


Region duplication detection based on image segmentation and keypoint contexts

Cong Lin^{1,2}  · Wei Lu^{1,3} · Wei Sun^{4,5} · Jinhua Zeng⁶ ·
Tianhua Xu⁵ · Jian-Huang Lai¹

Received: 15 December 2016 / Revised: 12 May 2017 / Accepted: 10 July 2017 /
Published online: 22 July 2017
© Springer Science+Business Media, LLC 2017

Abstract In this paper, a novel region duplication detection method is proposed based on image segmentation and keypoint contexts. The proposed method includes the primary

✉ Wei Lu
luwei3@mail.sysu.edu.cn

Cong Lin
lincong0310@gmail.com

Wei Sun
sunwei@mail.sysu.edu.cn

Jinhua Zeng
zengjh@ssfjd.cn

Tianhua Xu
thxu@bjtu.edu.cn

Jian-Huang Lai
stsljh@mail.sysu.edu.cn

¹ School of Data and Computer Science, Guangdong Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou 510006, China

² Educational Technology Center, Guangdong University of Finance and Economics, Guangzhou 510320, China

³ State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

⁴ School of Electronics and Information Technology, Key Laboratory of Information Technology (Ministry of Education), Sun Yat-sen University, Guangzhou 510006, China

⁵ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

⁶ Institute of Forensic Science, Ministry of Justice, Shanghai 200063, China

region duplication detection based on keypoints and the supplementary region duplication detection based on blocks. In the primary region duplication detection, an image is divided into non-overlapped patches by using SLIC. Furthermore, the keypoints are matched and clustered within the same patch as patch feature. Then the patches are matched and an affine transformation is tried to be estimated from a pair of patches. When the estimation fails, in the supplementary region duplication detection, a transformation matrix is tried to be estimated from a pair of keypoints by the proposed Keypoint Contexts (KC) approach. The experimental results indicate that the proposed method can achieve much better comprehensive performances than the state-of-the-art methods on the public databases, even under various challenging conditions.

Keywords Digital image forensics · Region duplication detection · Copy-move forgery · Image segmentation · Keypoint contexts

1 Introduction

As the development of photo editing software and Internet, digital image forensics aiming to reveal forgery operations in digital images is receiving more and more attentions [6]. Among the existing types of image tampering, a common manipulation is region duplication (copy-move forgery) [8], which is to paste one or more regions of an image into other parts of the same image. Many region duplication detection methods have been proposed. In fact, region duplication detection methods can be classified into two categories. The first category is the block-based region duplication detection methods. Fridrich et al. [8] proposed the first method, Discrete Cosine Transform (DCT) and lexicographic sorting are adopted in the method to detect the duplicated blocks. Huang et al. [12] proposed the method enhancing the work done by Fridrich et al. [8]. The experiments show that the method is not only fast, but also robust to JPEG compression, blurring and adding noise. Khan and Kulkarni [13] proposed a method based on Discrete Wavelet Transform (DWT). The method showed robustness to adding noise and JPEG compression but not to scaling and rotation. Popescu and Farid [17] proposed a method using Principal Component Analysis (PCA) to reduce the DCT representation of each image block. The method is robust against JPEG compression and adding noise. Ryu et al. [19, 20] proposed a method based on the Zernike Moment (ZM). The method starts by extracting ZMs of small image blocks. Then a novel algorithm based on Locality Sensitive Hashing (LSH) is proposed in the matching phase. The Random Sample Consensus (RANSAC) [7] algorithm is used to reduce mismatch. The experimental results showed that the method is robust to rotation, adding noise and JPEG compression but not to scaling. The block-based method is good at plain copy-move, but it cannot address significant geometrical transformations of the duplicated regions, and its computing time is expensive.

To overcome disadvantages of the block-based method, the keypoint-based method is proposed. The method usually extracts keypoints from the whole image, such as the Scale-Invariant Feature Transform (SIFT) [15] and the Speeded Up Robust Features (SURF) [4]. Huang et al. [11] proposed a preliminary method to utilize the SIFT in region duplication detection. Pan and Lyu [16] also proposed a method based on SIFT. As an important work, an approach to estimate the affine transformation is proposed in the paper. The method of Pan and Lyu [16] is robust when JPEG compression or adding noise is adopted as post-processing operations in their experiments, but it is not robust when regions have fewer

keypoints or images have intrinsically identical areas. Amerini et al. [2] also proposed a method based on SIFT. Before the estimation of the affine transformation, an Agglomerative Hierarchical Clustering (AHC) [10] is performed on the matched points. The experimental results showed very good performances in terms of a high True Positive Ratio (TPR) and a low False Positive Ratio (FPR). The method of Amerini et al. [2] gives a framework for methods based on keypoints. Later, Amerini et al. [3] also proposed a region duplication detection method similar to [2]. The main difference is the clustering method. This time, the J-Linkage clustering is used. Li et al. [14] proposed a method using image segmentation and keypoint matching. The matching approach in [14] consists of two stages. In the first stage, a transformation matrix is estimated roughly. In the second stage, the locations of keypoints are moved, and the transformation matrix is estimated iteratively by an EM-based algorithm. The image segmentation strategy is also adopted by Pun et al. [18].

The keypoint-based method is fast and robust against JPEG compression and adding noise, but it is not robust as smaller regions have fewer keypoints. The block-based method can obtain good performances in plain copy-move, but it is usually slow and cannot deal with significant geometrical transformations. Some methods can detect more duplicated regions than others, so their recall is higher than others. Meanwhile, some background is regarded as duplicated regions, which lead to the decrease of the precision. In fact, greater recall might decrease precision and vice versa. There must be a trade-off between the recall and the precision.

To address the above-mentioned issues, this paper is to propose a novel method for region duplication detection based on image segmentation and keypoint contexts. The proposed method integrates both keypoint-based and block-based region duplication detection schemes. Considering the trade-off between the recall and the precision, the aim of the paper is to find a method that has better F_1 scores. The proposed method is divided into two phases: the primary region duplication detection based on keypoints and the supplementary region duplication detection based on blocks. The final result is to combine the detection results of the two phases. In the primary region duplication detection, the input image is segmented into non-overlapped patches, each patch is a meaningful superpixel, then SIFT keypoints are extracted from the whole image, matching patches are processed after keypoints are matched. A transformation matrix is tried to be estimated from a pair of patches. When the estimation of affine transformation fails, in the supplementary region duplication detection, a Keypoint Contexts (KC) approach is proposed, which is similar to block-based approach, a transformation matrix is tried to be estimated from a pair of keypoints.

