DETECTING PHOTOGRAPHIC COMPOSITES USING SHADOWS

Wei Zhang, Xiaochun Cao, Jiawan Zhang

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

Jigui Zhu, Ping Wang

Tianjin University, China
{jiguizhu, wang_ping}@tju.edu.cn

ABSTRACT

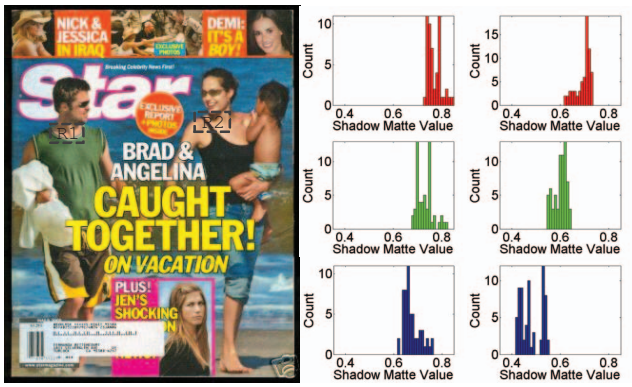
Image compositing technology has become popular for tampering with digital photographies. We describe how such composites can be detected by enforcing the geometric and photometric constraints from shadows. In particular, we explore (i) the imaged shadow relations that are modeled by the planar homology, and (ii) the color characteristics of the shadows measured by the shadow matte. Our approach efficiently extracts these constraints from a single image and makes use of them for the digital forgery detection. Experimental results on visually plausible images demonstrate the performance of the proposed method.

Index Terms— Digital Forensics, Planar Homology, Shadow Matte

1. INTRODUCTION

Matting and compositing are important operations in the production of special effects [1]. Since then, image region copy and paste has become one of the most common video editing and manipulating techniques due to its simplicity. Trivial tricks, including frame and region duplication across frames, can be detected by correlation based algorithms [2, 3], even when the camera moves [4]. However, the above methods assume the composited regions or frames are from known images or videos. They will fail in the case that the source and target images are different. For example, in Fig. 1 (*a*), Brad was photographed on a Caribbean island in 2005 while Angelina was taken in Virginia several years earlier.

Geometry and photometry based methods are also used as evidence of tampering. The former algorithms typically enforce the consistencies in the camera internal parameters including skew [5] and principal point [6]. The latter approaches [7, 8] use inconsistencies in the lighting for detecting forgeries. While a person's eyes are common in the images and can be used for the estimation of camera internal parameters [6] or lighting inconsistencies [8] when the eyes are sizable, they are not always available or detectable in the real images, e,g. people with sunglasses as shown in Fig. 1. Theoretically, the lighting direction inconsistency in Fig. 1 can be estimated using [7]. However, we attack this problem



(*a*) (*b*) (*c*)

Fig. 1. It is difficult for the current Image matting and compositing technology to satisfy both geometric and photometric constraints. (*a*) Star cover of Pitt and Jolie. (*b*) and (*c*) Shadow matte values in RGB channels (top to bottom) of the two regions marked in black dashed rectangles in (*a*). By the analysis of shadow matte values of the shadow regions as detailed in text, the tampering is obvious.

in both photometric and geometric points of view, and the geometric constraint used is also different from [7].

Shadows provide important visual cues for depth, shape, contact, and lighting in our perception of the world [9]. In this paper, we introduce a new framework to detect composites using geometric and photometric constraints from shadows. We show how to extract both geometric and photometric constraints from a single view of a target scene, and to make use of them for the digital forgery detection. Methods based on *shadow geometry* and *shadow photometry* are described in sections 2 and 3. Then we demonstrate the performance of our methods in section 4, and section 5 concludes this paper.

2. METHOD BASED ON SHADOW GEOMETRY

In this work, we utilize the planar homology [10] that encompasses the imaged shadow relationship as shown in Fig. 2 to detect photo composites. Note that the light source is not necessarily to be at infinity to keep the model by a planar ho-

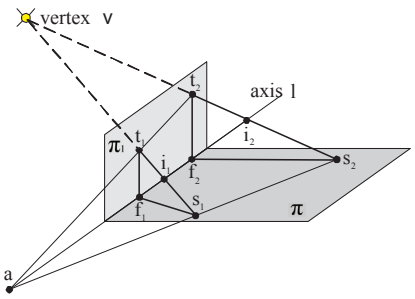


Fig. 2. Geometry of a planar homology. A plane, $\boldsymbol{\pi}_1$, and its shadow, illuminated by a point light source $\mathbf{v}$ and cast on a ground plane $\boldsymbol{\pi}$, are related by a planar homology.

mology, provided that the light source is a point light source, i.e. all light rays are concurrent.

