

DETECTING PHOTOGRAPHIC COMPOSITES USING TWO-VIEW GEOMETRICAL CONSTRAINTS

Wei Zhang, Xiaochun Cao, Zhiyong Feng, Jiawan Zhang

Ping Wang

School of Computer Science and Technology
Tianjin University, China
{wzhang, xcao, zyfeng, jwzhang}@tju.edu.cn

School of Science
Tianjin University, China
wang_ping@tju.edu.cn

ABSTRACT

In this work, we describe a new technique for detecting image composites by enforcing two-view geometrical constraints: \mathbf{H} and \mathbf{F} constraints on image pairs, where \mathbf{H} denotes the planar homography matrix and \mathbf{F} the fundamental matrix. Our approach detects fake regions efficiently on pictures taken at the same scene but with different camera configurations. Performance of this approach is demonstrated on real image pairs with visually plausible composites.

Index Terms— Digital Tampering, Planar Homography Constraint, Epipolar Constraint

1. INTRODUCTION

With the advancement of image and video editing tools, it is becoming easier and easier to make photorealistic composites into image sequences and videos, which makes it necessary to evaluate the authentication of images and videos. Digital watermarking [1, 2] has been proposed as a means to authenticate an image. However, a watermark must be inserted at the time of recording, which would limit this approach to specially equipped digital cameras.

Previous passive works mainly focus on detecting forgeries on single images using various features [3], including lighting conditions [4], camera parameters [5]. Although one can make visually plausible forgeries on a single image with care on various factors such as intricate boundaries and lighting conditions [6], it's more difficult to satisfy constraints between two images. Johnson and Farid [7] describe a technique to recover distortions by rectifying regions, which however can only deal with special known shapes of the region, such as polygons or circles. Wang and Farid [8] detect duplication in videos by evaluating the correlation coefficients of different regions, even for moving cameras, but can only deal with videos with small motion between frames. In addition, the performance of this method degenerates on shape/scale distortions caused by different camera viewpoints.

Real world would appear differently with various camera configurations, which makes it hard to evaluate their authenticity (Fig. 1). But appearances and positions of rigid objects



Fig. 1. It is difficult even for human to locate the photo composites in visually plausible image pairs.

in one image are related with another one monitoring the same scene. There are cases that two or more cameras are monitoring the same scene but with different configurations, e.g. the visual surveillance sites. When the two cameras are not co-centered, the two video feeds are related by a fundamental matrix. On the other hand, when the cameras are co-centered, e.g. a the leader and follower camera system, or large planes including walls and floors are monitored, the image pair is restricted by the stronger planar homography matrix. Pictures described above with composited regions could be difficult to be detected by human eyes, especially between pictures with significant distortions in shapes or scales (Fig. 1 (Row 1)). In this work, images are authenticated by enforcing the two-view geometrical constraints to detect photo composites.

The rest of this paper is organized as follows. In section 2, we review the two-view geometry briefly. Section 3 describes our method using two-view geometrical constraints to detect composites. Experimental results are shown in section 4, and section 5 concludes the paper.

2. GEOMETRICAL CONSTRAINTS

A real world camera can be modeled by a pinhole or perspective camera model. Theoretically, there exists the epipolar

geometry in any pair of images of the same scene, as long as the cameras' center are distinct. We call it the \mathbf{F} constraint, with \mathbf{F} denoting the fundamental matrix [9]. In the special cases, when cameras are co-centered or the scene is piecewise coplanar, there exists the planar homography between images of the same scene. We name it the \mathbf{H} constraint, with \mathbf{H} denoting the planar homography matrix.

2.1. Projecting 3D points onto 2D image

Under a pinhole camera model, mapping a 3D world point \mathbf{X} to the point \mathbf{x} on the 2D image plane is given by [9]:

$$\mathbf{x} = \mathbf{K}\mathbf{R}[\mathbf{I} | -\mathbf{C}]\mathbf{X}, \quad (1)$$

where \mathbf{K} is the camera intrinsic parameter matrix, \mathbf{R} is the rotation matrix, and \mathbf{C} denotes the location of camera.

