Effective and Efficient Global Context Verification for Image Copy Detection

PROF. ANIL KULKARNI1, MD SHUJAATH KHAN[2], MD IMRAN AHMED[3], BHAGYASHREE BAPURE[4], ZAHOOR SIQQIQUA[5]
[1, 2, 3, 4, 5] *Guru Nanak Dev Engineering College, Bidar*

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 schemeto filter false matches for copy detection. More specifically, after obtaining initial scale invariant feature transform (SIFT) matchesbetween images based on the BOW quantization, the overlappingregion-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.

Indexed Terms- Image copy detection, near-duplicate detection, partial-duplicate detection, global context, overlapping region.

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 distributedon networks [1], [2]. To protect owners against unauthorized(re)use of their content, detecting illegal copies of digitalmultimedia is a basic requirement.

In recentyears, content-based image copy detection has been researchedas a passive technology to detect illegal copies. Differentfrom watermarking, which uses previously embedded marks, this technology extracts content-based features from imagesand 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 informationand copy detection can be implemented after image distrib-ution [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], 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.

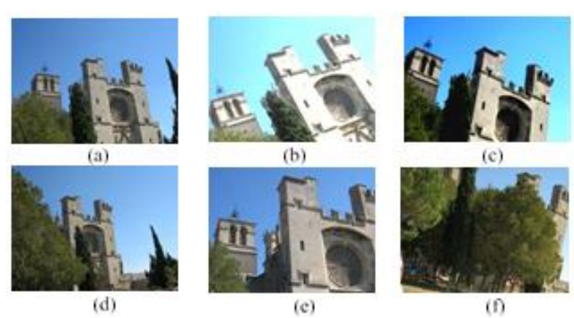


Fig. 1. The difference between image copies and similar images. (a) is an original image. (b) and (c)

are copies of the original image. (d)-(f) are similar images.

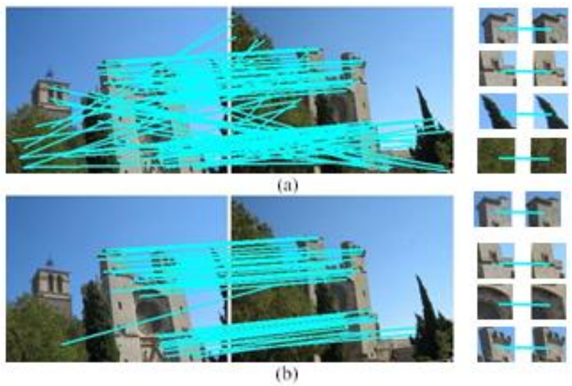


Fig. 2. The illustration of limited performances of BOW model and geometric verification. (a) The 83 false SIFT matches, which are generated by matching SIFT features based on the quantization of BOW model, where the size of visual codebook is set to 50K according to our experimental section.
(b) The 47 false SIFT matches, which still remain after removing most of the geometrically inconsistent matches by the algorithm with its default parameters. 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 thereare not only many copies but also a lot of similar images.

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-duplication detection methods.

In the literature, many copy detection methods based on local features have been proposed. To the best of our knowledge, these methods generally investigate the local features, such as SIFT and its extensions including principal component analysis on SIFT (PCA-SIFT), speeded-up robust feature (SURF), and multi-scale SIFT for detecting 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.

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. 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.

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. In Jegouetal. 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, 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 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, Zhouetal. 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 to 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.

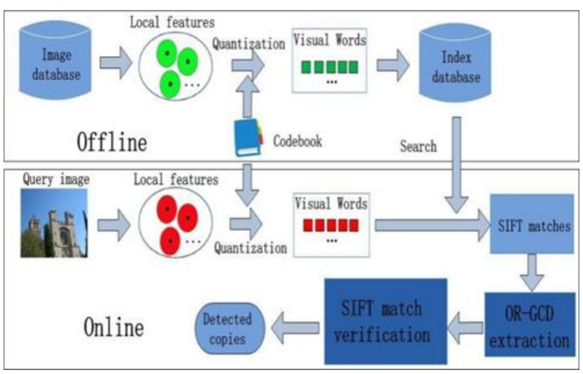


Fig. 4. The framework of our copy detection method.

Instead of resorting to the geometric consistency information, our motivation is to design an effective and efficient verification scheme by exploring the global context informationof matched local features. In our previous work, a convexregion-based global context (CRGC) feature is proposed for SIFT match verification. However, it is extracted by directly using the histogram of oriented gradients (HOG) , 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-GCDfor the verification of local matches to remove false matches.

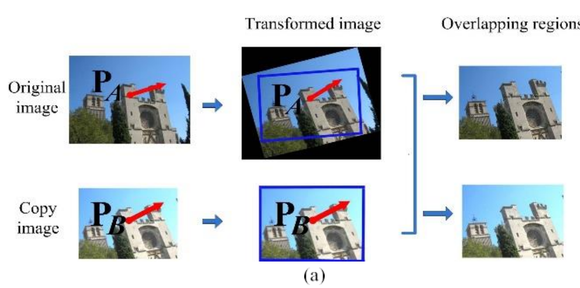
III. THE PROPOSED COPY DETECTION METHOD

The framework of our copy detection method is shownin 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 liein the later two components. In Section III-A, we match SIFT features between images based on the BOW quantization to obtain initial SIFT matches. 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 thenthe verification result can be further used to measure image similarities for copy detection.