The rest of the paper is organized as follows. In Section 2, the primary region duplication detection is described. In Section 3, the supplementary region duplication detection is described. The experimental results and discussions are given in Section 4. Finally, the conclusion is drawn in Section 5.

2 Primary region duplication detection

The proposed method is divided into two phases. The duplicated regions obtained in Section 2 are denoted as D_1 , then the duplicated regions obtained in Section 3 are denoted as D_2 . All the duplicated regions obtained by the proposed method are denoted as D (Where $D = D_1 \cup D_2$). The framework of the proposed method is shown in Fig. 1. Main steps of the proposed method are shown in Fig. 2. In the section, the primary region duplication detection is described in detail.

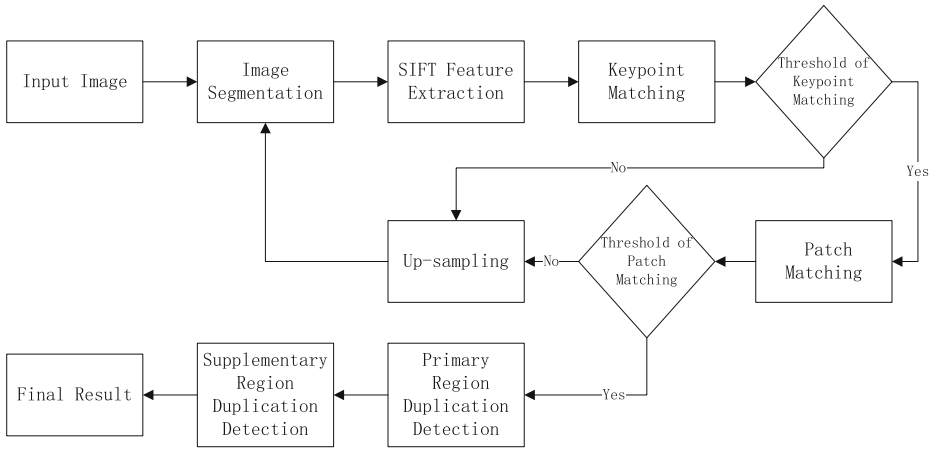


Fig. 1 The flowchart of the proposed method

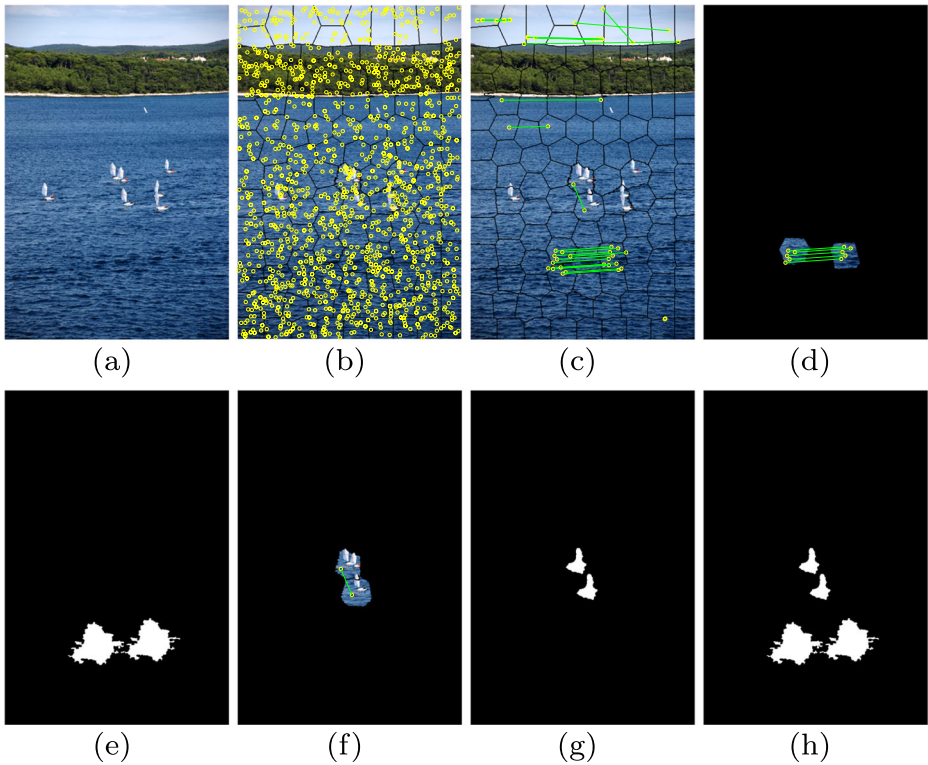


Fig. 2 Main steps of the proposed method: **a** the input image, **b** keypoints and segmentation of the input image, **c** all the matched keypoints and image segmentation, **d** the matched patches, **e** the duplicated regions of the matched patches, **f** a pair of keypoints and the corresponding image patches, **g** the duplicated regions of the pair of keypoints, **h** the final duplicated regions D

2.1 Image segmentation

As mentioned, a lot of block-based region duplication detection method have been proposed [8, 12, 13, 17, 19, 20] in recent years. The input image is first divided into overlapped regular blocks in this category. There are some disadvantages of the block-based methods. The first is that the block-based method can not address significant geometrical transformations of the duplicated regions. The second is that the block size is fixed to a constant value, but the duplicated regions usually have a certain meaning with irregular shapes. To overcome these shortcomings of the block-based methods, an image is divided into non-overlapped superpixels as image patches in the proposed method. There are a large number of image segmentation algorithms. In our implementation, the Simple Linear Iterative Clustering (SLIC) algorithm [1] is adopted to segment the input image into meaningful patches. In fact, each patch is a superpixel. The SLIC algorithm adopts a k-means clustering approach to efficiently generate patches. In spite of simplicity, the SLIC algorithm adheres to boundaries and can improve segmentation performance. Meanwhile, the SLIC algorithm is fast and efficient.

The initial size of the patches in SLIC is import to the duplicate region detection. In fact, the input images and the duplicated regions are of different sizes and have different contents. Different initial sizes of the patches can produce different detection results in the proposed method. When the size of duplicated region is large and the size of the patch is small, the computational cost will be very expensive. Otherwise, when the size of duplicated region is small and the size of the patch is large, the source region and the target region may be existed in the same patch. In our implementation, every image is empirically segmented into 100 patches.

