. *IEEE Access* **2019**, *7*, 24204–24213. [[CrossRef](#)]
29. Cho, J.H.; Park, C.G. Multiple Feature Aggregation Using Convolutional Neural Networks for SAR Image-Based Automatic Target Recognition. *IEEE Geosci. Remote. Sens. Lett.* **2018**, *15*, 1882–1886. [[CrossRef](#)]
30. Hoogi, A.; Subramaniam, A.; Veerapaneni, R.; Rubin, D.L. Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis. *IEEE Trans. Med Imaging* **2016**, *36*, 781–791. [[CrossRef](#)] [[PubMed](#)]
31. Leutenegger, S.; Chli, M.; Siegwart, R. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
32. Dou, J.; Qin, Q.; Tu, Z. Robust image matching based on the information of SIFT. *Optik* **2018**, *171*, 850–861. [[CrossRef](#)]
33. He, Y.; Deng, G.; Wang, Y.; Wei, L.; Yang, J.; Li, X.; Zhang, Y. Optimization of SIFT algorithm for fast-image feature extraction in line-scanning ophthalmoscope. *Optik* **2018**, *152*, 21–28. [[CrossRef](#)]
34. Chen, Z.; Sun, S.K. A Zernike moment phase-based descriptor for local image representation and matching. *IEEE Trans. Image Process.* **2009**, *19*, 205–219. [[CrossRef](#)] [[PubMed](#)]
35. Klare, B.; Jain, A.K. Heterogeneous Face Recognition: Matching NIR to Visible Light Images. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010.
36. Wang, T.; Kemaq, Q.; Seah, H.S.; Lin, F. A flexible heterogeneous real-time digital image correlation system. *Opt. Lasers Eng.* **2018**, *110*, 7–17. [[CrossRef](#)]
37. Lin, G.; Fan, C.; Zhu, H.; Miu, Y.; Kang, X. Visual feature coding based on heterogeneous structure fusion for image classification. *Inf. Fusion* **2017**, *36*, 275–283. [[CrossRef](#)]
38. Wang, N.; Li, J.; Tao, D.; Li, X.; Gao, X. Heterogeneous image transformation. *Pattern Recognit. Lett.* **2013**, *34*, 77–84. [[CrossRef](#)]
39. Peng, C.; Gao, X.; Wang, N.; Li, J. Graphical representation for heterogeneous face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 301–312. [[CrossRef](#)] [[PubMed](#)]
40. Witkin, A.P. Scale-space filtering. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Netherlands, 1987; pp. 329–332.
41. Wei, W.; Yang, X.L.; Zhou, B.; Feng, J.; Shen, P.Y. Combined energy minimization for image reconstruction from few views. *Math. Probl. Eng.* **2012**, *2012*, 154630. [[CrossRef](#)]
42. Wei, W.; Zhou, B.; Połap, D.; Woźniak, M. A regional adaptive variational PDE model for computed tomography image reconstruction. *Pattern Recognit.* **2019**, *92*, 64–81. [[CrossRef](#)]
43. Zhou, B.; Mu, C.L. Level set evolution for boundary extraction based on a p-Laplace equation. *Appl. Math. Model.* **2010**, *34*, 3910–3916. [[CrossRef](#)]
44. Jolliffe, I. *Principal Component Analysis*; Springer: Berlin/Heidelberg, Germany, 2011.
45. Fergus, R.; Perona, P.; Zisserman, A. Object class recognition by unsupervised scale-invariant learning. In *CVPR (2)*, Madison, WI, USA, 16–22 June 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 264–271.
46. Heinly, J.; Dunn, E.; Frahm, J.M. Comparative evaluation of binary features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 759–773.
47. Mikolajczyk, K.; Schmid, C. Scale & affine invariant interest point detectors. *Int. J. Comput. Vis.* **2004**, *60*, 63–86.

