

# Effective and Efficient Global Context Verification for Image Copy Detection

Zhili Zhou, Yunlong Wang, Q. M. Jonathan Wu, *Senior Member, IEEE*,  
 Ching-Nung Yang, and Xingming Sun, *Senior Member, IEEE*

**Abstract**—To detect illegal copies of copyrighted images, recent copy detection methods mostly rely on the bag-of-visual-words (BOW) model, in which local features are quantized into visual words for image matching. However, both the limited discriminability of local features and the BOW quantization errors will lead to many false local matches, which make it hard to distinguish similar images from copies. Geometric consistency verification is a popular technology for reducing the false matches, but it neglects global context information of local features and thus cannot solve this problem well. To address this problem, this paper proposes a global context verification scheme to filter false matches for copy detection. More specifically, after obtaining initial scale invariant feature transform (SIFT) matches between images based on the BOW quantization, the overlapping region-based global context descriptor (OR-GCD) is proposed for the verification of these matches to filter false matches. The OR-GCD not only encodes relatively rich global context information of SIFT features but also has good robustness and efficiency. Thus, it allows an effective and efficient verification. Furthermore, a fast image similarity measurement based on random verification is proposed to efficiently implement copy detection. In addition, we also extend the proposed method for partial-duplicate image detection. Extensive experiments demonstrate that our method achieves higher accuracy than the state-of-the-art methods, and has comparable efficiency to the baseline method based on the BOW quantization.

**Index Terms**— Image copy detection, near-duplicate detection, partial-duplicate detection, global context, overlapping region.

Manuscript received March 16, 2016; revised July 11, 2016; accepted August 9, 2016. Date of publication August 17, 2016; date of current version October 31, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61602253, Grant U1536206, Grant 61232016, Grant U1405254, Grant U1536203, Grant 61502242, and Grant 61572258, in part by the Jiangsu Basic Research Programs-Natural Science Foundation under Grant BK20150925, in part by the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology under Grant 2014r024, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Tubaro.

Z. Zhou, Y. Wang, and X. Sun are with the Jiangsu Engineering Center of Network Monitoring, School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: zhou\_zhili@163.com; youngon\_wyl@163.com; sunnudt@163.com).

Q. M. J. Wu is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: jwu@uwindsor.ca).

C.-N. Yang is with the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien City 974, Taiwan (e-mail: cnyang@mail.ndhu.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2016.2601065

## I. INTRODUCTION

WITH the rapid development of network technologies and the wide use of various powerful multimedia processing tools, digital multimedia (image, video and audio) is becoming easier to be replicated, modified and distributed on networks [1], [2]. To protect owners against unauthorized (re)use of their content, detecting illegal copies of digital multimedia is a basic requirement [1], [3], [4]. In recent years, content-based image copy detection has been researched as a passive technology to detect illegal copies. Different from watermarking, which uses previously embedded marks, this technology extracts content-based features from images and then searches for the copies by matching the extracted features. The main advantages of content-based image copy detection are that it does not need additional information and copy detection can be implemented after image distribution [1], [3], [5]. In addition, content-based copy detection technology can be applied to some emerging applications, such as automatic annotating [6], [7], redundancy elimination [8], and merchandize image retrieval [9].

Content-based copy detection is similar to near-duplicate detection. These technologies include two main parts: content-based feature extraction and image search based on the extracted features. However, according to the definitions in the literature [5], [10], there is a difference between the two technologies. The former aims to search for copies of an original (a copyrighted) image, which include both the exact duplicates and the transformed versions generated by various copy attacks such as rotation, scaling, cropping, intensity and contrast changes and noise addition. The original image and its copies are assumed to share the same digital source [5], [10]. The objective of the latter is not only to find the copies, but also similar images captured from the same scene or object by different capturing conditions such as different acquisition time, viewpoints and positions. Fig. 1 shows several examples to illustrate the difference between copies and similar images. Fig. 1(b) and (c) are copies derived from Fig. 1(a) by some copy attacks, while Fig. 1(d), (e) and (f) are similar images of Fig. 1(a) obtained by different capturing conditions. The three similar images are near-duplicates but not copies of Fig. 1(a), as they do not share the same source with Fig 1(a).

In practice, among the huge number of near-duplicate images distributed on the networks, there are a lot of similar images. It is worth noting that similar images that are not

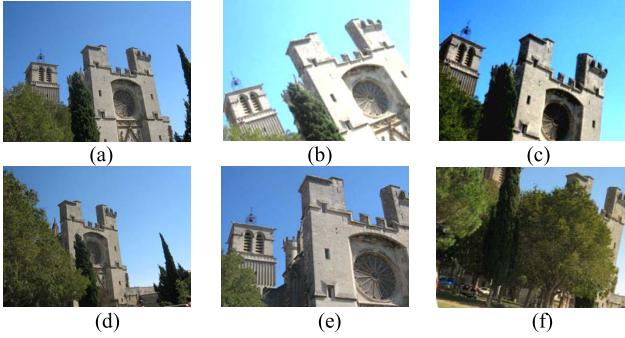


Fig. 1. The difference between image copies and similar images. (a) is an original image. (b) and (c) are copies of the original image. (d)-(f) are similar images.

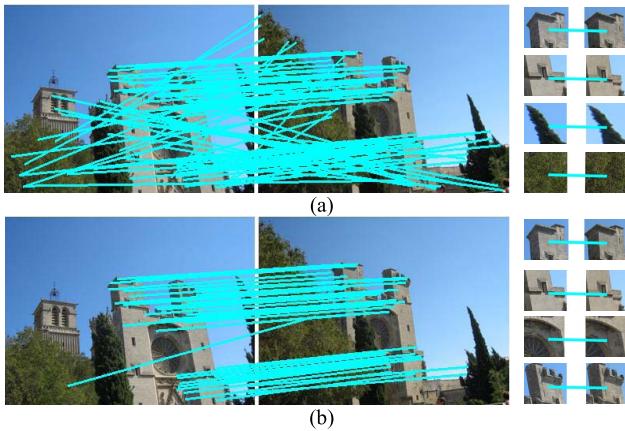


Fig. 2. The illustration of limited performances of BOW model and geometric verification. (a) The 83 false SIFT matches, which are generated by matching SIFT features based on the quantization of BOW model, where the size of visual codebook is set to 50K according to our experimental section. (b) The 47 false SIFT matches, which still remain after removing most of the geometrically inconsistent matches by the algorithm [17] with its default parameters. Several pairs of visually similar local patches (non-copies) corresponding to the falsely matched SIFT features are also illustrated on the right side.

copies could be more visually similar to each other than the copies generated by strong copy attacks [5], which makes copy detection quite challenging. Thus, in this paper, we focus on how to effectively and efficiently detect copies of a given original (query) image in a large-scale database, in which there are not only many copies but also a lot of similar images.

In the past few years, the bag-of-visual-words (BOW) model [11] has become very popular for large-scale image retrieval tasks. Generally, local features extracted from images are quantized into visual words, and then they are indexed with inverted file structure for image search. While favorable for efficiency and scalability, the BOW model has two drawbacks when directly applied to copy detection. The first is that local features do not encode enough spatial information and thus have limited discriminability. The second is that the BOW quantization errors will further degrade the discriminability of local features. Both the two drawbacks will lead to many false matches when there are many visually similar local patches, typically found between the similar images. That is illustrated by the toy example in Fig. 2(a). The two similar images, which share many visually similar local patches, have

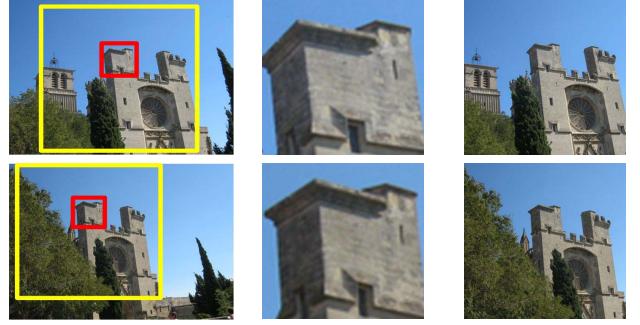


Fig. 3. The corresponding local patches and global context regions of a pair of falsely matched SIFT features between the two similar images.

83 scale invariant feature transform (SIFT) matches generated by matching their SIFT features based on the quantization of BOW model. Note that the two corresponding local patches of each match are visually quite similar but not copies of each other, and thus all of these SIFT matches can be regarded as false matches. The existence of these false matches will cause many similar images to be falsely detected as copies, resulting in low accuracy for copy detection.

Recently, many geometric verification strategies [7], [12]–[17] have been proposed to reduce false matches. After obtaining local matches between images based on the BOW quantization, these strategies check geometric consistency among the matches to reduce the number of false matches. This process can be quite efficient, since the number of matched features is much smaller than the number of extracted features [16], [17]. However, these strategies are originally designed for near-duplication detection but not for copy detection. Many similar images will be falsely detected as copies when they are employed for copy detection. That is because many matches between similar images may satisfy geometric consistency, and thus cannot be effectively removed. That can be illustrated by the example shown in Fig. 2(b). After removing most of the geometrically inconsistent false matches by the geometric verification algorithm [17], the two similar images still have 47 SIFT matches, most of which satisfy the geometric consistency. Therefore, when the geometric consistency verification strategies are applied to copy detection, it will result in limited improvement of detection accuracy.

Nevertheless, since the variations of capture conditions may lead to some changes in the global appearances of images, in most cases global context information of falsely matched local features between similar images differs to some extent. This can be illustrated by the toy example in Fig. 3. This figure shows the corresponding local patches and global context regions of a pair of falsely matched SIFT features between the two similar images in Fig. 2. The global context regions, which almost cover the whole image, are created by expanding the local patches proportionally. It can be clearly observed that, although the corresponding local patches of the two falsely matched SIFT features are visually similar, their global context is somewhat different. Thus, the global context could provide an overall reference to identify false matches.

Therefore, we propose an effective and efficient global context verification scheme, which explores the global context

information of matched SIFT features to identify and remove the false matches for copy detection. More specifically, after obtaining initial SIFT matches between images based on the BOW quantization, we extract the overlapping region-based global context descriptor (OR-GCD) for each matched SIFT feature to describe its global context information. Then, we compare the corresponding OR-GCDs of each pair of matched SIFT features to filter the false matches. Finally, the verification result is used to measure image similarities to implement copy detection. To improve the detection efficiency, we also propose a fast image similarity measurement based on random verification. In addition, we extend the proposed copy detection method for partial-duplicate image detection. The main contributions of our method are concluded as follows:

- 1) The proposed global context descriptor (OR-GCD) for verification of SIFT matches. In our previous work [18], we propose a convex region-based global context (CRGC) feature for SIFT match verification. However, since the feature is extracted by directly using the histogram of oriented gradients (HOG) [19], it is not discriminative and efficient enough for SIFT match verification. To address this issue, we propose a novel global context descriptor, *i.e.*, OR-GCD. It does not only encode relatively rich global context information of each matched SIFT feature, but also has good robustness and efficiency. Thus, OR-GCD allows an effective and efficient global context verification of SIFT matches for copy detection.