After the image is segmented into superpixels as patches, the feature of the patch is extracted. Unlike the traditional block-based methods, the patches are irregular shapes. However, the keypoint-based method is robust to scaling, rotation, adding noise and JPEG compression. So keypoints are extracted from the image, each patch have some keypoints, the patch feature are described by the features of the keypoints which belong to the patch.

2.2 Keypoints extraction and matching

After the image is segmented, keypoints and their corresponding descriptors are extracted from the image, then the keypoints are matched. There are plenty of methods to detect and describe local features in computer vision. Among them, the Scale-Invariant Feature Transform (SIFT) [15] and the Speeded Up Robust Features (SURF) [4] have been widely used. According to Li et al. [14], SIFT is more robust than SURF, especially when the image is resized. In our implementation, SIFT keypoints are extracted from the whole image, each keypoint only belongs to one patch. Therefore, each patches are characterized by the SIFT keypoints which are in the corresponding patches. The image segmentation as well as keypoints in the image are shown in Fig. 2b.

Given a keypoint, we define a similarity vector $D = \{d_1, d_2, \dots, d_n\}$ that represents the sorted Euclidean distances with respect to the other descriptors of the whole image, where d_1 represents the Euclidean distances of the features between the given keypoint and its closest neighbor, and d_2 represents the Euclidean distances of the features between the given keypoint and its second-closest neighbor, etc. Thus, the given keypoint and its closest neighbor are matched if the following constraint is satisfied:

$$d_1/d_2 \leq T_d. \quad (1)$$

In other words, the two keypoints are matched if the ratio between the distance of the closest neighbor and that of the second-closest one is less than or equal to a threshold T_d . The procedure is known as 2NN test. In our implementation, we set $T_d = 0.5$. Therefore, given a keypoint, we can only find no more than one keypoint which is matched with it. There is an obvious disadvantage in the matching approach if the same region is copied and pasted more than once. To overcome the above shortcomings, the generalized 2 nearest neighbor (g2NN) [2] which is a generalization of (1) is adopted. The matching approach is similar, but the new constraint is $d_i/d_{i+1} \leq T_d$ (Where $1 \leq i < n$). If the value is j when the constraint is satisfied, a keypoint set K_s which is in correspondence to a distance in $\{d_1, \dots, d_j\}$ (where $1 \leq j < n$) is obtained. Any keypoint in the set K_s is matched with the given keypoint. In addition, given a keypoint, we search for n nearest neighbors from the whole image ($n = 10$). Now the set of matched keypoints are acquired. Please note that matched keypoints cannot come from the same patch, for those cannot help us to obtain matched patches. All the matched keypoints and image segmentation are shown in Fig. 2c. The image is up-sampled if the number of matched keypoints is less than a threshold (we set it equal to 15), then we restart our method from the beginning. When an image is up-sampled, more keypoints are acquired, some undetectable duplicated regions can be detected now. Meanwhile, some detectable duplicated regions will be more accurately located.

2.3 Patch feature matching

Next, we find the feature of patch, then patch feature are matched. Each keypoint only belongs to a certain patch, and the patch feature consists of the features of keypoints. For instance, there are six keypoints in patch E, each keypoints have a descriptor, then the feature of patch E consists of the six descriptors. After the set of matched keypoints is obtained, we find the suspicious pairs of patches, each patch is compared with the rest. Patch A and B are considered to be a pair of patches if the number of matched keypoints between patch A and B is greater than a threshold T_p . In our implementation, we set $T_p = 3$. The image is up-sampled if we cannot obtain any matched patches, then the proposed method is restarted from the beginning. Note that the up-sample will be performed no more than once when an image is processed.

2.4 Estimation of affine transformation

Once matched patches are found, then the affine transformation between the source region and the target region is estimated. Given two corresponding keypoints $\hat{x}_i = (x_i, y_i, 1)^T$ and $\hat{x}'_i = (x'_i, y'_i, 1)^T$ from the source region and the target region, respectively, their affine transformation is represented by a 3×3 matrix H :

$$\hat{x}'_i = H\hat{x}_i = \begin{pmatrix} h_{11} & h_{12} & t_x \\ h_{21} & h_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \hat{x}_i \quad (2)$$

Where t_x and t_y are the translation factors, while h_{11} , h_{12} , h_{21} and h_{22} are the scaling and rotation directions deformation. There are six transform parameters in the transformation matrix H , so three or more pairs of keypoints are selected randomly that are not collinear to estimate the transformation matrix. According to [9], the transformation matrix H is the one for which the following error is minimized:

$$\sum_i d(\hat{x}'_i, H\hat{x}_i)^2 \quad . \quad (3)$$

In fact, Random Sample Consensus (RANSAC) [7] algorithm which can lead to the largest number of matched keypoints and the minimum error in (3) is employed. Therefore, we get an affine transformation between the source region and the target region. A transformation matrix is estimated from a pair of patches in the proposed method. Next, we use the approach which is similar to [16] to locate the duplicated regions. In order to reduce the errors, some transformation matrices with mistakes are abandoned before the duplicated regions are located. The matched patches and the corresponding duplicated regions are shown in Fig. 2d and e. All the duplicated regions obtained in Section 2 are combined together and denoted as D_1 .

3 Supplementary region duplication detection

In the primary region duplication detection, we have obtained the matched patches, then a transformation matrix is estimated from a pair of patches. The drawbacks of the traditional keypoint-based methods are discussed as follows. For one thing, transformation matrix cannot be estimated when there are only one or two pairs of keypoints between the two patches; for another, there are more than three pairs of keypoints between the two patches, but affine transformation can not be acquired from those matched patches. The above mentioned conditions can not be solved by the traditional keypoint-based methods. Some duplicated regions with few keypoints can not be detected. Since keypoints are few, we consider the points around the keypoints. Li et al. [14] propose an approach to exploit all the points in the matched patches to find out a more accurate estimation of transformation matrix. The dense SIFT descriptor [22] is adopted to describe the points belonging to the duplicated regions. The method of Li et al. [14] is used for moving the locations of the keypoints, then a transformation matrix is re-estimated. The problem remains the same because the number of keypoints is not increased. In most cases, we believe that there is no forgery when the pairs of keypoints are less than three between the two patches. Thus, some duplicated regions with less than three pairs of keypoints cannot be detected by the traditional keypoint-based method. In the section, we consider the points around the keypoint and call them Keypoint Contexts (KC). A transformation matrix is tried to be estimated from a pair of keypoints by the KC approach.