2.2. H Constraint

When a camera rotates for an angle, corresponding points \mathbf{x}_1 and \mathbf{x}_2 on two image planes are related by:

$$\mathbf{x}_2 = \mathbf{K}[\mathbf{R}|0]\mathbf{X} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{x}_1, \quad (2)$$

while in case of pure zooming, corresponding points hold:

$$\mathbf{x}_2 = \mathbf{K}'[\mathbf{I}|0]\mathbf{X} = \mathbf{K}'\mathbf{K}^{-1}\mathbf{x}_1, \quad (3)$$

where \mathbf{K} and \mathbf{K}' are the two internal parameter matrices. When points are locally coplanar, correspondences can also be related by a planar homography matrix, regardless the motion of camera.

Therefore, we can draw the conclusion that pictures taken before and after camera motion are constrained by a planar homography, \mathbf{H} , if any of the assumptions holds: (1) camera does not change its position; (2) scenes viewed are locally coplanar. The planar homography matrix: \mathbf{H} , can be estimated with corresponding points of the two images, which will be discussed below.

2.3. F Constraint

When a camera moves generally and points are not coplanar, they still can be related with a Fundamental Matrix, \mathbf{F} , which maps a point \mathbf{x}_1 on one image to a line l_1 on the other image:

$$\mathbf{x}_2^T l_1 = \mathbf{x}_2^T \mathbf{F} \mathbf{x}_1 = 0, \quad (4)$$

where \mathbf{x}_2 is the corresponding point of \mathbf{x}_1 .

3. METHODS

In this section, we leverage the two-view geometrical constraints discussed in section 2.2 and 2.3 to highlight photo-realistic composites. In this work, we use SIFT [10] to find initial matches for the estimation of \mathbf{H} and \mathbf{F} .

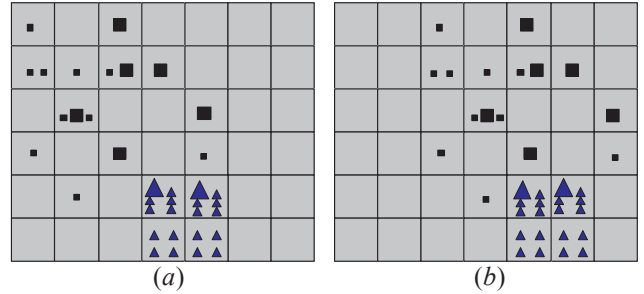


Fig. 2. Bucketing technique alleviates pollution from fake regions with dense feature points. (a, b) A pair of images with corresponding feature points. Black squares are feature points from the original images, while blue triangles denote points from fake regions. Selected points are in bigger size.

3.1. Detecting Composites Using H Constraint

3.1.1. Purification

Theoretically, \mathbf{H} can be estimated precisely with un-tampered images and correct matching points. However, when two pictures with complex fake regions are to be examined, SIFT may pay great attention on the fake area (blue triangles in Fig. 2). To handle this “pollution”, as well as wrong matches found by SIFT, we combine bucketing technique [11] and RANSAC for purification.

Bucketing Technique is to make global spatial optimization. As shown in Fig. 2, images are divided into $M \times N$ buckets, and corresponding points are located in each bucket. Buckets with at least one corresponding point are indexed for later selection. When it needs a more pair of corresponding points, a bucket is selected randomly, and then this bucket votes a point randomly as a candidate to calculate \mathbf{H} . Selected buckets can not be selected again during one calculation of \mathbf{H} . After bucketing, the hit ratio of real points has been increased sharply, as shown in Fig.2. After this, RANSAC is used to remove wrong matches during the estimation of \mathbf{H} . In this work, \mathbf{H} is estimated using “Golden Standard Algorithm”[9].