As shown in Fig. 2, a planar homology is a planar projective transformation $\mathbf{H}$ which has a line $\mathbf{l}$ of fixed points, called the *axis*, and a distinct fixed point $\mathbf{v}$, not on the axis $\mathbf{l}$, called the *vertex* of the homology,

$$\mathbf{H} = \mathbf{I} + (\mu - 1)\frac{\mathbf{v}\mathbf{l}^T}{\mathbf{v}^T\mathbf{l}}, \tag{1}$$

where μ is the cross ratio that will be discussed later. In our case, the vertex $\mathbf{v}$ is the image of the light source, and the axis, $\mathbf{l}$, is the image of the intersection between planes $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}$. Each point off the axis, e.g. $\mathbf{t}_2$, lies on a fixed line $\mathbf{t}_2\mathbf{s}_2$ through $\mathbf{v}$ intersecting the axis at $\mathbf{i}_2$ and is mapped to another point $\mathbf{s}_2$ on the line. Note that $\mathbf{i}_2$ is the intersection in the image plane, although the light ray $\mathbf{t}_2\mathbf{s}_2$ and the axis, $\mathbf{l}$, are unlikely to intersect in 3D real world.

One important property of a planar homology is that the corresponding lines intersect with the axis, e.g. the lines $\mathbf{t}_1\mathbf{t}_2$ and $\mathbf{s}_1\mathbf{s}_2$ intersect at $\mathbf{a}$ on $\mathbf{l}$. Another important property of a planar homology is that the cross ratio, μ, defined by the vertex, $\mathbf{v}$, the corresponding points, $\mathbf{t}_i$ and $\mathbf{s}_i$, and the intersection point, $\mathbf{i}_i$, is the characteristic invariance of the homology, and thus is the same for all corresponding points. For example, the cross ratios $\{\mathbf{v}, \mathbf{t}_1; \mathbf{s}_1, \mathbf{i}_1\}$ and $\{\mathbf{v}, \mathbf{t}_2; \mathbf{s}_2, \mathbf{i}_2\}$ are equal. The two constraints can be expressed as

$$((\mathbf{t}_2 \times \mathbf{t}_1) \times (\mathbf{s}_2 \times \mathbf{s}_1)) \cdot (\mathbf{f}_2 \times \mathbf{f}_1) = 0, \tag{2}$$

$$\{\mathbf{v}, \mathbf{t}_1; \mathbf{s}_1, \mathbf{i}_1\} = \{\mathbf{v}, \mathbf{t}_2; \mathbf{s}_2, \mathbf{i}_2\}, \tag{3}$$

where

$$\mathbf{v} = (\mathbf{t}_2 \times \mathbf{s}_2) \times (\mathbf{t}_1 \times \mathbf{s}_1). \tag{4}$$

Therefore, we use this two constraints to detect composities in a nature image. Notice that $\mathbf{t}_1, \mathbf{f}_1, \mathbf{t}_2, \mathbf{f}_2$ have to be coplanar and $\mathbf{f}_1, \mathbf{f}_2$ on the intersection of plane $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}$. In real world, vertical objects standing on the ground satisfy this assumption, such as standing people, street lamps, trees and buildings. In addition, people usually are interested in inserting a new actor, which is mostly standing and vertical, into some target scene.

3. METHOD BASED ON SHADOW PHOTOMETRY

Besides the geometric constraints, the color characteristics of the shadows are also powerful cues for detecting composites from one image. In this work, we adopt shading image values [11] as the color characteristics. The shading image (or illumination image), $\mathbf{S}(x, y)$, together with the reflectance image, $\mathbf{R}(x, y)$, are called the intrinsic images. Generally, the observed image, $\mathbf{I}(x, y)$, can be modelled as:

$$\mathbf{I}(x, y) = \mathbf{S}(x, y)\mathbf{R}(x, y). \tag{5}$$

Therefore, the problem of detecting image composites reduces to recovering the shading image, $\mathbf{S}(x, y)$, from the input image $\mathbf{I}(x, y)$.