- 2) A fast image similarity measurement strategy based on random verification. To further improve efficiency for copy detection, we also propose a fast similarity measurement strategy based on random verification and show that this strategy can accurately measure image similarities for copy detection using the verification result obtained by verifying only a small number of randomly selected SIFT matches.

- 3) The extended method for partial-duplicate image detection. To deal with the task of partial-duplicate image detection, we also extend the proposed copy detection method by adding a stage of potential duplicated region location after obtaining initial SIFT matches between images.

The rest of this paper is organized as follows. Section II introduces the related works. Section III presents the proposed copy detection method in detail. Section IV presents the fast image similarity measurement strategy based on random verification. Section V details the extended method for partial-duplicate image detection. Section VI presents and discusses the experimental results. Conclusions are drawn in Section VII.

## II. RELATED WORK

Content-based copy detection is similar to near-duplicate detection. Thus, in this section, we not only review the copy detection literature, but also state-of-the-art near-duplicate detection methods.

In the literature, many copy detection methods based on local features have been proposed [10], [20]–[23]. To the best of our knowledge, these methods generally investigate the local features, such as SIFT [24] and its extensions including principal component analysis on SIFT (PCA-SIFT) [25], speeded-up robust feature (SURF) [26], and multi-scale SIFT [10] for detecting image copies. These local features

have shown good robustness to various common types of copy attacks, such as rotation, scaling, cropping, intensity and contrast changes and noise addition. Therefore, these methods perform well in detecting the copies generated by those copy attacks. However, the robustness generally comes at the expense of detection efficiency, because matching these local features between images is usually computationally intensive due to their high-dimensionality. More importantly, since local features are extracted from small local patches and thus encode less spatial context information, they have limited discriminability. As a result, many false matches will occur when there are many visually similar local patches, typically found between the similar images [20]. As a result, some similar images will be falsely detected as image copies, which will significantly affect detection accuracy. Recently, the BOW model has become popular for large-scale content-based image retrieval. It quantizes the extracted local features into visual words and then indexes images using an inverted file structure for image search. To improve efficiency, many copy detection methods relying on the BOW model have been proposed [27]–[29]. Although they can achieve efficiency, more false matches will occur between similar images, since the BOW quantization errors will further degrade the discriminability of local features. Consequently, the accuracy of copy detection will be further decreased.

State-of-the-art near-duplicate detection methods such as [7], [12], [16], [17], [30]–[34] also rely on the BOW model. To reduce false matches for near-duplicate detection, two main kinds of strategies have been proposed: incorporating spatial information into image representation and geometric consistency verification.

Incorporating spatial information entails encoding spatial information around local features into image representation to improve the discriminability of local features. In [7], Zhang *et al.* propose a geometry-preserving visual phrase (GVP) to encode more spatial information, which are the co-occurrences and spatial layout of local features. In [31], Wang *et al.* employ the statistics in the local neighborhood of each local feature as its spatial context to improve its discriminability. In [32], Zheng *et al.* propose a multi-IDF scheme to embed the binary color information of local features into the inverted index file. In [33], Yao *et al.* design a contextual descriptor that encodes the relative spatial relationship between each local feature and its neighbor features to strengthen the discriminability of local features. Since the spatial information within local areas instead of the whole image plane is considered, the above strategies do not encode enough spatial information, resulting in limited discriminability improvement. Inspired by shape context [35], [36], Mortensen *et al.* [37] augment SIFT feature with a high-dimensional global context vector, which describes curvilinear shape information from a much larger neighborhood. In [38], Liu *et al.* propose an improved global context vector to enhance the discriminability of SIFT features. However, these global context vectors are not fully scale invariant and are very sensitive to cropping attacks. Therefore, these methods cannot detect the images after scaling and cropping attacks. Moreover, since high-dimensional global context vectors are extracted for each SIFT feature

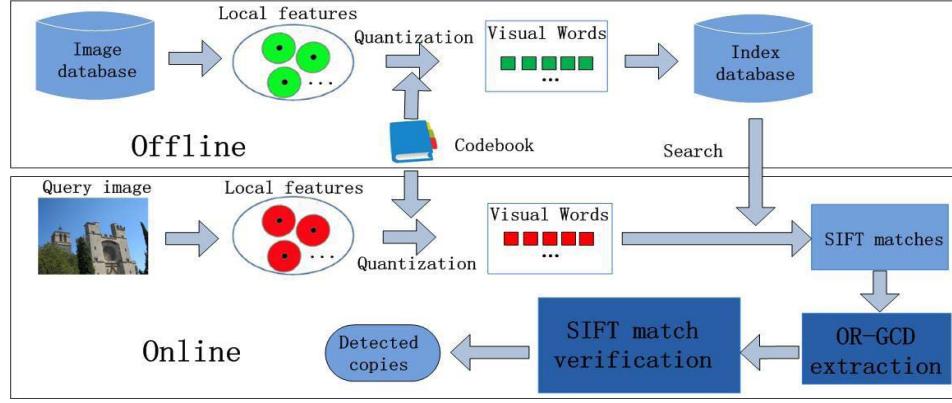


Fig. 4. The framework of our copy detection method.

before feature matching, these methods make image representation very complex, resulting in high computational cost in the steps of feature extraction and matching.

The second kind of strategy focuses on implementing geometric verification to filter false matches. Different from the first kind, it does not change the image representation or image-matching algorithm. Instead, local matches are first obtained between images, and the geometric consistency among the matches is then utilized for verification of these matches to filter the false matches that are geometrically inconsistent. Since the number of matched features is much smaller than that of extracted features in images, these strategies can be quite efficient for near-duplication detection [16], [17]. In [12], Jegou *et al.* propose a weak geometric consistency (WGC) scheme. It verifies the consistency of the angle and scale parameters for matched features to filter the false matches. With an additional assumption that the correct matches also follow consistent translation transformation, Zhao *et al.* [7] improve the WGC by adding translation information. To fully capture geometric relationships of local features, a global geometric consistency verification strategy, *i.e.*, RANSAC [13]–[15], is very popular for this task. It randomly samples several pairs of local matches many times to estimate the affine transformations between images, and then verifies the geometric consistency of local matches to filter out those that are false. However, RANSAC can only be applied to a small number of top-ranked candidate images due to its high computational complexity. In [16], Zhou *et al.* propose a spatial coding method to remove false matches based on spatial maps. Unfortunately, it cannot handle rotation transformation well. To address this problem, Zhou *et al.* [17] propose a novel geometric coding scheme, which describes the geometric relationships among SIFT features in three geo-maps for verification of SIFT matches. It can efficiently and effectively filter the false matches that are geometrically inconsistent under rotation and scaling transformations, partial-occlusion, and background clutter. Note that all of those strategies are originally designed for near-duplication but not for copy detection. Many false matches between similar images may satisfy geometric consistency, and thus they cannot be removed effectively by those strategies. Therefore, if those strategies are directly applied to copy detection, it will result in limited improvement of accuracy.

Instead of resorting to the geometric consistency information, our motivation is to design an effective and efficient verification scheme by exploring the global context information of matched local features. In our previous work [18], a convex region-based global context (CRGC) feature is proposed for SIFT match verification. However, it is extracted by directly using the histogram of oriented gradients (HOG) [19], which has high-dimensionality and does not preserve enough spatial information. Consequently, the CRGC feature is not discriminative and efficient enough for local match verification. This paper proposes a novel global context descriptor, *i.e.*, OR-GCD for the verification of local matches to remove false matches. The OR-GCD not only encodes the global context information of local features but also has good robustness and efficiency, which ensures that false matches can be effectively and efficiently identified and removed. Thus, our method can achieve desirable performances in both accuracy and efficiency for copy detection.

### III. THE PROPOSED COPY DETECTION METHOD

The framework of our copy detection method is shown in Fig. 4. It consists of three main components, which are SIFT feature matching, OR-GCD extraction, and verification of SIFT matches. The main contributions of our method lie in the later two components. In Section III-A, we match SIFT features between images based on the BOW quantization to obtain initial SIFT matches. In Section III-B, we extract the OR-GCD to describe the global context of each matched SIFT feature. In Section III-C, the matched SIFT features are verified by comparing their corresponding OR-GCDs, and then the verification result can be further used to measure image similarities for copy detection.

#### A. SIFT Feature Matching

We extract hundreds of SIFT features from each image by using the SIFT algorithm [24]. The extracted SIFT features are then efficiently matched between images based on the BOW quantization, which is detailed as follows.

A visual codebook including a lot of visual words is first generated through clustering a sample set of SIFT features by the hierarchical visual vocabulary tree approach [39]. Then, each extracted SIFT feature is quantized to its nearest

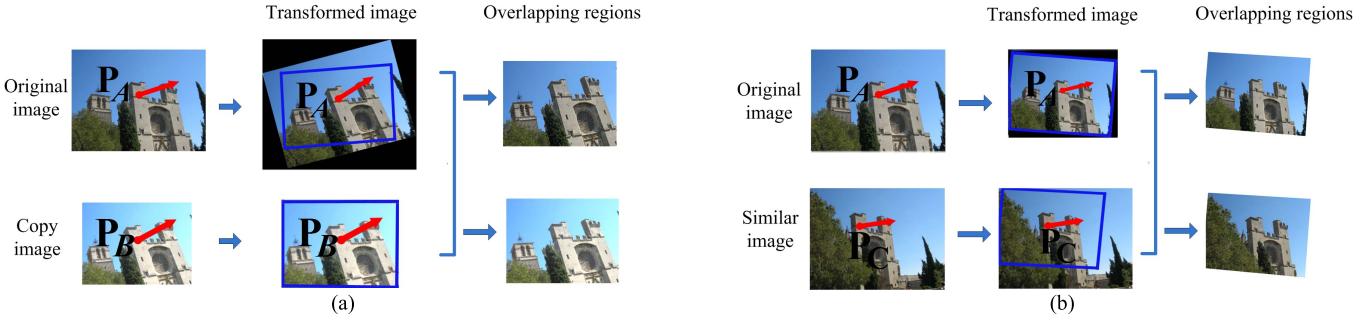


Fig. 5. The illustration of overlapping region construction. (a) The construction between an original image and its copy. (b) The construction between the two similar images.

visual word. Next, these SIFT features are indexed according to their visual words to obtain an inverted index file, in which each indexed feature records the ID of the image to which it belongs, and its dominant orientation  $\phi$ , characteristic scale  $s$ , and coordinates  $(x, y)$ . Note that these information will be further used to extract the OR-GCDs for verification of SIFT matches. By using the inverted index file, any two SIFT features from two images quantized to the same visual word are regarded as a local match between the two images.