3.1.2. Locating Fake Regions

After the calculation of \mathbf{H} , we can recover the rectified image as \mathbf{I}_1' from the original image \mathbf{I}_1 . Since the two images (\mathbf{I}_1' and \mathbf{I}_2) should be similar after this rectification, we simply minus them to produce the difference matrix: \mathbf{D} .

Regions with high difference \mathbf{D} are selected to produce the binary map, which highlights fake regions. The threshold of cutting the difference is given by:

$$t = \max(\mathbf{D}) - c \quad (5)$$

where \mathbf{D} denotes the difference of frame \mathbf{I}_2 and \mathbf{I}_1' at every pixel, and the constant value c locates in $[0.3, 0.6]$ through our experiments.

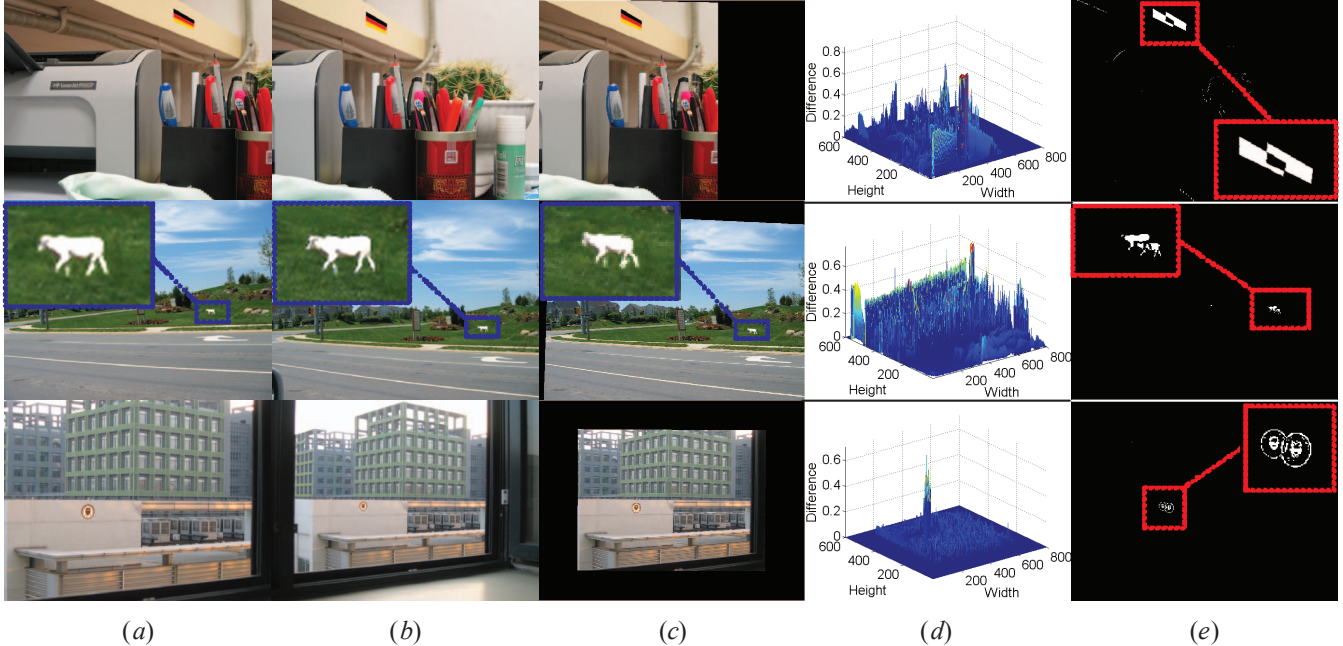


Fig. 3. Detecting composites by enforcing the \mathbf{H} constraint. (a, b) Original image pairs. (c) Images rectified from (a) using the estimated \mathbf{H} . (d) Difference maps between (b) and (c) based on correlation. (e) Binary masks with fake regions in white. Rows 1 and 2 are images taken with rotation, and row 3 is a zooming case.