Many approaches [12, 13] have been proposed to derive illumination image and reflectance image from images or videos. Theoretically, those decomposed light maps could be used as shadow mattes in our work. However, the strategy [12] requires the knowledge about the structure of the surface in the target scene and how it appears when illuminated, which is not practical for our case since we are provided only a single view of the target scene, which is inaccessible. On the other hand, the method [13] is not able to handle the compression effects such as JPEG effects. Nevertheless, our challenge is to detect composited objects from given images or video frames, which are typically compressed.

Our approach takes the advantage of the property that changes in color between pixels indicate either reflectance changes or shading effects [12]. In other words, it is unlikely that significant shading boundaries and reflectance edges occur at the same point. Therefore, we make the assumption that every color change along the shadow boundaries, the edges caused by illumination difference only, is caused by shading only, i.e. the reflectance image colors across the shading boundaries should be the same or similar. In practice, considering the gradual change along the normal direction of the shadow boundaries, due to either compression effects or soft shadows, the input image pixel value, $\hat{\mathbf{I}}(x, y)$, and the reflectance image pixel value, $\hat{\mathbf{R}}(x, y)$, of each boundary pixel, (x, y), are estimated as:

$$\hat{\mathbf{I}}(x, y) = median\{\mathbf{I}(m, n) : (m, n) \in \mathcal{N}_i\} \tag{6}$$

$$\hat{\mathbf{R}}(x, y) = median\{\mathbf{I}(m, n) : (m, n) \in \mathcal{N}_o\} \tag{7}$$

where $\mathcal{N}_i$ and $\mathcal{N}_o$ are subsets of the set of neighbor pixels of (x, y) inside a grid with size $M \times M$, and the subscripts denote whether the pixels are inside ($\mathcal{N}_i$) or outside ($\mathcal{N}_o$) of the shadows. From Eq. (6, 7), we can compute the shading image value as, $\hat{\mathbf{S}}(x, y) = \hat{\mathbf{I}}(x, y)/\hat{\mathbf{R}}(x, y)$, for each pixel (x, y) along the shadow boundaries. For example, the histogram of estimated $\hat{\mathbf{S}}(x, y)$ along the boundaries of the shadow patch (Fig. 3 (*b*)) is plotted in Fig. 3 (c). Size of the grid ($M \times M$) should be dependent on the size of the shadow area. As M grows, the median shadow matte values are becoming more

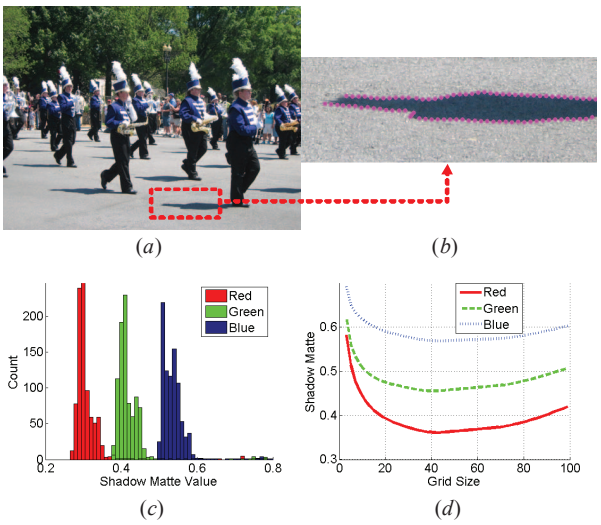


Fig. 3. Shadow matte demo. (*a*) A natural image. (*b*) The shadow patch (528×188 pixels) of the red dashed rectangular region in (*a*) with "Canny" edges in magenta dotted lines. (*c*) The histogram of shadow matte values of (*b*) in red, green and blue channels. (*d*) The curve of median shadow matte value with respect to different grid size as detailed in text.

and more stable, but finally, too large M exceeds the scope of shadow, which pollutes the shadow matte value estimation (Fig. 3 (*d*)). Generally, high quality images with large shadow areas are preferred to make precise estimation. However, the difference of shadow matte values of authentic and composited shadow areas are still obvious for various M in our experiments. We in this work use M=25 for all experiments, including Fig. 3 (*c*).