1-SIFT Feature Matching

We extract hundreds of SIFT features from each image by using the SIFT algorithm. The extracted SIFT features arethen 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. Then, each extracted SIFT feature is quantized to its nearest



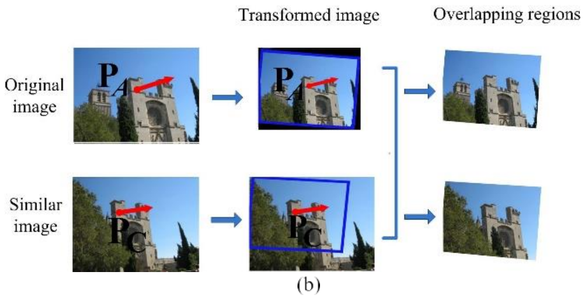


Fig. 5. The illustration of overlapping region construction. (a) The construction between an original image and its copy. (b) The construction between the two similar images.

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 φ, characteristic scale s, and coordinates (x, y). Note that this 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.

2: -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 ofSIFT matches might be a feasible way to identify and remove the false ones.

IV. LITERATURE SURVEY

Bharat M. Prajapati, Nirav P. Desai [8] - Digital images have been widely used from the last few years in various applications such as forensic evidence, medical, insurance and military etc. With easy availability of low - cost image modification and editing software such as Adobe Photoshop, GIMP (GNU Image Manipulation Program), paint etc. digital image content is not considered as safe. There are various types of image tampering techniques but Copy - move is the most used. In this technique, some part of the image is copied, then it is pasted on the same Image, which changes the visual contents of the image.

An efficient algorithm for image copy-move forgery detection based on DWT and SVD [2014] In this paper, an efficient algorithm is presented for image copy-move forgery detection and localization based on DWT and SVD. Experimental results demonstrate that our proposed algorithm can effectively detect multiple copy-move forgery and precisely locate the duplicated regions, even when an image was distorted by Gaussian blurring, JPEG compression and their mixed operations [1]

Extensive research has been done to device methods to detect copy-move forgery in both intensity domain and frequency domain. Various image analysis techniques using image moments, dimensionality reduction, texture analysis etc. Has been experimented. This paper presents a study of various image forgery techniques and a survey of various attempts in copy-move forgery detection. A comparative analysis of major techniques is also presented.[2]

V. SYSTEM ANALYSIS

For feature-level fusion, we concatenated the individual feature sets extracted from inertial sensor data and the corresponding RGB and/or depth video sequence after min-max normalization. Although feature-level fusion seems to be simple and straightforward, it suffers from some several deficiencies. First, the increase in the dimensionality of the fused feature vector raises the computational complexity of classification. Second, the dimensionality of the RGB/depth features is typically much higher than the features extracted for inertial sensor data, which ultimately degrades the fusion purpose. We address these issues using a variable length feature vector for RGB and depth data sequences. The size of the feature vector obtained from inertial sensor data is equal to 1×12. On the other hand, the length of each feature vector extracted for RGB/depth video sequence is equal to the number of clusters k in BoWs representation of dense HOG features.

CONCLUSION

Identifying whether or not an image is fake is essential. In this project, we'll look at a few methods for spotting fake images that are employed in forensics. A popular sort of forgery is copy-move forgery, in which the culprit moves objects from photographs to locations where they shouldn't be. In the project, copy move forgery is covered in more detail. We will examine numerous forensics techniques that can be used to identify copy-move forgery.

REFERENCES

[1] "Feng Liu, Hangzhou" and Hao Feng "An efficient algorithm for image copy-move forgery detection based on DWT and SVD" International Journal of Security and Its Applications Vol.8, No.5 (2014), pp.377-390 http://dx.doi.org/10.14257/ijsia.2014.8.5.33

[2] "Rani Susan Oommen Sree Buddha" "A Survey of CopyMove Forgery Detection Techniques for Digital Images" International journal of innovations in engineering and technology (IJIET).

[3] "Bharat M. Prajapati and Nirav P.Desai" "Forensic analysis of digital image tampering" International Journal For Technological Research In Engineering Volume 2, Issue 10, June-2015.

[4] "S. A. Fattah M.M.I. Ullah. M.Ahmed I. Ahmmed and C. Shahnaz" "A Scheme for Copy-Move Forgery Detection in Digital Images Based on 2D-DWT"

[5] "K. Kiruthika, S. Devi Mahalakshmi, K. Vijayalakshm" Detecting Multiple Copies of Copy-Move Forgery Based on SURF" ISSN (Online): 2319 -8753ISSN (Print): 2347 -6710 International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue ,3, March 2014

[6] "Rajdeep Kaur and Amandeep Kaur" A Review on Copy Move Forgery Techniques" International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 .6, No.2, Mar-April 2016.

[7] "Muhammad Hussain and Sahar Qasem " "Evaluation of Image Forgery Detection Using Multi-scale Weber Local Descriptors [2015]" International Journal of Artificial Intelligence Tools 24(4) · August 2015 with 24 Reads DOI: 10.1142/s0218213015400163

[8] "Hitesh Batra, Dr. Sanjay Badjate", A Review on Copy Move Forgery Techniques, International Journal of Advanced Research in Computer and Communication Engineering" Vol. 4, Issue 7, July 2015.