3.2. Detecting Composites using \mathbf{F} Constraint

The estimation of fundamental matrix also involves the same “pollution” problem occurred in \mathbf{H} estimation, and the same bucketing technique in Section. 3.1.1 is utilized which will not be discussed again. Theoretically, 7 pairs of corresponding points are sufficient to estimate \mathbf{F} , since \mathbf{F} has the DOF of 7. However, at least 8 pairs of points are required to achieve a linear estimation. In this work, the “Gold Standard” algorithm [9] is applied to estimate \mathbf{F} . Unfortunately, we can not get a rectified image from one of the frames as Section 3.1.2 does, since there is no one-to-one mapping function before and after general camera motion. Instead, point \mathbf{x}_1 from frame \mathbf{I}_1 maps to an epipolar line on frame \mathbf{I}_2 . Distance between \mathbf{x}_2 and the epipolar line $l = \mathbf{F}\mathbf{x}_1$ is used as the metric,

$$d(\mathbf{x}_2, \mathbf{F}\mathbf{x}_1) = \sqrt{(\mathbf{x}_2^T \mathbf{F}\mathbf{x}_1)^2 / ((\mathbf{F}\mathbf{x}_1)_x^2 + (\mathbf{F}\mathbf{x}_1)_y^2)}. \quad (6)$$

The distance measurements in Eq. (6) provide a candidate set, $\Psi = \{(\mathbf{x}_{1i}, \mathbf{x}_{2i}) | d(\mathbf{x}_{2i}, \mathbf{F}\mathbf{x}_{1i}) > t\}$, of features inside the potential fake regions. Our method takes the advantage of the property that unmatched features inside fake regions are dense and close to each other, while wrong SIFT matches are random and sparse. Therefore, we make the assumption that clustered feature points in Ψ indicate fake regions. As a result, we dilate the points in Ψ with morphological operation to highlight a region including dense fake points.

4. EXPERIMENTAL RESULTS

In this section, various visually plausible pictures with composites are used to test our proposed methods.

4.1. \mathbf{H} Constraint Cases

In Fig. 3 (Rows 1, 2), both indoor and outdoor scenes are taken before and after the camera rotation. The faked flag and cow are detected as forgeries in Fig. 3 (e). Row 3 shows another case with pure zooming, and our method also managed to locate the fake region. Notice that the detections are slightly easier for larger regions (e.g. Fig. 3 (Row 1)), particularly at high SNR values, but that even very small regions are still detectable, as shown in Fig. 3 (Rows 2, 3).

Fig. 4 demonstrates an example with Bucketing technique. In this case, bucket size is configured as 50×50 (in pixels) on a 800×600 image. After Bucketing Technique, 67 (Fig. 4 (b)) out of 467 (Fig. 4 (a)) pairs are voted, most of which are from “real region”, thus hit ratio of real points is increased, and the impact of dense distribution of feature points in fake regions is also alleviated, as shown in Fig. 4 (d).

4.2. \mathbf{F} Constraint Cases

Fig. 5 (a) shows two image pairs under the general camera motion. Notice that Fig. 5 (b) only plots points and their epipolar lines which don’t satisfy the \mathbf{F} constraint. Using

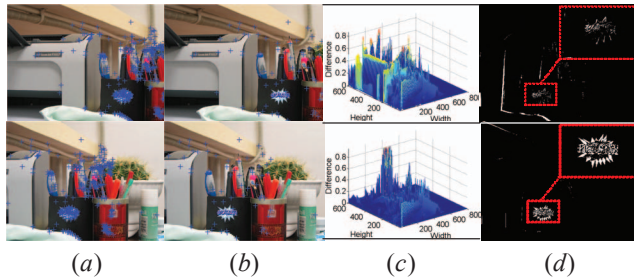


Fig. 4. Performance improvement by Bucketing Technique. (a) A natural image pair with original SIFT matches superimposed in blue crosses. (b) Refined matches using Bucketing Technique. (c) Difference images before (upper) and after (lower) bucketing techniques. (d) Detected fake regions before (upper) and after (lower) bucketing.

the method described in Section 2.3, the fake region is highlighted in Fig. 5 (c). Zoomed in versions of detected fake regions are also shown, where the estimated fake pixels are emphasized by setting the green channel values to zero.