### B. OR-GCD Extraction

After the previous stage, we can obtain the initial SIFT matches between images. As indicated in Section I, exploring the global context of matched SIFT features for verification of SIFT matches might be a feasible way to identify and remove the false ones.

Therefore, a corresponding global context descriptor of each matched feature is required to be extracted for the verification. An ideal global context descriptor is expected to encode rich global context information of a matched feature, so that the false matches can be easily distinguished from the correct matches. Meanwhile, the descriptor should be robust to a variety of possible copy attacks, such as rotation, scaling, cropping, intensity and contrast changes, and noise addition, so that the false matches can be easily identified and removed under these attacks. Moreover, it should also have high efficiency, *i.e.*, extraction and comparison should be efficient to compute, so that the verification process is less computationally complex.

To meet the above requirements, we propose a novel global context descriptor, *i.e.*, OR-GCD. Different from the traditional global context vectors [37], [38], which are directly extracted from the whole image region, for each pair of matched SIFT features between two images we use their three property values—the dominant orientations, characteristic scale, and coordinates—to construct the overlapping regions between the images. Then, we generate a pair of OR-GCDs corresponding to the two matched features by extracting and concatenating two kinds of binary vectors—one based on intensities and the other on gradients from the constructed overlapping regions. The extraction of OR-GCD is detailed as follows.

1) *The Construction of the Overlapping Regions Between Images:* To achieve robustness to the common geometric transformations including rotation, scaling, and cropping, we

construct the overlapping regions of images for the global context descriptor extraction. For each pair of matched SIFT features of two images, we use their property values—the dominant orientations and characteristic scales—to adjust the orientations and scales of the two images to make them consistent. We then employ their keypoint coordinates to compute the overlapping regions between the images. The examples in Fig. 5 illustrate the construction of overlapping regions.

Suppose that the two matched SIFT features  $f_A$  and  $f_B$  are a correct match between an original image  $A$  and its copy  $B$ , which is generated by transforming image  $A$  with a combination of the three geometric attacks, which are rotation, scaling, and cropping. We denote the corresponding keypoints of the two features as  $p_A$  and  $p_B$ , their orientation angles as  $\phi_A$  and  $\phi_B$ , and scales as  $s_A$  and  $s_B$ . In Fig. 5, the arrows are used to represent the property values of the features. More specifically, the origin positions of the arrows represent the keypoint positions, and the lengths and orientations of the arrows indicate the dominant orientations and characteristic scales of the features, respectively.

From [24], we know that as SIFT features are extracted based on image properties, their property values—the dominant orientations and characteristic scales—change covariantly with the rotation and scaling transformations. Therefore, we can adjust the orientations and scales of images  $A$  and  $B$  to be consistent according to the dominant orientations and characteristic scales of their features  $f_A$  and  $f_B$ . Taking  $f_B$  as a reference feature and setting the keypoint coordinates of feature  $f_A$  as the origin in image  $A$ , we first rotate image  $A$  by aligning the orientation of  $f_A$  to be the same as that of  $f_B$ . Then, the scale of image  $A$  is adjusted to be the same as that of image  $B$  by multiplying with the ratio between scales  $s_B$  and  $s_A$ . If scale  $s_A$  is larger than scale  $s_B$ , the image is smoothed by a Gaussian kernel before the transformation, where the standard derivation of Gaussian used for smoothing is set to be the ratio of  $s_A$  and  $s_B$  [24]. The transformation can be represented by

$$X'_A = s_B/s_A \begin{pmatrix} \cos(\phi_B - \phi_A) & -\sin(\phi_B - \phi_A) \\ \sin(\phi_B - \phi_A) & \cos(\phi_B - \phi_A) \end{pmatrix} X_A \quad (1)$$

where  $X_A$  denotes the coordinates of any pixel point in image  $A$ , and  $X'_A$  is the adjusted coordinates of the point in the transformed image. Consequently, the orientations and scales of image  $A$  are adjusted to be consistent with those of

image B, as shown in Fig. 5(a). In the following, the images are the transformed ones and the coordinates of all of their pixel points are the adjusted coordinates.

Then, by using the coordinates of keypoints, we compute the overlapping regions between the images. For the pair of matched SIFT features  $f_A$  and  $f_B$  between images A and B, we first align the coordinates of keypoints  $p_A$  and  $p_B$ . Then we compute the overlapping regions of the two images by

$$R_A = \{p_A^i \in A : I(p_A^i) > 0 \text{ } \& \text{ } I(p_B^i) > 0\} \quad (2)$$

$$R_B = \{p_B^i \in B : I(p_A^i) > 0 \text{ } \& \text{ } I(p_B^i) > 0\} \quad (3)$$

where  $I(p_A^i)$  and  $I(p_B^i)$  are intensities of the pixel points  $p_A^i$  and  $p_B^i$ , and  $R_A$  and  $R_B$  are the overlapping regions of images A and B, respectively. Since the overlapping regions are the common regions between the two images, they can cover the same image content under cropping transformation.

After the above steps, as features  $f_A$  and  $f_B$  are a correct match between the original image A and its copy B, the orientations, scales, and content of the constructed overlapping regions can be consistent. That enables us to extract the corresponding global context descriptors of the matched features that are robust to the geometric transformations including scaling, rotation, cropping, and their combinations. From Fig. 5(b), it can be also observed that, for the falsely matched features  $f_A$  and  $f_C$  between image A and its similar image C, the content of constructed overlapping regions of the two similar images cannot be adjusted to be consistent in the same manner. Thus, the corresponding global context descriptors of the two features will be different to some extent, which is helpful in identifying the match as false.

2) *Generation of OR-GCD*: Similar to the shape context vector [35]–[37], OR-GCD is generated using the histograms created in the polar coordinate system. The main difference is that we create the histograms from the constructed overlapping regions instead of the whole image regions for OR-GCD generation.

Note that the overlapping regions might be irregular in shape after the above construction procedure. To facilitate OR-GCD generation, we extend each overlapping region to a rectangle region through padding the black pixels whose intensities are equal to 0, and then normalize it to the square region with a certain size, as shown in Fig. 6(b). As all of the query images used in our experimental section are resized to no larger than  $400 \times 400$ , the size of square regions is set as  $400 \times 400$ .

Next, we create the intensity and gradient histograms in the polar coordinate system, and then extract two kinds of binary vectors—an intensity-based vector and a gradient-based vector—from the two histograms, respectively. The two extracted vectors can be further used to form an OR-GCD. The intensity histogram denoted as  $H_I$  is created by accumulating the intensities of the pixels into  $M \times N$  bins. To create the histogram, we set the widths of  $M$  bins at the radial direction and  $N$  bins at the azimuthal direction in the polar coordinate system by

$$wr_i = \left( \frac{\sqrt{i} - \sqrt{i-1}}{\sqrt{M}} \right) \times r \quad (4)$$

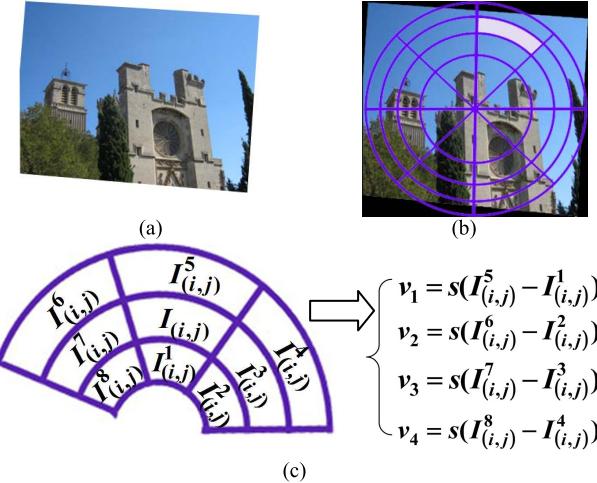


Fig. 6. The generation of the binary vector of each histogram bin. (a) An overlapping region. (b) Its normalized version overlaid with an intensity histogram, which is created by accumulating the pixel intensities into  $4 \times 8$  bins. (c) A bin and its eight neighbors selected from the intensity histogram, and a four-dimensional binary vector generated by comparing the intensity differences of these bins.

$$waj = \frac{2\pi}{N} \quad (5)$$

Where  $r$  is equal to half of the width of the square region, *i.e.*, 200, and  $wri$  and  $waj$  represent the width of  $i$ -th bin in the radial direction and that of  $j$ -th bin in the azimuthal direction, respectively. That can ensure that all of the  $M \times N$  bins are with the same size, which is equal to  $\pi r^2 / MN$ . The parameters  $M$  and  $N$  are set according to our experimental section. The example of a created intensity histogram with  $4 \times 8$  bins in the polar coordinates is shown in Fig. 7(b).

Then, we use the intensity differences of the neighboring bins of each bin to extract the intensity-based vector. For each bin  $b_{(i,j)}$ , its intensity is denoted as  $I_{(i,j)}$ , and the intensities of its eight neighboring bins are denoted as  $(I_{(i,j)}^1, I_{(i,j)}^2, \dots, I_{(i,j)}^8)$ . For a bin located at the edge of the histogram, some of its neighbors do not exist. Thus their intensities are set to zero. To achieve the invariance to the linear intensity and gradient changes, similar to CS-LBP [40], we compute the signs of intensity differences of the center-symmetric bins to form a four-dimensional binary vector represented as  $V_{(i,j)} = (v_1, v_2, v_3, v_4)$ , as shown in Fig. 6(c), where

$$\begin{cases} v_1 = s(I_{(i,j)}^5 - I_{(i,j)}^1) \\ v_2 = s(I_{(i,j)}^6 - I_{(i,j)}^2) \\ v_3 = s(I_{(i,j)}^7 - I_{(i,j)}^3) \\ v_4 = s(I_{(i,j)}^8 - I_{(i,j)}^4) \end{cases} \quad (6)$$

and

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (7)$$

Then, we can concatenate the four-dimensional binary vector of each bin to form the final intensity-based vector, denoted

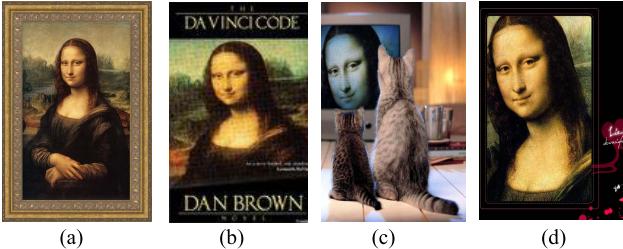


Fig. 7. Examples of partial-duplicate images. (a) is an original image. (b)-(d) are partial duplicates of the original image.

as  $IV = (Iv_1, Iv_2, \dots, Iv_{M \times N \times 4})$ , which has the length of  $M \times N \times 4$ .