The proposed KC approach is described in detail at present. Since the keypoints are few between the two patches, we consider the points around the keypoints. The Keypoint Contexts (KC) approach is shown in Fig. 3. Since \hat{x}_1 and \hat{x}'_1 are the matched SIFT keypoints, the point \mathbf{x}_i is around the keypoint \hat{x}_1 , then the point which is matched with the point \mathbf{x}_i should also be around the keypoint \hat{x}'_1 . We select one region (i.e., Ω) whose center is \hat{x}_1 and whose size is $(2r + 1) \times (2r + 1)$, the other region (i.e., Ω') whose center is \hat{x}'_1 and whose size is $(2r + 1) \times (2r + 1)$. We set $T_3 = (2r + 1) \times (2r + 1)$. In our implementation, we set $r = 5$, then we have $T_3 = 121$. Given a point \mathbf{x}_i of Ω , we extract a Zernike moment (where the order is 5) from a 8×8 neighborhood around the point \mathbf{x}_i and call it ZM_i . For the point \mathbf{x}'_j of Ω' (where $1 \leq j \leq T_3$), we also extract a Zernike moment (where the order is 5) from a 8×8 neighborhood around the point \mathbf{x}'_j and call it ZM'_j . The feature ZM_i of the given point \mathbf{x}_i from Ω is compared with the features of any points from Ω' within the Euclidean distances which they are denoted as $\|ZM_i, ZM'_j\|_2$. It is well know that the matched points have the similar features. So the point \mathbf{x}_i and its matched point \mathbf{x}'_j should satisfy the following condition:

$$\|ZM_i, ZM'_j\|_2 \leq T_1. \quad (4)$$

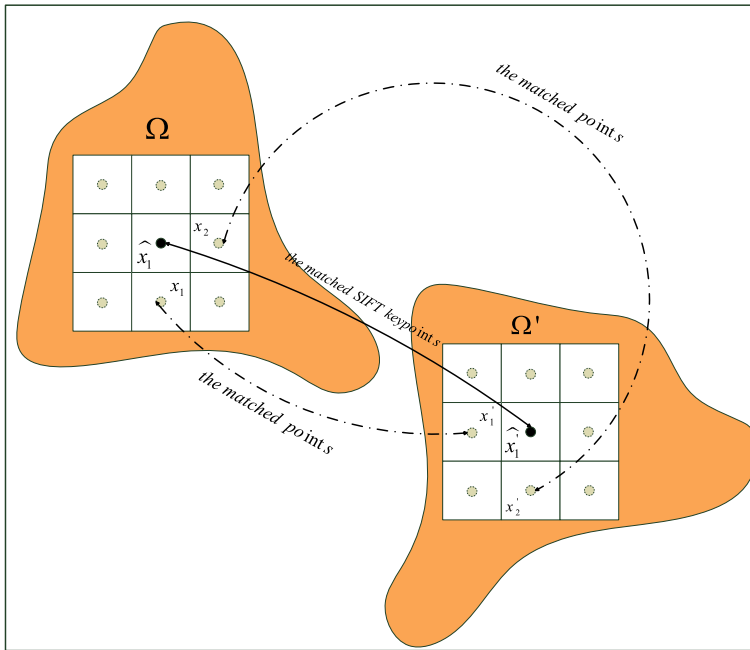


Fig. 3 The Keypoint Contexts (KC) approach. The *black points* are the SIFT keypoints. The *green points* are the points around the keypoints. Among them, \hat{x}_1 and \hat{x}'_1 are the matched SIFT keypoints, which are connected by a *solid line*. There are two pair of points. i.e. $(\mathbf{x}_1, \mathbf{x}'_1)$ and $(\mathbf{x}_2, \mathbf{x}'_2)$, which are connected by a *dash-dotted line*. Then a transformation matrix is estimated from the three pair of points, including a pair of SIFT keypoints

In our implementation, we empirically set $T_1 = 300$ [20]. Because the feature of point in close spatial are usually similar, the spatial distance of the point \mathbf{x}_i and its matched point \mathbf{x}'_j should satisfy the following condition:

$$\|\mathbf{x}_i, \mathbf{x}'_j\|_2 \geq T_2. \tag{5}$$

Since the adjacent points might have similar features, the distance threshold T_2 is defined as 20. Then a pair of points is obtained, which is denoted as $(\mathbf{x}_i, \mathbf{x}'_j)$ and connected by a dash-dotted line, as shown in Fig. 3. In fact, for each point in region Ω , using the same approach, its matched point in region Ω' may be obtained. Finally, a transformation matrix is tried to be estimated if the number of pairs of points is larger than three.

The important meaning of KC approach is that a transformation matrix can be estimated from only a pair of keypoints, and some small duplicated regions with few keypoints can be detected by the KC approach. The matched keypoints in the unmatched patches are processed in order. Thus, all the suspicious duplicated regions are processed. A pair of keypoints and the corresponding duplicated regions are shown in Fig. 2f and g. In our experiment, the images in the dataset are resized to no more than 800 pixels. There are only a pair of keypoints between the two patches in Fig. 2f. It is obvious that this situation cannot be detected by the traditional keypoint-based methods, for there are less than three pairs of keypoints between the two patches. But the duplicated regions can be detected by the KC approach, as shown in Fig. 2g.

As discussed above, the proposed method integrates the ideas of both keypoint-based and block-based schemes. It is obviously that the keypoint-based idea is adopted in the primary region duplication detection. Then a Keypoint Contexts (KC) scheme, which is similar to the block-based scheme, is adopted in the supplementary region duplication detection. A transformation matrix is tried to be estimated from a pair of keypoints by the KC scheme. We select a region Ω around a SIFT keypoint and a region Ω' around its matched SIFT keypoint. A feature is extracted from the neighborhood of a point. Each point in the region Ω and Ω' is represented by a feature, then the points are matched between the region Ω and Ω' . The idea of block-based schemes is adopted by these steps.

All the duplicated regions obtained in Section 3 are combined together and denoted as D_2 . The final duplicated regions (denote as D) are to combine the detection results of the Section 2 (denote as D_1) and the Section 3 (denote as D_2). In Fig. 2h, the final duplicated regions D are illustrated.