After the statistics on shadow matte values of shadow patches, a Bhattacharyya coefficient based similarity comparison is used to identify fake regions.

$$\mathbf{B}(\mathbf{H}_\alpha, \mathbf{H}_\beta) = \sum_{m=1}^{N} \sqrt{h_{\alpha,m} h_{\beta,m}}, \tag{8}$$

where $\mathbf{H}_x = (h_{x,1}, h_{x,2}, \ldots h_{x,N})$ is the histogram of shadow matte values along the edges of shadow area x. In our implementation, N is 100. Shadow matte is assumed approximately constant in one picture with similar lighting condition, thus, the region with low similarity coefficient is faked. Provided that the ground surface are partially under shadow, which are mostly true in real world, our experiments show that this approximation works well.

4. RESULTS

4.1. Composite Detection Based on Shadow Geometry

The estimation of the planar homology constraints requires three pairs of points. These six points are manually selected as shown in Fig. 4 in black squares, crosses and circles. First, we check the constraint in Eq. (2). The two corresponding lines of the two shadows in Fig. 4 (Row 2) don't intersect on the axis, while those in Row 3 do. Take Fig. 4 (left) for example, the first constraint shows that region 1 is not consistent with region 2, but regions 2 and 3 are. We therefore suspect region 1 to be faked. Table 1 shows the cross ratios between region pairs in Fig. 4. Without surprise, the cross ratios that involve fake regions 1 and 6 are different (around 21% on average), while the cross ratios involve all authentic regions (regions 2∼5) are within the relative difference of 1%. Note that cross ratios of the same shadow are different in different shadow pairs, as different pair defines a different plane $\boldsymbol{\pi}_1$ in Fig. 2, and thus results in a different location of $\mathbf{i}$.

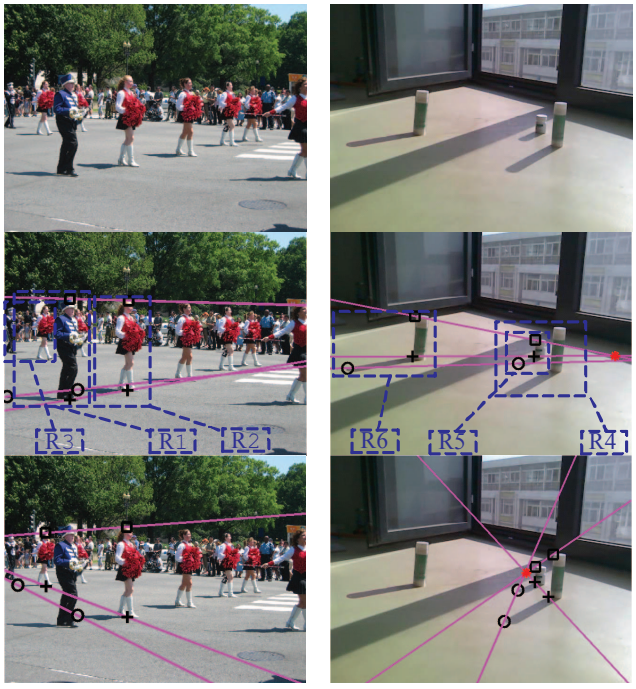


Fig. 4. Composite detection based on shadow geometry. **Top row**: Two nature images with composited regions. **Middle row**: The corresponding lines that involve composited regions (R1 and R6) don't intersect on the axis, i.e. disobeying the constraint in Eq. (2). In addition, they dissatisfy the characteristic invariance constraint in Eq. (3) (see Table 1). **Bottom row**: The imaged shadow relationship of authentic objects can be modeled by a planar homology. Black squares, crosses and circles denote the locations of $\mathbf{t}$, $\mathbf{f}$ and $\mathbf{s}$ in Fig. 2 respectively.