The extraction of the gradient-based vector is quite similar to that of the intensity vector. The only difference is that, instead of employing the intensities, we create the gradient histogram  $H_G$  by accumulating the gradient magnitudes of the pixels into the bins, and compute the signs of gradient magnitude differences to form the gradient-based vector, denoted as  $GV = (Gv_1, Gv_2, \dots, Gv_{M \times N \times 4})$ . Finally, we concatenate the two vectors to form a  $M \times N \times 8$  dimensional OR-GCD, represented as  $D = (IV, GV)$ .

### 3) Property Analysis of OR-GCD:

a) *Discriminability*: As analyzed in Section I, since the global context of falsely matched features is usually different to some extent, global context information is quite helpful in identifying false matches. Thus, the overlapping regions constructed between images, which nearly cover the whole image region and contain the global context information of matched features, are suitable for extracting discriminative global context descriptor. In our method, we create the intensity and gradient histograms from the overlapping regions in the polar coordinate system, and then use the two histograms to extract the global context descriptor, *i.e.*, OR-GCD. Since these histograms contain the intensity and gradient information of the histogram bins, and also preserve their relative spatial positions, OR-GCD can encode relatively rich spatial context information, and thus has high discriminability.

b) *Robustness*: Instead of directly employing the whole image region, we construct the overlapping regions of images for global context descriptor extraction. For each pair of matched SIFT features, their three properties are used to construct the overlapping regions. If the two features are a correct match, the scale, orientation, and content of the constructed overlapping regions between images are consistent. Therefore, OR-GCD can be invariant to the common geometric transformations, such as rotation, scaling, cropping, and their combinations.

OR-GCD also has the invariance to linear intensity and contrast changes for the following reasons. A linear change in image intensity in which the intensity of each pixel is added with a constant will not change the intensity differences of the histogram bins. Also, it will not change the gradients as they are computed from intensity differences, and thus the gradient differences of the histogram bins also remain the same. Therefore, OR-GCD is invariant to the linear intensity change, as it is computed using the signs of intensity and gradient differences of histogram bins. A linear contrast change will

make the intensity and gradient of each pixel to be multiplied by a constant. The intensity and contrast differences of the histogram bins will be also be multiplied by the same constant. However, the signs of intensity and contrast differences will not be affected. Therefore, OR-GCD is also invariant to linear contrast change.

In addition, the non-linear intensity and contrast changes and the noise-like attacks, such as Gaussian blurring, compression, and noise addition, will significantly change the intensity and gradient differences of individual pixels. However, instead of employing the intensities and gradients of individual pixels, we accumulate the intensities and gradients of the pixels into histogram bins and use the signs of their intensity and gradient differences for the OR-GCD extraction. The effect caused by the non-linear intensity and contrast changes will be effectively reduced in this manner. As a result, OR-GCD has good robustness to those attacks.

c) *Efficiency*: An OR-GCD is extracted for each matched feature instead of each extracted feature. Note that the number of matched features is much smaller than that of all the extracted features. In addition, OR-GCD is extracted in the spatial domain by only utilizing image intensities and gradients. Therefore, the extraction of OR-GCD can be efficient. On the other hand, according to our experimental part, OR-GCD is only a 256 bit vector, when  $M$  and  $N$  are set appropriately ( $M = 4$ , and  $N = 8$ ). Thus, OR-GCD has good compactness, so that the OR-GCD comparison is also efficient.

All of the above properties of OR-GCD ensure that the verification of SIFT matches can be implemented effectively and efficiently.

### C. Verification of SIFT Matches

In this subsection, we verify each pair of matched SIFT features by comparing their corresponding OR-GCDs, and then use the verification result to measure image similarities to implement copy detection.

For two matched features  $f_A$  and  $f_B$ , their corresponding OR-GCDs are represented as  $D(f_A) = (IV(f_A), GV(f_A))$  and  $D(f_B) = (IV(f_B), GV(f_B))$ . We compute their distance  $Dis(D(f_A), D(f_B))$  and determine whether they are a correct match or not by

$$Dis(D(f_A), D(f_B)) = \alpha \frac{\sum_{i=1}^{M \times N \times 4} |Iv_i(f_A) - Iv_i(f_B)|}{M \times N \times 4} + (1 - \alpha) \frac{\sum_{i=1}^{M \times N \times 4} |Gv_i(f_A) - Gv_i(f_B)|}{M \times N \times 4} \quad (8)$$

$$Dis(D(f_A), D(f_B)) \leq Dis_{TH} \quad (9)$$

where  $Iv_i(f_A)$ ,  $Gv_i(f_A)$ ,  $Iv_i(f_B)$ , and  $Gv_i(f_B)$  are  $i$ -th elements of  $IV(f_A)$ ,  $GV(f_A)$ ,  $IV(f_B)$ , and  $GV(f_B)$ , respectively,  $\alpha$  is a weighting factor, and  $Dis_{TH}$  is a predefined threshold. The two parameters are experimentally set. If  $Dis(D(f_A), D(f_B))$  is no greater than  $Dis_{TH}$ ,  $f_A$  and  $f_B$  can be determined as a correct match; otherwise, they are regarded as a false match and then are removed.

After removing the false matches, we measure the image similarities for copy detection. The traditional cardinality-based similarity measurement strategy has been widely

adopted in many copy detection and near-duplicate detection methods [16], [17], [20], [33], [34]. This strategy directly uses the number of matches judged to be correct, *i.e.*, the number of the surviving matches, as the image similarity. However, since the number of these matches is often dependent on image scene complexity, this strategy usually suffers from the sensitivity problem. To avoid this problem, we use the ratio of the number of the matches judged to be correct as the image similarity for copy detection. Let the number of the matches judged to be correct between a given query (original) image  $Q$  and a test image  $T$  be  $n(Q, T)$ , and the number of all the initial matches between the two images be  $N(Q, T)$ . The image similarity between the two images can be measured by the ratio of  $n(Q, T)$  and  $N(Q, T)$ , which is in the range of 0 to 1. Then, we can determine the test image  $T$  is a copy of the query image  $Q$ , if

$$\frac{n(Q, T)}{N(Q, T)} \geq Rat_{TH} \quad (10)$$

where  $Rat_{TH}$  is a preset threshold that ranges from 0 to 1.

#### IV. FAST IMAGE SIMILARITY MEASUREMENT BASED ON RANDOM VERIFICATION

In the process of verification of SIFT matches, both the extraction and comparison of OR-GCDs have low computational complexity. However, if we verify all initial SIFT matches and then use the verification result to measure image similarities, the detection efficiency will be affected to some extent. In this section, we propose a fast image similarity measurement based on random verification, which verifies a certain small number of randomly selected SIFT matches and then uses the verification result to measure image similarity. Also, we theoretically prove that the proposed similarity measurement strategy can significantly improve efficiency while maintaining high accuracy for copy detection.

##### A. The Fast Image Similarity Measurement

Suppose the set of SIFT matches between a given query (original) image  $Q$  and a test image  $T$  is  $S(Q, T)$ . We randomly select a SIFT match from the set  $S(Q, T)$  and extract the two corresponding OR-GCDs. We then compare them to verify the correctness of the match by Eq. (8) and (9). This process is iterated a certain number of times, which is denoted as  $NR(Q, T)$ . Consequently,  $NR(Q, T)$  matches are randomly selected from the set  $S(Q, T)$  in total and they are verified.

Then, we use the ratio between the number of randomly selected matches judged to be correct, denoted as  $nr(Q, T)$ , and that of all the randomly selected matches, *i.e.*,  $NR(Q, T)$ , as the similarity between the two images. The range of the similarity is from 0 to 1. Similarly, we can determine test image  $T$  is a copy of query image  $Q$ , if

$$\frac{nr(Q, T)}{NR(Q, T)} \geq Rat_{TH} \quad (11)$$

where  $Rat_{TH}$  is a preset threshold that also ranges from 0 to 1.

##### B. Property Analysis of Fast Similarity Measurement

The property analysis of the fast similarity measurement is given as follows. By Eq. (8) and (9), we can determine whether two matched SIFT features is a correct match or not. For the SIFT matches between any two image copies, *i.e.*, one is a copy of the other, we suppose the probability that a match is judged to be correct is  $p_1$ . For the SIFT matches between any two non-copies, we suppose the probability that a match is judged to be correct is  $p_2$ . After verifying  $NR$  randomly selected matches between images, two conclusions can be obtained as follows.

If the two images are copies, the number of the randomly selected matches that are judged to be correct, denoted as  $nr_1$ , will satisfy a binomial distribution denoted as  $B_1(NR, p_1)$ . The probability density function (PDF) of  $B_1(NR, p_1)$  can be represented as

$$pdf_1(nr_1) = \binom{NR}{nr_1} p_1^{nr_1} (1 - p_1)^{NR - nr_1} \quad (12)$$

If the two images are non-copies of each other, the number of the randomly selected matches that judged to be correct, denoted as  $nr_2$ , will also satisfy a binomial distribution denoted as  $B_2(NR, p_2)$ . The PDF of  $B_2(NR, p_2)$  can be represented as

$$pdf_2(nr_2) = \binom{NR}{nr_2} p_2^{nr_2} (1 - p_2)^{NR - nr_2} \quad (13)$$

To obtain  $p_1$  and  $p_2$ , 7,500 pairs of image copies and 4M pairs of non-copies are chosen from the Holidays dataset [41], which is used in our experiments. Then, for each SIFT match, we extract and compare the corresponding OR-GCDs to verify its correctness, where the parameters—the number of histogram bins  $M \times N$ , the weighting factor  $\alpha$ , and the threshold  $Dis_{TH}$ —are set according to our experimental section. As a result,  $p_1$  and  $p_2$  are estimated to be 0.92 and 0.24, respectively.