4 Experiments and discussions

4.1 Test dataset

The Image Manipulation Dataset (IMD) constructed by Christlein et al. [5] is used to test the proposed method. It is formed based on 48 high-resolution uncompressed base images and the size of images is from 800×533 to 3888×2592 . In the experiment, five kinds of attacks are considered, including plain copy-move, scaling, rotation, adding noise and JPEG compression. In summary, the dataset has 1488 images in total.

1. *Plain copy-move*: There are 48 original images and 48 plain copy-move images. In this case, we will test a total of 96 images.
2. *Scaling*: The copied regions are scaled from 0.91 to 1.09 of its original size, with a step of 0.02. In this case, we will test a total of 480 images.
3. *Rotation*: The copied regions are rotated, and the rotation angle is from 2° to 10° , with a step of 2° . In this case, we will test a total of 240 images.
4. *Adding noise*: Zero-mean Gaussian noise with standard deviations is from 0.02 to 0.1, with a step of 0.02. In this case, we will test a total of 240 images.
5. *JPEG compression*: The quality factors vary from 20 to 100, with a step of 10. In this case, we will test a total of 432 images.

The smaller the image, the less the number of keypoints. Thus, the forgery is more difficult to be detected. Many images in the internet are usually small. So in our experiment, the width and the height of the images are set to no larger than 800 pixels. The proposed method is rather challenging for the keypoint-based schemes. Each image is segmented into 100 patches by using SLIC algorithm. The detection results of the proposed method are shown in Fig. 4, the first row is the forgery images that are selected from the dataset, the second row is the ground truth of the corresponding images, and the third row shows the detection results of the proposed method.

4.2 Evaluation criteria

In the following experiments, we focus on performances at two levels: the image level and the pixel level. The former focuses on whether an image has been tampered or not, and the later focuses on the veracity of the tampered regions. In plain copy-move, we have 48

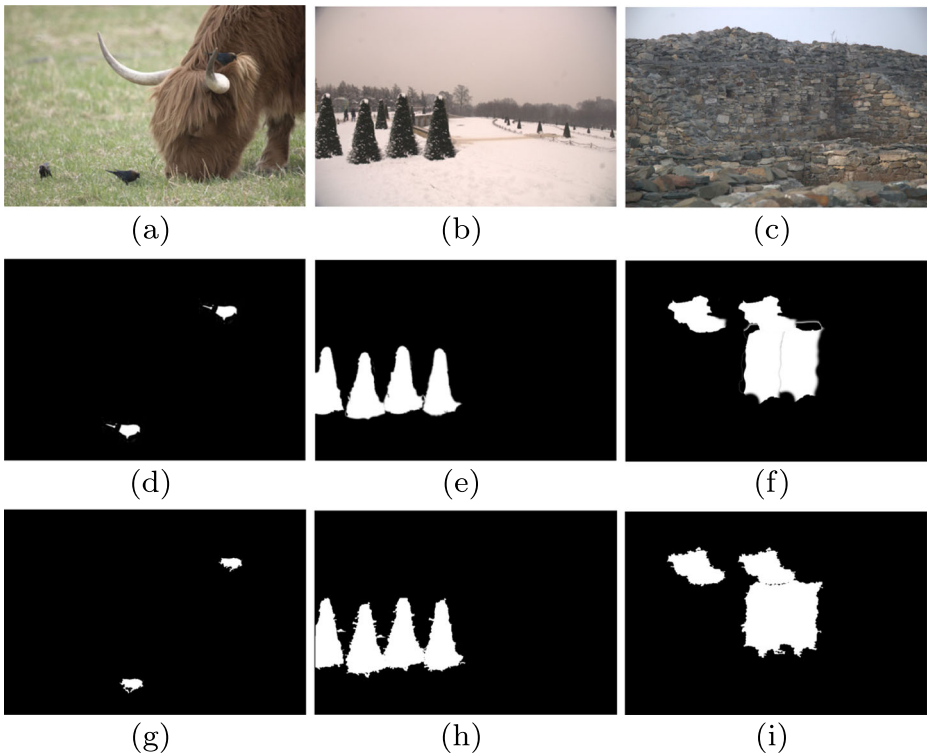


Fig. 4 Region duplication detection results of the proposed method: **a**, **b** and **c** are the forgery images; **d**, **e** and **f** are the ground truth of these images; **g**, **h** and **i** are the detection results of the proposed method

original images as well as 48 images with plain copy-move forgery. Because there are both negative samples and positive samples in the experiment, the evaluation criteria at the image level is measured by the False Positive Ratio (FPR) and the False Negative Rate (FNR) [14]. The FPR and the FNR are defined as:

$$FNR = \frac{FN}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

True Positive (TP) is a result which indicates that a duplication has been detected, when it actually has been duplicated. True Negative (TN) is a result which indicates that a duplication has not been detected, when it actually has not been duplicated. False Positive (FP)

Table 1 Detection results for plain copy-move at the image level

Methods	FNR (%)	FPR (%)
SIFT [2, 16]	54.17	18.75
SURF [21]	56.25	16.67
Li [14]	16.67	35.42
proposed	8.33	14.58

is a result which indicates that a duplication has been detected, when it actually has not been duplicated and False Negative (*FN*) is a result which indicates that a duplication has not been detected, when it actually has been duplicated. As all images are tampered in the robust test, the evaluation criteria at the pixel level is measured by the *precision*, the *recall*, and the *F₁* score [5]. The *precision* denotes the probability that a detected forgery is truly a forgery, while the *recall* shows the probability that a forged is detected. The *precision* and the *recall* are defined as following:

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

The *F₁* score is a measure that combines the *precision* and the *recall*, when the *precision* and the *recall* are evenly weighted. The *F₁* score is defined as following:

$$F_1 = \frac{2 * precision * recall}{precision + recall} \tag{10}$$

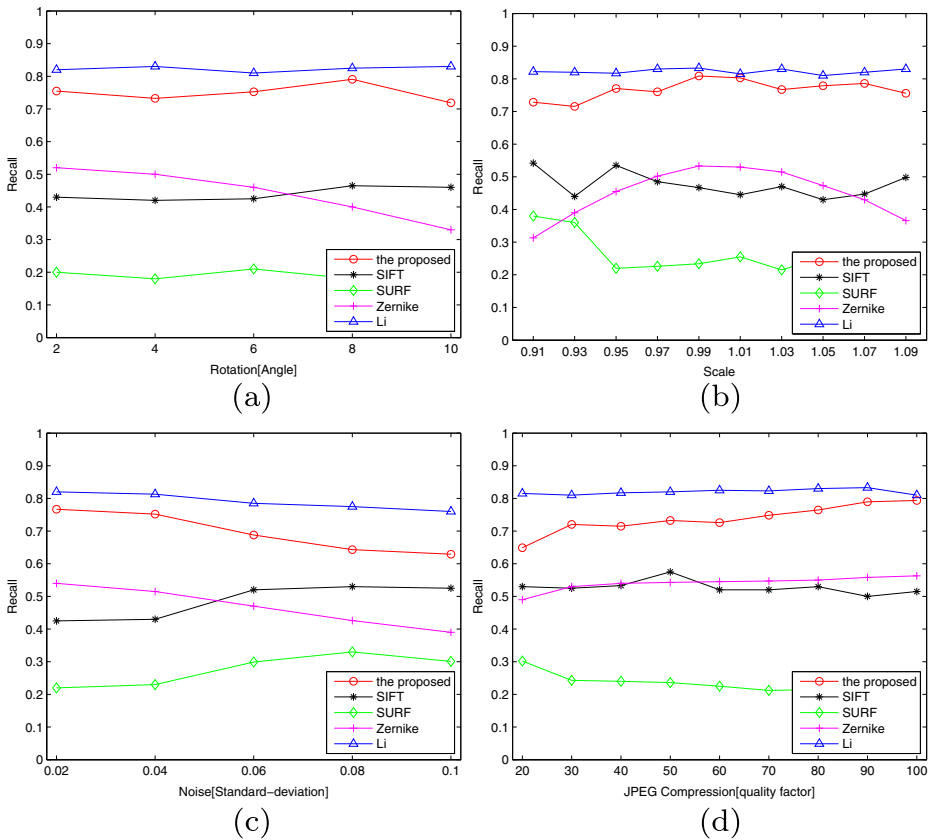


Fig. 5 Recall results at the pixel level: **a** Rotation, **b** Scale, **c** Adding noise and **d** JPEG Compression (see text for details)

4.3 Results on the dataset

4.3.1 Detection results under plain copy-move

Table 1 lists the experimental results under plain copy-move at the image level. The proposed method is compared with SIFT [2, 16], SURF [21] and Li et al. [14]. Considering the image resizing, the results are different with Christlein et al. [5]. Comparing with the state-of-the-art methods, the *FNR* of the proposed method is the smallest, which means that the ratio of the missing detection to the forged images of the proposed method is the lowest. Meanwhile, the *FPR* of the proposed method is also the smallest, which means that the ratio of the false alarm to the original images of the proposed method is the lowest. So both the *FNR* and the *FPR* of the proposed method are much better than that of the existing methods.

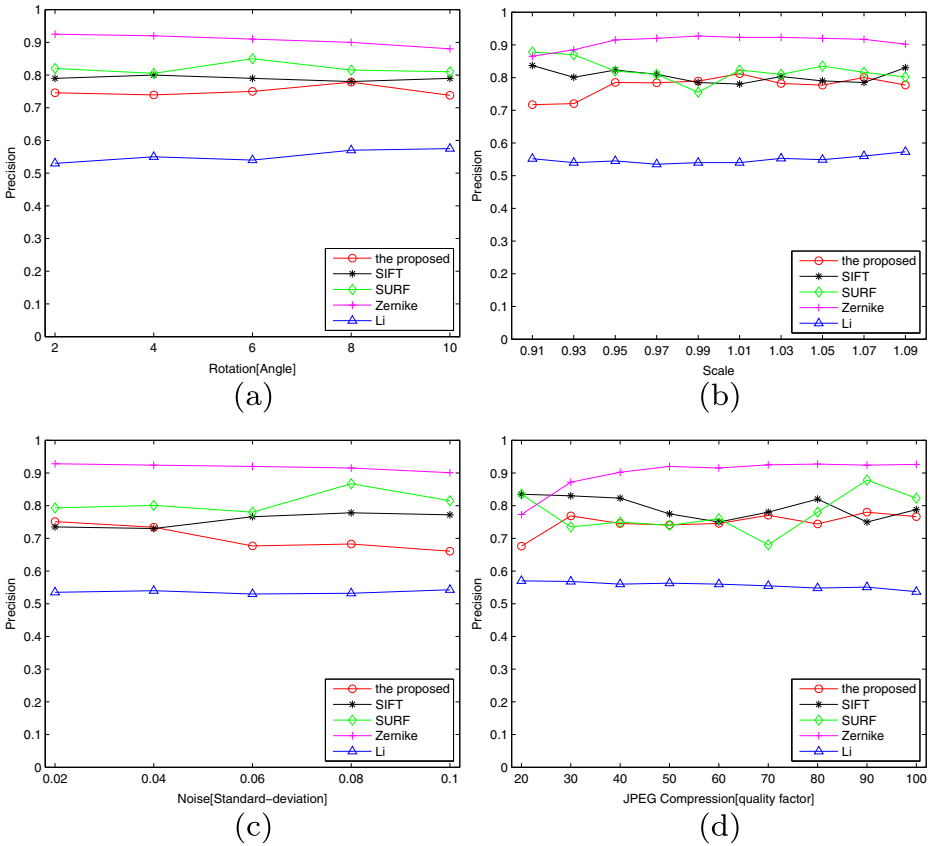


Fig. 6 Precision results at the pixel level: **a** Rotation, **b** Scale **c** Adding noise and **d** JPEG Compression (see text for details)

4.3.2 Detection results under other attackers

Now we test the robustness of the proposed method against various attacks. In this test, four kinds of attacks are evaluated, including scaling, rotation, adding noise and JPEG compression. Besides the above-mentioned methods in plain copy-move, the proposed method is compared with Zernike [20] at the pixel level. Considering the image resizing, the results are different with Christlein et al. [5]. In Figs. 5, 6, and 7, the x-axis in (a) represents the rotation angle, (b) represents the scale factor, (c) represents the white Gaussian noise, and (d) represents the quality factor. The recall results at the pixel level are shown in Fig. 5. It can be observed that the recall of Li et al. [14] is the highest among all the tested methods, followed by the recall of the proposed method. It means that the proposed method can find the second largest number of duplicated regions. Similarly, the precision results at the pixel level are shown in Fig. 6. Among all the tested methods, the precision of Zernike is the highest, the precision of the proposed method is in the middle, and the precision of Li is the lowest. The F_1 scores at the pixel level are shown in Fig. 7. Obviously, the proposed