5. CONCLUSION

We leveraged the two-view geometrical constraints from computer vision for the purpose of photographic composite detection. Both constraints, the stronger \mathbf{H} and the weaker \mathbf{F} , have their own advantages. The \mathbf{H} constraint has the favorable one-to-one mapping property with more assumptions, and the \mathbf{F} constraint can be used widely in general cases although it only provides weaker relationship between frames. Future research directions include extending this method to non-grid objects or dynamic scenes.

6. ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundations of China (No. 50735003 and No.60673196), Tianjin University 985 research fund, Natural Science Foundation of Tianjin (No. 07F2030), and State Key Laboratory of Precision Measuring Technology and Instruments open fund.

7. REFERENCES

- [1] I.J. Cox, M.L. Miller, and J.A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2002.
- [2] H. Liu, J. Rao, and X. Yao, "Feature based watermarking scheme for image authentication," in *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, 2008, pp. 229–232.
- [3] T. V. Lanh, K. Chong, S. Emmanuel, and M. S. Kankanhalli, "A survey on digital camera image forensic methods," *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 16–19.

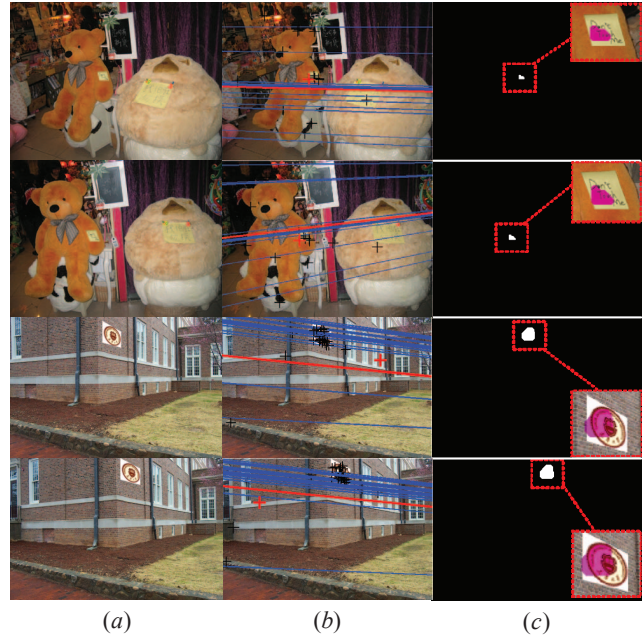


Fig. 5. Detecting composites by enforcing the \mathbf{F} constraint. (a) Two natural image pairs with visually plausible fake regions. (b) Points (black crosses) with large distance to their corresponding epipolar lines (blue lines). Some examples are highlighted with larger size in red. (c) Binary map highlighting the fake regions. The zoomed in versions of detected fake regions are also shown, where the estimated fake pixels are emphasized by setting the green channel values to zero.

- [4] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," *ACM Proceedings of the 7th Workshop on Multimedia and Security*, 2005, pp. 1–9.
- [5] W. Wang and H. Farid, "Detecting re-projected video," *International Workshop on Information Hiding*, 2008.
- [6] X. Cao, Y. Shen, M. Shah, and H. Foroosh, "Single view compositing with shadows," *The Visual Computer*, 2005, vol. 21, pp. 639–648.
- [7] M. K. Johnson and H. Farid, "Metric measurements on a plane from a single image," *Technical Report TR2006-579*, Department of Computer Science, Dartmouth College, 2006.
- [8] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," *ACM Multimedia and Security Workshop*, 2007.
- [9] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," Cambridge University Press, 2004.
- [10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 2004, pp. 91–110.
- [11] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, 1998, vol. 27, pp. 161–198.