According to the estimated  $p_1$  and  $p_2$ , if we use the proposed image similarity measurement strategy for copy detection, we can compute the true positive rate (TPR)  $P_T$  and false positive rate (FPR)  $P_F$  by

$$P_T = \sum_{NR \times Rat_{TH}}^{NR} \binom{NR}{nr_1} p_1^{nr_1} (1 - p_1)^{NR - nr_1} \quad (14)$$

$$P_F = \sum_{NR \times Rat_{TH}}^{NR} \binom{NR}{nr_2} p_2^{nr_2} (1 - p_2)^{NR - nr_2} \quad (15)$$

Table I lists the computed  $P_T$  and  $P_F$  values when we use different values of  $NR$  and  $Rat_{TH}$ . From Table I, we can see that when  $NR = 15$  and  $Rat_{TH} = 0.7$ , we can obtain satisfactory true positive and false positive rates, *i.e.*, 0.9950 and 0.0001. This demonstrates that our strategy can theoretically obtain a high level of accuracy by verifying only a small number of randomly selected SIFT matches for copy detection.

Thus, the proposed fast similarity measurement strategy based on random verification can significantly improve the efficiency, while maintaining the high accuracy levels for our copy detection method. That can be also illustrated by the experimental results in Section VI-C.

TABLE I  
THE  $P_T$  AND  $P_F$  VALUES FOR DIFFERENT  $NR$  AND  $Rat_{TH}$

$NR \backslash Rat_{TH}$	0.1	0.3	0.5	0.7
1	0.9200	0.2400	0.9200	0.2400
2	0.9936	0.4224	0.9936	0.4224
5	1.0000	0.7464	0.9998	0.3461
10	1.0000	0.9357	1.0000	0.4442
15	1.0000	0.9065	1.0000	0.2810

## V. THE EXTENDED METHOD FOR PARTIAL-DUPLICATE IMAGE DETECTION

Partial-duplicate images are the images sharing some duplicated local regions, which are cropped from the same original images with some copy attacks [16], [17]. Thus, these images can be viewed as special cases of image copies. Fig. 7 shows some examples of partial duplicates of a given original image. From this figure, we can see that the partial-duplicate images share some duplicated local regions but they have different global appearances. As a result, the corresponding OR-GCDs of matched local features between partial-duplicate images will be quite different. Thus, the proposed copy detection cannot be directly applied to partial-duplicate image detection.

To deal with the task of partial-duplicate image detection, we extend our method by adding a stage of potential duplicated region location after obtaining the initial SIFT matches between images. In this stage, a strict geometric verification (SGV) strategy is proposed to effectively filter the geometrically inconsistent matches between images, and then the coordinates of the remaining matches are used for potential duplicated region location. Finally, by cropping the located regions and treating them as full-size images, the proposed copy detection method is implemented to further confirm whether the located regions are duplicated regions of each other to obtain the partial-duplicate image detection result.

Most of the existing geometric verification strategies, such as RANSAC [13]–[15] and geometric coding [17], only utilize the relative spatial positions of local features to check the geometric consistency of local matches. Thus, they are not strict enough to filter geometrically inconsistent matches. As a result, if we use the coordinates of the remaining matches to locate duplicated regions, the location accuracy will be affected. To address this issue, we propose the SGV strategy, which not only considers the relative positions of local features but also their characteristic relationships to filter geometrically inconsistent matches for potential duplicated region location. The SGV strategy is detailed as follows.

As shown in Fig. 8(a), the coordinates, scale, and dominant orientation of SIFT feature  $f$  in a given image are denoted as  $(x, y)$ ,  $s$ , and  $\phi$ , respectively. By setting each matched SIFT feature as a reference feature  $f_r$ , we extract four geometric context features of  $f_r$  to describe the relative positions and characteristic relationships between the reference feature  $f_r$  and each of the other features  $f_i$ . These features are the relative position  $pr_i = \sqrt{(x_i - x_r)^2 + (y_i - y_r)^2}/s_r$  in radial

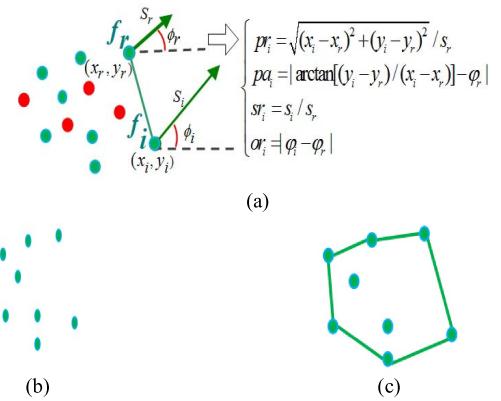


Fig. 8. The potential duplicated region location. (a) Four geometric context features of feature  $f_r$ , which describe the relative positions and characteristic relationships between a reference feature  $f_r$  and each of the other features  $f_i$  in a given image. (b) The remaining features after the strict geometric verification. (c) The located region constructed using the coordinates of these features.

direction, the relative position  $pa_i = |\arctan [(y_i - y_r)/(x_i - x_r)] - \phi_r|$  in azimuthal direction, the scale relationship  $sr_i = s_i / s_r$ , and the dominant orientation relationship  $or_i = |\phi_i - \phi_r|$ . Consequently, we obtain  $4 \times (N - 1)$  geometric context features of  $f_r$ . Where  $N$  represents the total number of all matches between the two images.

We denote the initially matched features of  $f_r$  and  $f_i$  in a database image as  $f'_r$  and  $f'_i$ , respectively. In the same manner, we also compute four geometric context features between the reference feature  $f'_r$  and each of other features  $f'_i$ . These features are denoted as  $pr'_i$ ,  $pa'_i$ ,  $sr'_i$ , and  $or'_i$ . As a result,  $4 \times (N - 1)$  geometric context features of  $f'_r$  are also generated.

Then, we measure inconsistency between the features  $f_r$  and  $f'_r$ , denoted as  $IC(f_r, f'_r)$ , by computing the normalized mean difference between all of their corresponding geometric context features according to Eq. (16).

$$IC(f_r, f'_r) = \frac{1}{4(N-1)} \sum_{i=1}^{N-1} \left( \frac{|pr_i - pr'_i|}{|pr_i + pr'_i|} + \frac{|pa_i - pa'_i|}{|pa_i + pa'_i|} \right. \\ \left. + \frac{|sr_i - sr'_i|}{|sr_i + sr'_i|} + \frac{|or_i - or'_i|}{|or_i + or'_i|} \right) \quad (16)$$

Note that, if all the matches are true positives,  $IC(f_r, f'_r)$  will be close to zero. If  $f_r$  and  $f'_r$  are a geometrically inconsistent match,  $IC(f_r, f'_r)$  will be a relatively large value. Similar to [17], we find the match with the highest inconsistency value, and remove it if its inconsistency value is larger than a preset threshold  $IC_{TH}$ . We iterate this process until the highest inconsistency value of the remaining matches is no larger than the threshold.

In most cases the meaningful objects are duplicated between images, and thus the duplicated regions are usually not with a regular shape. Thus, to be more close to the duplicated object shape, by the algorithm of convex hull computation [42], we construct a convex region using the coordinates of the remaining matches to obtain the potential duplicated regions, as shown in Fig. 8(c). When two images have multiple

duplicated regions, which usually suffer different scale and rotation transformations, we can locate the most dominant region.

It is worth noting that some similar objects, which are not partial duplicates, may have some geometric consistency matches, and they will be located as potential duplicated regions between images. Thus, we crop the located regions from images and extend them to rectangle regions through padding the black pixels whose intensities are equal to 0. Then, by treating them as full-size images, we adopt the proposed copy detection method to further confirm whether the located regions are duplicates of each other to obtain the final partial duplicate detection result.

## VI. EXPERIMENTS

In this section, first, the datasets and evaluation criteria adopted in our experiments are introduced. Second, the parameters—the number of histogram bins  $M \times N$ , the weighting factor  $\alpha$ , and the threshold  $Dist_{TH}$ —are experimentally determined. Third, the validity of the proposed fast image similarity measurement strategy is evaluated and its performance is compared to that of the traditional cardinality-based similarity measurement strategy. Finally, the performances of our method are tested and compared with those of state-of-the-art methods for the tasks of copy detection and partial duplicate detection.

### A. Data Set and Evaluation Criteria

In our experiments, we adopt four datasets: Holidays dataset [41], Copydays dataset [43], a challenging ground-truth dataset [44], and DupImage dataset [45].

1) *Holidays Dataset*: This dataset is originally composed of 1,491 personal holiday photos. In this dataset, there are 500 image groups, each of which represents a distinct scene or object. Each group consists of several similar images, which are obtained by different capture conditions, such as different acquisition time, viewpoints, and positions. To conduct experiments on this dataset, we choose the first image of each group as the query image and treat the rest similar images as non-copies. All the 500 query images are modified by Stirmark tool [46] using the 15 different copy attacks listed in Table II to generate 7,500 copies, which are inserted into the dataset. As a result, the dataset contains 8,491 test images, which include 7,500 copies and 991 similar images (non-copies).

2) *Copydays Dataset*: This dataset contains 3,212 images, including 157 original images and 3,055 copies generated by three kinds of copy attacks, including JPEG compression, cropping and “strong”. For each original image, it has 9 copies generated by JPEG compression with 9 different quality factors, 9 cropped versions generated by removing the image surface from 10% to 80%, and 2 to 6 copies generated by “strong” attacks, which includes a variety of manipulations such as scanning, blurring, rotating, and their combinations. We choose the 175 original images as query images for copy detection.

TABLE II  
THE 15 DIFFERENT IMAGE COPY ATTACKS

Rotation by 30 degrees
Rotation by 60 degrees
Scaling by the factor of 25%
Scaling by the factor of 50%
Cropping the image by 50%
Cropping the image by 75%
Rotation by 30 degrees + cropping the image by 50%
Rotation by 60 degrees + cropping the image by 50%
Changing intensity to 130%
Changing contrast to 130%
Gaussian blurring with the template [1 2 1; 2 4 2; 1 2 1]
Gaussian sharpening with the template [0 -1 0; -1 10 -1; 0 -1 0]
JPEG compression with the quality factor of 35%
Watermark embedding with the strength of 30
Noise addition with the factor of 15

3) *Challenging Ground-Truth Dataset*: We built a challenging ground-truth dataset to further test the copy detection performance of different methods. This dataset contains 10K images from 500 image groups. In each image group, there are 20 similar images, which are captured from the same scene or object but under slightly different capture conditions including acquisition time, viewpoints, and positions. The first image of each group is chosen as query image, and all the 500 query images are modified by the 15 different copy attacks listed in Table II to generate 7,500 copies, which are inserted into the dataset. Note that the challenging ground-truth dataset contains a lot of similar images, which are mixed together with the copies. These images are non-copies but are quite similar to the query images, which makes it challenging to successfully identify the copies from the dataset. The first image of each group is treated as query image.