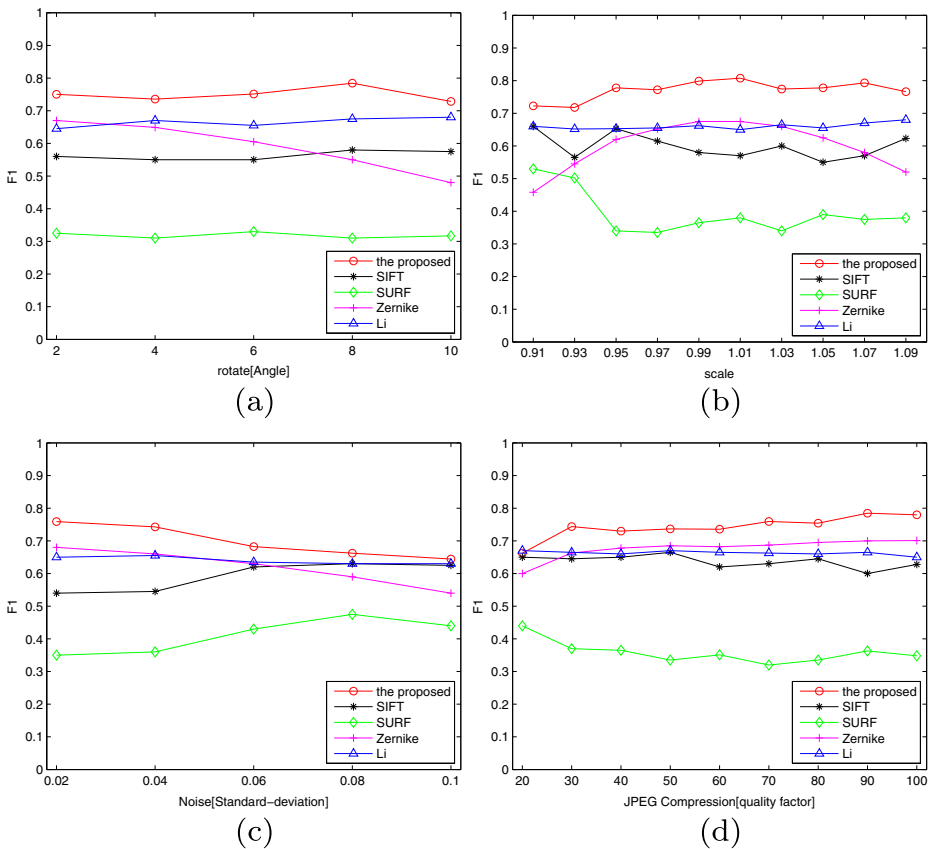


Fig. 7 F1 results at the pixel level: **a** Rotation, **b** Scale, **c** Adding noise and **d** JPEG Compression (see text for details)

method outperforms the state-of-the-art methods in term of F_1 scores under various challenging conditions. The F_1 score combines both the *precision* and the *recall* into a single value. It is a comprehensive evaluation. Therefore, the proposed method is the best among all the tested methods under the various attacks.

The *recall* of Li is the highest among all the tested methods, but the *precision* of Li is the lowest, so the F_1 scores of Li is in the middle among all the tested methods. The *precision* of Zernike is the highest among all the test methods, but the *recall* of Zernike is worse than that of many methods, so the F_1 of Zernike is in the middle among all the test methods. The recall of the proposed method is greater than or equal to the average of the recall, and so is the precision. Considering the trade-off between the precision and the recall, the F_1 score of the proposed method is better than that of Li and Zernike. The *recall* of SIFT and SURF is lower than the proposed, and the *precision* of them are as good as that of the proposed method, so the F_1 scores of them are lower than that of the proposed method. The proposed method is with the best F_1 scores among all the existing state-of-the-art methods under the various attacks. The F_1 score is a comprehensive evaluation. Therefore, the proposed method is the best among all the tested methods under various challenging conditions.

5 Conclusion

In this paper, a novel method based on image segmentation and keypoint contexts (KC) is proposed. The proposed method is divided into two phases. In the primary region duplication detection, a transformation matrix is tried to be estimated from a pair of patches and the duplicated regions D_1 are obtained. When the estimation of affine transformation fails, in the supplementary region duplication detection, a transformation matrix is tried to be estimated from a pair of keypoints by the proposed Keypoints Contexts (KC) approach and the duplicated regions D_2 are obtained. Finally, the total duplicated regions are $D = D_1 \cup D_2$.

The aim of the proposed method is to improve the F_1 scores. As a comprehensive evaluation, the F_1 score is a measure that combines the precision and the recall. If we want to improve the F_1 score, we must improve the recall and the precision first. In order to improve the recall, a Keypoint Contexts (KC) approach is proposed in the supplementary region duplication detection. A transformation matrix is tried to be estimated from only a pair of keypoints by the KC approach. But a transformation matrix is tried to be estimated from at least three pairs of keypoints in the traditional keypoint-based methods. Hence, some duplicated regions with less than three pairs of keypoints cannot be detected by those methods. As a result, the recall is improved by the proposed method. At the same time, some transformation matrices with mistakes are abandoned before the duplicated regions are located, so the precision is improved by the proposed method. Both the recall and the precision are improved, so the F_1 score is improved by the proposed method.

The main contributions of the proposed method can be concluded as the following aspects. First, the proposed method integrates both keypoint-based and block-based region duplication detection schemes, the primary region duplication detection is based on keypoints and the supplementary region duplication detection is based on blocks. Second, a Keypoint Contexts (KC) approach is proposed when affine transformation can not be estimated by the traditional keypoint-based methods, and a transformation matrix is tried to be estimated from a pair of keypoints in the unmatched patches. Third, up-sampling is used in the primary region duplication detection when there are fewer pairs of keypoints in the image or none pair of patch is obtained. Therefore, more keypoints are acquired to estimate affine transformations and the detection results will be more accurate. The proposed method

is robust and adopted to small images with fewer keypoints. The experimental results show that the proposed method is the best among all the tested methods when under the various attacks.