4) *DupImage Dataset*: We use this dataset to test the performances of different methods for partial duplicate detection. This dataset consists of 1,104 partial-duplicate images downloaded from the networks, which are put into 33 groups. In each group, images are partial duplicates of each other. For partial-duplicate image detection, 100 representative images are randomly selected from this dataset as queries.

The last three datasets are used to test performances of different methods. However, their sizes are relatively small. Thus, we download 1M Internet images, which is used as distracting images, and add these images into each of these datasets. To make SIFT feature extraction and matching more efficient, all query images in the datasets are rescaled to no larger than 400×400. In addition, we find that the percentage of the image copies that have no fewer than 5 matched features is up to 98.13%. Thus, to enhance efficiency of the verification process, the candidate images that have fewer than 5 matched features will be skipped. We adopt Mean Average Precision (MAP), which represents the average precision across all different recall levels, to evaluate the detection performances of different methods.

### B. Parameter Determination

In our copy detection method, there are four important parameters:  $M$  and  $N$ , which control the number of the bins

TABLE III  
THE EFFECTS OF  $M$  AND  $N$  WHEN  $DisTH = 0.27$

$N \backslash M$	2	4	8
4	0.9397	0.9439	0.9429
8	0.9443	<b>0.9771</b>	0.9765
16	0.9764	0.9769	0.9761
32	0.9761	0.9755	0.9742

of the created histograms;  $\alpha$ , which is the weighting factor for the computation of distance between OR-GCDs; and  $DisTH$ , which represents the distance threshold for judging whether a SIFT match is correct or not. Note that, if the values of  $M$ ,  $N$ , and  $DisTH$  are too small or too large, it will cause an imbalance between the robustness and the discriminability for OR-GCD, and thus the performance of our method will be affected to some extent. Also, when the value of  $\alpha$  is equal to 1 or 0, only the intensity-based vector or the gradient-based vector in OR-GCD plays a role in copy detection, and thus we cannot obtain the optimal detection performance. Thus, we test several proper values for each of those parameters. More specifically, we test 3 values (2, 4, and 8) for  $M$ , 4 values (4, 8, 16, and 32) for  $N$ , 3 values (0.27, 0.3, and 0.32) for  $DisTH$ , and 5 values (0.1, 0.25, 0.5, 0.75, and 0.9) for  $\alpha$ . In this experiment, we set the size of visual codebook, *i.e.*, the number of visual words, as 50K to verify all the SIFT matches between images, and use the ratio of the matches judged to be correct as the image similarity for copy detection. We test the effects of these parameters on the MAP values of our method using the Holidays dataset.

However, if we use all of the combinations of these values to find the optimal parameter setting, it is required to repeat the experiment  $3 \times 4 \times 3 \times 5$  times, which is too time-consuming. According to the experimental setting of the SIFT algorithm [24], we assume  $\alpha$  is independent to the three other parameters. Therefore, we set  $\alpha$  to a default value, *i.e.*, 0.5, to test the other parameters and then fix the three other parameters to test  $\alpha$ .

The effects of  $M$  and  $N$  when using different values of  $DisTH$  are illustrated in Table III to Table V. From these tables, it is clear that larger  $M$  and  $N$  lead to better detection performance. That is mainly because the created histograms with a larger number of bins can encode more detailed global context information, which makes OR-GCD more discriminative. However, increasing  $M$  and  $N$  does not consistently improve performance. That might be because a larger number of histogram bins causes a smaller number of pixels in each bin, which will make OR-GCD more sensitive to nosing-like attacks such as Gaussian blurring, compression, and noise addition.

The effects of  $DisTH$  can be also observed in Table III to Table V. When  $DisTH$  is too small or too large, the detection performance degrades. The main reason is that when  $DisTH$  is too small, it causes many correct matches to be detected as false, and when  $DisTH$  is too large, it causes many false matches to be judged as correct. According

TABLE IV  
THE EFFECTS OF  $M$  AND  $N$  WHEN  $DisTH = 0.3$

$N \backslash M$	2	4	8
4	0.9428	0.9439	0.9449
8	0.9757	<b>0.9787</b>	0.9469
16	0.9458	0.9780	0.9761
32	0.9452	0.9768	0.9750

TABLE V  
THE EFFECTS OF  $M$  AND  $N$  WHEN  $DisTH = 0.32$

$N \backslash M$	2	4	8
4	0.9437	0.9446	0.9439
8	0.9445	0.9761	0.9765
16	0.9454	<b>0.9774</b>	0.9771
32	0.9458	0.9772	0.9770

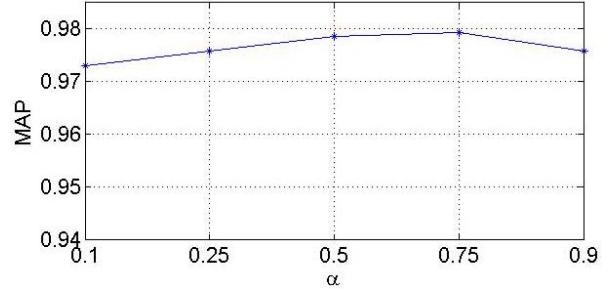


Fig. 9. The effects of  $\alpha$  when  $M = 4$ ,  $N = 8$  and  $DisTH = 0.3$ .

to these tables, when  $M$ ,  $N$ , and  $DisTH$  are equal to 4, 8, and 0.3, respectively, we can obtain desirable detection performance.

We test the effects of  $\alpha$  by fixing the three other parameters as the above values. Fig. 9 illustrates the effects of  $\alpha$ . According to Eq. (8), larger  $\alpha$  means the intensity-based vector of OR-GCD plays a greater role in copy detection, while smaller  $\alpha$  indicates the gradient-based vector of OR-GCD has more influence on copy detection. From Fig. 9, it can be observed that  $\alpha = 0.75$  provides the best performance, *i.e.*, 0.9793 MAP value. Therefore, we set  $\alpha$  as 0.75 and the three other parameters— $M$ ,  $N$ , and  $DisTH$ —as 4, 8, and 0.3, respectively, in the following experiments.

### C. Validity of Fast Image Similarity Measurement

After selecting the parameters, we test the validity of the fast image similarity measurement strategy based on random verification, and compare its performance with that of the traditional cardinality-based similarity measurement strategy.

To this end, we verify different numbers of randomly selected SIFT matches between images (5, 10, 15, and 20) by the proposed global context verification scheme, and then measure the image similarity using the ratio of the randomly selected matches judged to be correct. We denote

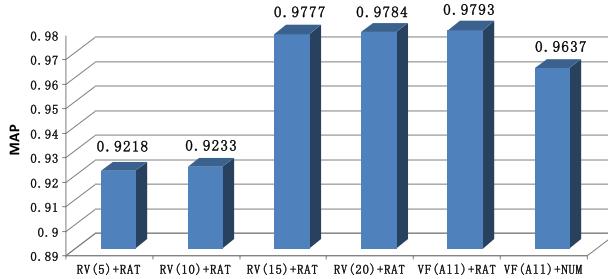


Fig. 10. The MAP values when verifying different numbers of matches.

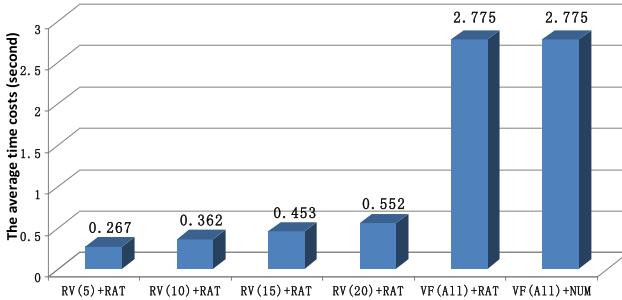


Fig. 11. The average detection time when verifying different numbers of matches.

the corresponding methods as RV(5)+RAT, RV(10)+RAT, RV(15)+RAT, and RV(20)+RAT. Also, we verify all the initial matches, and measure the image similarity using the ratio value or the number of matches judged to be correct. The two corresponding methods are denoted as VF(All)+RAT and VF(All)+NM. Here, VF(All)+NM can be also regarded as the method that uses the traditional cardinality-based strategy for copy detection. Then, we test and compare the performances of all of these methods for accuracy and efficiency. The experiments are also conducted on the Holidays dataset, and the size of visual codebook is set as 50K.

Fig. 10 shows the MAP values of these methods. As more SIFT matches are verified for image similarity measurement, the better detection accuracy is obtained. This indicates that verifying more initial matches is helpful for performance improvement. It is clear that RV(15)+RAT and RV(20)+RAT achieve comparable performances with VF(All)+RAT, which illustrates that it is unnecessary to verify all the matches between images; desirable performance can be obtained by verifying only a small number of matches. We also observe that VF(All)+RAT outperforms VF(All)+NM. The main reason is that using the ratio of the matches judged to be correct for similarity measurement can avoid the sensitive problem of the traditional cardinality-based strategy, since the ratio value is independent of image scene complexity.

The average detection time of different methods are shown in Fig. 11. It is clear that as more initial matches are verified for similarity measurement, the time cost per query image increases proportionally. However, the time costs of RV(5)+RAT, RV(10)+RAT, RV(15)+RAT, and RV(20)+RAT are significantly lower than those of VF(All)+RAT and VF(All)+NM. This is because the four methods only need to verify a small number of matches, far smaller than that of all

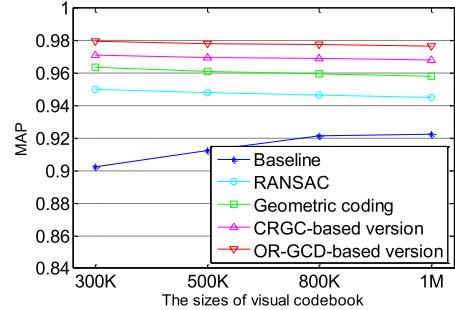


Fig. 12. The MAP values of different methods on the Copydays dataset.

the initial matches. As indicated in Fig. 10 and Fig. 11, when we use RV(15)+RAT for copy detection, a good trade-off between accuracy and efficiency can be obtained, *i.e.*, 0.9777 MAP value and 0.453 second per query on average. Therefore, the proposed similarity measurement strategy can significantly improve efficiency while maintaining high accuracy for copy detection. In the following experiments, we verify 15 randomly selected matches, and use the ratio of the matches judged to be correct as image similarity for copy detection.

#### D. Experiments on Copy Detection

In this subsection, we test our method on the Copydays dataset and our challenging ground-truth dataset, and make comparisons with the four other methods in the aspects of both accuracy and efficiency for copy detection. These methods are listed as follows.

(1) Baseline: The baseline is the method based on the BOW quantization. It obtains initial SIFT matches between images by the steps described in Section III-A, and then uses the number of these matches as image similarity for copy detection.