Acknowledgments This work is supported by the Natural Science Foundation of Guangdong (No. 2016A030313350), the Special Funds for Science and Technology Development of Guangdong (No. 2016KZ010103), the Fundamental Research Funds for the Central Universities (No. 16lgjc83), Scientific and Technological Achievements Transformation Plan of Sun Yat-sen University, the Research Project of Guangdong University of Finance and Economics (No. 10GL52001), Shanghai Sailing Program (No. 17YF1420000), the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (Contract No. RCS2016K007), and the Science and Technology Development Fund of Macao SAR (No. 097/2013/A3).

References

- Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
- Amerini I, Ballan L, Caldelli R, Bimbo AD, Serra G (2011) A SIFT-based forensic method for copy-move attack detection and transformation recovery. *IEEE Trans Inf Forensics Secur* 6(3):1099–1110
- Amerini I, Ballan L, Caldelli R, Bimbo AD, Tongo LD, Serra G (2013) Copy-move forgery detection and localization by means of robust clustering with J-Linkage. *Signal Process Image Commun* 28(6):659–669
- Bay H, Ess A, Tuytelaars T, Gool LV (2008) SURF: speeded up robust features. *Comput Vis Image Underst* 110(3):346–359
- Christlein V, Riess C, Jordan J, Riess C, Angelopoulou E (2012) An evaluation of popular copy-move forgery detection approaches. *IEEE Trans Inf Forensics Secur* 7(6):1841–1854
- Farid H (2009) Photo fakery and forensics. *Adv Comput* 77:1–55
- Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24(6):381–395
- Fridrich J, Soukal D, Lukáš J (2003) Detection of copy-move forgery in digital images. In: *Proceeding of digital forensic research workshop*, pp 19–23. Cleveland, OH, USA
- Hartley R, Zisserman A (2004) *Multiple view geometry in computer vision*. Cambridge University Press
- Hastie T, Tibshirani R, Friedman J (2003) *The elements of statistical learning*. Springer Series in Statistics, Berlin
- Huang H, Guo W, Zhang Y (2008) Detection of copy-move forgery in digital images using SIFT algorithm. In: *IEEE Pacific-Asia workshop on computational intelligence and industrial application*, vol 2, pp 272–276
- Huang Y, Lu W, Sun W, Long D (2011) Improved DCT-based detection of copy-move forgery in images. *Forensic Sci Int* 206(1–3):178–184
- Khan S, Kulkarni A (2010) Reduced time complexity for detection of copy-move forgery using discrete wavelet transform. *Int J Comput Appl* 6(7):31–36
- Li J, Li X, Yang B, Sun X (2015) Segmentation-based image copy-move forgery detection scheme. *IEEE Trans Inf Forensics Secur* 10(3):507–518
- Lowe DG (2004) Distinctive image features from scale-invariant keypoints. In: *International journal of computer vision*, pp. 91–110
- Pan X, Lyu S (2010) Region duplication detection using image feature matching. *IEEE Trans Inf Forensics Secur* 5(4):857–867
- Popescu AC, Farid H (2004) Exposing digital forgeries by detecting duplicated image regions. *Tech. Rep. TR2004-515*, Department of Computer Science, Dartmouth College
- Pun CM, Yuan XC, Bi XL (2015) Image forgery detection using adaptive over-segmentation and feature points matching. *IEEE Trans Inf Forensics Secur* 10(8):1705–1716
- Ryu SJ, Kirchner M, Lee MJ, Lee HK (2013) Rotation invariant localization of duplicated image regions based on Zernike moments. *IEEE Trans Inf Forensics Secur* 8(8):1355–1370
- Ryu SJ, Lee MJ, Lee HK (2010) Detection of copy-rotate-move forgery using Zernike moments. In: *IEEE International workshop on information hiding (IH)*. Springer, Berlin, pp 51–65
- Shivakumar BL, Baboo S (2011) Detection of region duplication forgery in digital images using SURF. *Int J Comput Sci Issues* 8(4):199–205
- Vedaldi A, Fulkerson B (2010) *Vfeat: an open and portable library of computer vision algorithms*. In: *International conference on multimedia 2010*, Firenze, Italy, October, pp 1469–1472



Cong Lin received the B.S. degree in computer science from Tsinghua University, Beijing, China in 2005, the M.S. degree in computer science from Sun Yat-sen University, Guangzhou, China in 2009. He is currently working toward the Ph.D. degree in computer science at the Sun Yat-sen University, Guangzhou, China. His research interests include multimedia forensics and security.



Wei Lu received the B.S. degree in Automation from Northeast University, China in 2002, the M.S. degree and the Ph.D. degree in Computer Science from Shanghai Jiao Tong University, China in 2005 and 2007 respectively. He was a research assistant at Hong Kong Polytechnic University from 2006 to 2007. He is currently an Associate Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include multimedia forensics and security, multimedia signal processing, image/video intelligent analysis.



Wei Sun received the Ph.D. degree in computer science from Sun Yat-sen University, Guangzhou, China in 2004, where he is currently a Professor with the School of Electronics and Information Engineering. His current research interests include information security and computer graphics.



Jinhua Zeng received the Ph.D. degree from Tongji University, Shanghai, China in 2013. He is currently a senior engineer, assistant professor and judicial expert in multimedia forensics in the Institute of Forensic Science, Ministry of Justice, China. His main research interests include multimedia forensics, computer vision, and intelligent human-computer interaction.



Tianhua Xu received the B.S. degree from Xi'an Jiaotong University, Xi'an, China; the M.S. degree from Shanghai Jiaotong University, Shanghai, China; and the Ph.D. degree from Xidian University, Xi'an, in 1993, 2002, and 2005, respectively, all in electrical and information engineering. Between 2005 and 2007, he was a Postdoctoral Fellow with the State Key Laboratory for Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China. from 2009 to 2010, he was a Visiting Scholar with the Department of Computer Science, The University of York, York, U.K. He is currently an Associate Professor with Beijing Jiaotong University. His current research interests include computer control, fault diagnosis, discrete-event systems, and hybrid systems.



Jian-Huang Lai received the Ph.D. degree in mathematics from Sun Yat-sen University in 1999. He is currently a Professor of the School of Data and Computer Science, Sun Yat-sen University. He has authored over 100 scientific papers in international journals and conferences, including the IEEE TPAMI, the IEEE TNN, the IEEE TIP, the IEEE TSMC-B, PR, ICCV, CVPR, and ICDM. His current research interests are in the areas of digital image processing, pattern recognition, multimedia communication, and wavelet and its applications.