(2) RANSAC: After obtaining initial SIFT matches based on the BOW quantization, this method adopts a variant of RANSAC algorithm [14] as used in [15] to verify SIFT matches and to identify and remove false matches. The number of matches judged to be correct, *i.e.*, surviving matches, is used as image similarity for copy detection.

(3) Geometric coding: Different from RANSAC, this method adopts the geometric coding algorithm [17] for the verification of SIFT matches to remove the false matches that are geometrically inconsistent.

(4) CRGC-based version and (5) OR-GCD-based version of our method: To compare the performances of our method when using two different global context features, *i.e.*, the CRGC feature proposed in our previous work [18] and OR-GCD, we test two corresponding versions of our method, denoted as CRGC-based version and OR-GCD-based version for copy detection.

We adopt different sizes of visual codebook, *i.e.*, 300K, 500K, 800K and 1M to test the detection performances of those methods. The parameters of the first three methods and the CRGC feature are set according to the suggestion in the corresponding papers.

Fig. 12 and Fig. 13 show the MAP values of these methods on the two datasets, respectively. It is clear that all of these

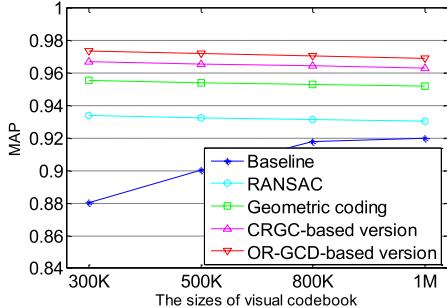


Fig. 13. The MAP values of different methods on the challenging ground-truth dataset.

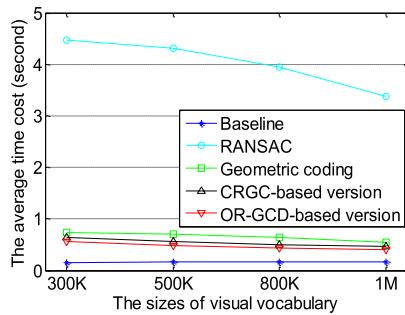


Fig. 14. The average detection time of different methods.

methods perform better on the Copydays dataset than on the challenging ground-truth dataset. That is mainly because the challenging ground-truth dataset contains many similar images, which are hard to be distinguished from image copies. From the two figures, we can also see that OR-GCD-based version achieves the highest accuracy among all of these methods. OR-GCD-based version outperforms RANSAC and Geometric coding, mainly because it explores the global context information for verification of SIFT matches, which is more effective for filtering false matches for copy detection than the geometric consistency information among SIFT matches. Also, OR-GCD-based version performs better than CRGC-based version. That is mainly because OR-GCD is more discriminative than CRGC feature, since OR-GCD encodes relatively rich spatial information. The baseline method performs worse than all of the other methods, because the baseline method does not make any extra efforts to reduce the false matches between images. In addition, when using a smaller visual codebook, the accuracy of the baseline method decreases. The main reason is fewer visual words lead to more quantization errors. However, the performances of the other methods are better when using a smaller visual codebook. That might be because the features extracted from image copies are more likely to be quantized to the same visual word. Although more false matches will occur at the same time, a considerable number of them can be identified and removed by the verification process.

The average time costs of these methods on the two datasets are shown in Fig. 14. It can be observed that the efficiency of OR-GCD-based version is comparable to that of the baseline method and higher than those of the four other methods. That is the case for the following reasons: 1) the extraction

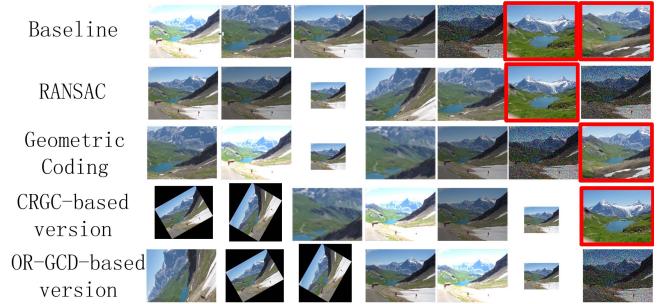


Fig. 15. Detection results of different methods for a same query. The images that are falsely detected as copies are highlighted by color boxes.

and comparison of OR-GCDs are quite efficient due to the good efficiency of OR-GCDs; 2) since only a small number of SIFT matches need to be verified, a small number of OR-GCDs corresponding to the matched features need to be extracted. Thus, OR-GCD-based version requires only a little extra computation cost for verification of SIFT matches. The average time cost of RANSAC is highest, because estimating an optimal affine transformation model for filtering false matches is a time-consuming process. In addition, when using larger visual codebook the average time cost of the baseline method increases slightly. That is because a larger number of visual words lead to more time cost in feature quantization. However, as the size of the visual codebook increases, the average time costs of the other methods decrease. The main reason is that fewer features are quantized to the same visual word, resulting in less time cost in the verification process.

Fig. 15 presents an example of the detection results of these methods on challenging ground-truth dataset. For the same query, the ninth to fifteenth images detected by different methods are shown in this figure. Although there are a lot of similar images in the dataset, our method, *i.e.*, OR-GCD-based version, successfully detects all the copies generated by a variety of possible copy attacks. The color boxes highlight the similar images that the other methods falsely detect as copies.

As indicated in Fig. 12 and Fig. 13, when the size of visual codebook is equal to 1M, the MAP values of OR-GCD-based version on the two datasets are 0.9767 and 0.9692, respectively, which are significantly higher than those of the other methods. From Fig. 14, the average time cost on the two datasets is only 0.4014 second, which is slightly higher than that of the baseline method. In conclusion, our method can achieve higher accuracy than all of the other methods with comparable efficiency to the baseline method.

### E. Experiments on Partial-Duplicate Image Detection

In our extended method, the potential duplicated region location is a key stage for partial-duplicate image detection. Thus, we first test the performance of the proposed location method, named as SVC-based location method, and compare with those of two other location methods. The two methods are denoted as RANSAC-based and Geometric coding-based location methods. They adopt RANSAC [14] and geometric

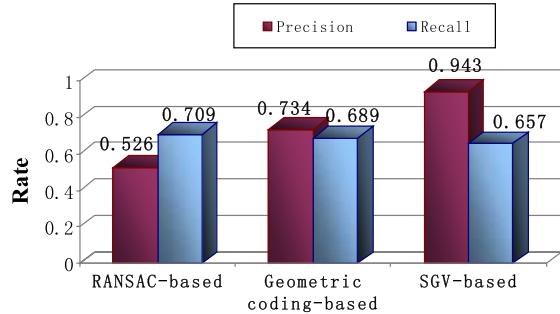


Fig. 16. The precision and recall rates of different location methods.

coding [17] strategies to filter geometrically inconsistent SIFT matches, respectively, and then both of them locate potential duplicated regions in the manner described in Section V. The parameters of RANSAC and geometric coding strategies used in the two location methods are set according to the suggestion in the corresponding papers. The size of visual codebook is set to 1M. Since at least three matches are required to locate the regions between images, the images that have less than three remaining matched features will be skipped.

We randomly choose 2,000 pairs of partial-duplicate images from the DupImage dataset to test the location performances of different methods. We adopt the recall rate and the precision rate to evaluate the location performances of different methods. Let  $R_l$  be the located region, and  $R_a$  actual duplicated region. At a certain threshold  $\tau$ , the precision rate and the recall rate are defined as follows:

$$\text{Precision}(\tau) = \frac{R_l \cap R_a}{R_l} \quad (17)$$

$$\text{Recall}(\tau) = \frac{R_l \cap R_a}{R_a} \quad (18)$$

A good region location method is expected to have high precision rate, so that the content of the corresponding located regions of a pair of partial-duplicate images can be almost consistent and thus the robustness of the OR-GCDs extracted from the regions will not be significantly affected. Meanwhile, it should not have low recall rate to ensure the located regions contain relatively rich spatial information, and thus the OR-GCDs extracted from the regions can maintain the good discriminability. By experiment, when the threshold  $IC_{TH}$  is set to 0.05, the precision rate and the recall rate of SVC-based location method can be up to 94.3% and 65.7%, respectively.

Fig. 16 shows the precision and recall rates of different location methods. From this figure, it is clear that the SVC-based location method can achieve higher precision rate and comparable recall rate compared with two other location methods. That is because the SVC-based location method considers both the relative spatial positions and the characteristic relationships of local features to filter geometrically inconsistent matches, and thus the accuracy of potential duplicated region location will be less affected.

For partial-duplicate image detection, we test the performances of six partial duplicate detection methods in the aspects of both accuracy and efficiency. These methods are

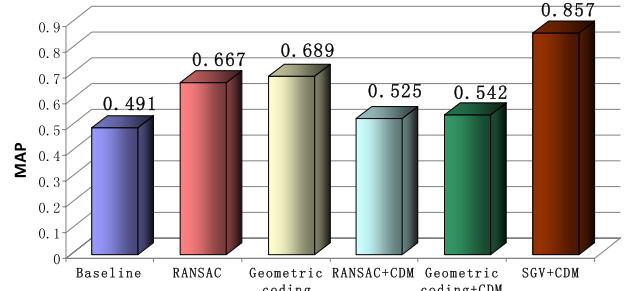


Fig. 17. The MAP values of different methods on the DupImage dataset.

denoted as: (1) Baseline, (2) RANSAC, (3) Geometric coding, (4) RANSAC+CDM, (5) Geometric coding+CDM, and (6) SGV+CDM. The first three methods are same to the corresponding methods described in Section VI-D. These methods can be directly applied to partial-duplicate image detection by using the number of the initial SIFT matches or that of the remaining SIFT matches after geometric verification. The next three methods are those methods, which locate the potential duplicated regions by adopting the RANSAC-based location method, the Geometric coding-based location method, and the proposed SGV-based location method, respectively. Then, all of the next three methods use the proposed copy detection method to confirm whether the located regions are duplicated regions or not to obtain partial duplicate detection result. The parameters of the first three methods and those of the three location methods and our copy detection method are set the same as mentioned above. The size of visual codebook is also set to 1M.

Fig. 17 shows MAP values of the six methods. From this figure, it is clear that our extended method, *i.e.*, SGV+CDM achieves highest MAP value, *i.e.*, 0.857. SGV+CDM outperforms both RANSAC+CDM and Geometric coding+CDM, mainly because the SGV-based location method can locate the duplicated regions more accurately than the two other location methods, as shown in Fig. 16. RANSAC+CDM and Geometric coding+CDM perform worse than RANSAC and Geometric coding, respectively, for the following reason. Although RANSAC+CDM and Geometric coding+CDM use both the geometric context information and global context information of matched local features for partial duplicate detection, due to the low precision rates of the RANSAC-based and Geometric coding-based location methods, the detection accuracies will be significantly affected.

Table VI shows the average time cost of each stage of the six methods and the total time costs per query of these methods. The stages in these methods include SIFT feature matching, geometric verification, potential duplicated region construction, and global context verification, which are denoted as FM, GV, PDRC, and GCV, respectively. It is clear that the total time cost per query of SGV+CDM, *i.e.*, our extended method, is 0.572 second, which is much lower than those of RANSAC+CDM and Geometric coding+CDM. The main reason is that RANSAC needs to estimate the affine transformation model beforehand and geometric coding contains the step of adjusting the coordinate of each matched feature.

TABLE VI  
THE AVERAGE TIME COSTS OF DIFFERENT METHODS (SECONDS)

Methods \ Stages	FM	GV	PDRC	GCV	Total
Baseline	0.158	-	-	-	0.158
RANSAC	0.158	6.475	-	-	6.633
Geometric coding	0.158	0.708	-	-	0.866
RANSAC+CDM	0.158	6.475	0.001	0.236	6.870
Geometric coding +CDM	0.158	0.708	0.001	0.236	1.103
SGV+CDM	0.158	<b>0.177</b>	0.001	0.236	<b>0.572</b>

Instead, SGV directly filter geometrically inconsistent matches without the two time-consuming steps. The total time costs per query of RANSAC+CDM and Geometric coding+CDM are higher than those of RANSAC and Geometric coding, respectively, and much higher than that of Baseline. That is because the stages of PDRC, GCV, and GV will lead to additional computation.

In conclusion, our extended method can achieve higher accuracy than all of the other methods, and slightly lower efficiency than the baseline method for partial-duplicate image detection.

## VII. CONCLUSION

In this paper, we present an effective and efficient global context verification scheme for image copy detection. Instead of resorting to geometric consistency, the global context information of SIFT features is explored for verification of SIFT matches to remove false matches to improve the detection performance. In addition, we extend the proposed copy detection method to deal with the task of partial-duplicate image detection.

Practically, there are many near-duplicate images distributed on the networks, in which the copies are usually mixed together with the similar images. These copies need to be distinguished from similar images for copyright protection purposes. The experimental results demonstrate that our method can achieve desirable performances in both accuracy and efficiency even when there are many similar images in the dataset. In addition, the extended version of the proposed method also achieves good performance for partial-duplicate image detection. Hence, we can conclude that our method has significance in the area of copyright protection, and is also quite useful in many other tasks, such as automatic annotating, content-based web links creation and redundancy elimination.

## REFERENCES

- [1] C. Kim, "Content-based image copy detection," *Signal Process. Image Commun.*, vol. 18, no. 3, pp. 169–184, Mar. 2003.
- [2] Z. Xia, X. Wang, X. Sun, and B. Wang, "Steganalysis of least significant bit matching using multi-order differences," *Secur. Commun. Netw.*, vol. 7, no. 8, pp. 1283–1291, Aug. 2014.
- [3] J. H. Hsiao, C. S. Chen, L. F. Chien, and M. S. Chen, "A new approach to image copy detection based on extended feature sets," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2069–2079, Aug. 2007.
- [4] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 507–518, Mar. 2015.
- [5] A. Joly, O. Buisson, and C. Frelicot, "Content-based copy retrieval using distortion-based probabilistic similarity search," *IEEE Trans. Multimedia*, vol. 9, no. 8, pp. 293–305, Feb. 2007.
- [6] J. Tang, H. Li, G. Qi, and T. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Trans. Multimedia*, vol. 12, no. 2, pp. 131–141, Feb. 2010.
- [7] W. Zhao, X. Wu, and C. Ngo, "On the annotation of Web videos by efficient near-duplicate search," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 448–461, Aug. 2010.
- [8] X. Wu, C.-W. Ngo, A. G. Hauptmann, and H.-K. Tan, "Real-time near-duplicate elimination for Web video search with content and context," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 196–207, Feb. 2009.
- [9] X.-Y. Wei, Z.-Q. Yang, C.-W. Ngo, and W. Zhang, "Visual typo correction by collocative optimization: A case study on merchandize images," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 527–540, Feb. 2014.
- [10] H. Ling, H. Cheng, Q. Ma, F. Zou, and W. Yan, "Efficient image copy detection using multiscale fingerprints," *IEEE Multimedia*, vol. 19, no. 1, pp. 60–69, Jan. 2012.
- [11] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.
- [12] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 304–317.
- [13] M. A. Fischle and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [14] O. Chum, J. Matas, and S. Obdrzalek, "Enhancing RANSAC by generalized model optimization," in *Proc. Asian. Conf. Comput. Vis.*, Jan. 2004, pp. 812–817.
- [15] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [16] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, "Spatial coding for large scale partial-duplicate Web image search," in *Proc. Int. Conf. Multimedia*, Oct. 2010, pp. 511–520.
- [17] W. Zhou, H. Li, Y. Lu, and Q. Tian, "SIFT match verification by geometric coding for large-scale partial-duplicate Web image search," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 1, Feb. 2013, Art. no. 4.
- [18] Z. Zhou, X. Sun, Y. Wang, Z. Fu, and Y. Shi, "Combination of SIFT feature and convex region-based global context feature for image copy detection," in *Proc. 14th Int. Workshop Digital-Forensics Watermarking*, 2015, pp. 60–71.
- [19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [20] K. Yan and L. Huston, "Efficient near-duplicate detection and sub-image retrieval," in *Proc. 12th ACM Int. Conf. Multimedia*, Jan. 2004, pp. 869–876.
- [21] L.-W. Kang, C.-Y. Hsu, H.-W. Chen, and C.-S. Lu, "Secure SIFT-based sparse representation for image copy detection and recognition," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2010, pp. 1248–1253.
- [22] Z. Xu, H. Ling, F. Zou, Z. Lu, and P. Li, "A novel image copy detection scheme based on the local multi-resolution histogram descriptor," *Multimedia Tools Appl.*, vol. 52, nos. 2–3, pp. 445–463, Apr. 2011.
- [23] H. Ling, L. Wang, F. Zou, and W. Yan, "Fine-search for image copy detection based on local affine-invariant descriptor and spatial dependent matching," *Multimedia Tools. Appl.*, vol. 52, nos. 2–3, pp. 551–568, Apr. 2011.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [25] K. Yan and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun./Jul. 2004, pp. II-506–II-513.
- [26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
- [27] H. Ling, L. Yan, F. Zou, C. Liu, and H. Feng, "Fast image copy detection approach based on local fingerprint defined visual words," *Signal Process.*, vol. 93, no. 8, pp. 2328–2338, Aug. 2013.

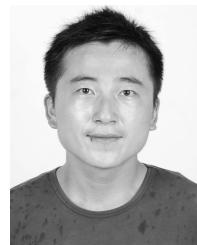
- [28] S. Wu and M. Lew, "Salient features for visual word based image copy detection," in *Proc. 4th ACM Int. Conf. Multimedia Retr.*, 2014, pp. 475–478.
- [29] L. Yan, F. Zou, R. Guo, L. Gao, K. Zhou, and C. Wang, "Feature aggregating hashing for image copy detection," *World Wide Web*, vol. 19, no. 2, pp. 217–229, May 2015.
- [30] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry-preserving visual phrases," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 809–816.
- [31] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, and T. Han, "Contextual weighting for vocabulary tree based image retrieval," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 209–216.
- [32] L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3368–3380, Aug. 2014.
- [33] J. Yao, B. Yang, and Q. Zhu, "Near-duplicate image retrieval based on contextual descriptor," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1404–1408, Sep. 2015.
- [34] S. Zhang, Q. Tian, K. Lu, Q. Huang, and W. Gao, "Edge-SIFT: Discriminative binary descriptor for scalable partial-duplicate mobile search," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2889–2902, Jul. 2013.
- [35] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
- [36] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1832–1837, Nov. 2005.
- [37] E. N. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 184–190.
- [38] J.-N. Liu and G.-H. Zeng, "Improved global context descriptor for describing interest regions," *J. Shanghai Jiaotong Univ.*, vol. 17, no. 2, pp. 147–152, Apr. 2012.
- [39] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2161–2168.
- [40] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, May 2009.
- [41] (2008). *Holidays*. [Online]. Available: <http://lear.inrialpes.fr/~jegou/data.php>
- [42] R. L. Graham, "An efficient algorithm for determining the convex hull of a finite planar set," *Inf. Process. Lett.*, vol. 1, no. 4, pp. 132–133, Jun. 1972.
- [43] (2008). *Copydays*. [Online]. Available: <http://lear.inrialpes.fr/~jegou/data.php>
- [44] (2015). *Challenging Ground-Truth*. [Online]. Available: <http://pan.baidu.com/s/1kUFVozX>
- [45] (2011). *DupImage*. [Online]. Available: <http://www.cs.utsa.edu/~wzhou/data/DupGroundTruthDataset.tgz>
- [46] F. A. P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 58–64, Sep. 2000.



**Zhili Zhou** received the B.S. degree in communication engineering from Hubei University in 2007, and the M.S. and Ph.D. degrees in computer application from the School of Information Science and Engineering, Hunan University, in 2010 and 2014, respectively.

He joined Nanjing University of Information Science and Technology, China, as an Assistant Professor in 2014. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Windsor, Canada.

His current research interests include near-duplicate image/video detection, image/video copy detection, coverless information hiding, digital forensics, and image processing.



**Yunlong Wang** received the B.S. degree in computer application from the Nanjing University of Information Science and Technology in 2014.

He is currently pursuing the M.S. degree in computer science. His current research interests include digital forensics, image/video copy detection, near-duplicate detection, and image retrieval.



**Q. M. Jonathan Wu** (M'92–SM'09) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990.

He was with the National Research Council of Canada for ten years from 1995 to 2004, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.



**Ching-Nung Yang** received the B.S. and M.S. degrees in telecommunication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1983 and 1985, respectively, and the Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 1997.

He is currently a Full Professor with the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien, Taiwan. His research interests include coding theory, information security, and cryptography.



**Xingming Sun** (SM'07) received the B.S. degree in mathematics from Hunan Normal University, China, in 1984, the M.S. degree in computing science from the Dalian University of Science and Technology, China, in 1988, and the Ph.D. degree in computer science from Fudan University, China, in 2001.

He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, China. In 2006, he visited University College London, U.K. He was a Visiting Professor with the University of Warwick, U.K., in 2008 and 2010. His research interests include network and information security, database security, and natural language processing.