4.2. Composite Detection Based on Shadow Photometry

Shadow photometry constraints are especially useful in cases where it is not easy to estimate the planar homology (e.g. Fig. 1), and where the planar homology constraints are ap-

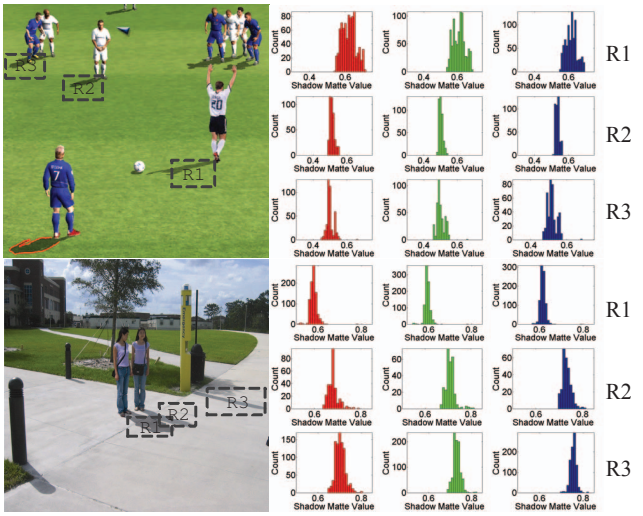


Fig. 5. Composite detection based on shadow photometry. (a) Two natural images with composited regions. $(b)\sim(d)$ Shadow matte values in RGB channels of the different regions marked in black dashed rectangles in (a). By the analysis of shadow matte values of the shadow regions, the regions labeled as $R1$ are clearly faked.

Table 1. Cross ratios of planar homologies in Fig. 4.

Region A	Region B	μ_A	μ_B	Diff Ratio
R1	R2	0.1741	0.1231	29.2589%
R2	R3	0.1587	0.1573	0.8794%
R1	R3	0.4454	0.4966	11.5145%
R4	R5	0.6298	0.6352	0.8647%
R4	R6	0.4473	0.3384	24.3526%
R5	R6	0.3237	0.2625	18.9191%

Table 2. Bhattacharyya coefficients for shadow regions.

Figure	B(R1,R2)	B(R1,R3)	B(R2,R3)
Fig. 1	0.0143	N/A	N/A
Fig. 5 Upper	0.0440	0.0823	0.7555
Fig. 5 Bottom	0.0035	0.0023	0.7688

proximately satisfied, e.g. the two images in Fig. 5. The power of constraints from shadow photometry for the detection of composites is demonstrated in Fig. 5. Obviously, the region labeled as $R1$ is faked by a glance at the histograms of shadow matte plots in Fig. 5 (right). These observations are also consistent with the Bhattacharyya coefficient based similarity measurements as shown in Table 2.

5. CONCLUSION

This paper presents a new framework for detecting image composites based on estimated shadow geometry and photometry. This method is especially useful when the scene is a wide area and the calibration objects such as a person's eyes are not measurable. The experimental results demonstrate that this method is efficient and can be applied to a variety of target scenes. As a pragmatic and flexible framework, it is also simple and easy to implement.

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] A. R. Smith and J. F. Blinn, "Blue screen matting," Proc. ACM SIGGRAPH, 1996, pp. 259–268.

[2] J. Fridrich, D. Soukal, and J. Lukas, "Detection of copy-move forgery in digital images," Proceedings of Digital Forensic Research Workshop, 2003.

[3] T.V. Lanh, K. Chong, S. Emmanuel, and M.S. Kankanhalli, "A survey on digital camera image forensic methods," Proceedings of the ICME, 2007, pp. 16–19.

[4] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," Proc. ACM MM&SEC, 2007.

[5] W. Wang and H. Farid, "Detecting re-projected video," International Workshop on Information Hiding, 2008.

[6] M.K. Johnson and H. Farid, "Detecting photographic composites of people," Proc. IWDW, 2007.

[7] M.K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," ACM Multimedia and Security Workshop, New York, NY, 2005, pp. 1–9.

[8] M.K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," 9th International Workshop on Information Hiding, Saint Malo, France, 2007.

[9] D. Kersten, P. Mamassian, and D. C. Knill, "Moving cast shadows induce apparent motion in depth," Perception, 1997, vol. 26, pp. 171–192.

[10] C. E. Springer, "Geometry and analysis of projective spaces," Freeman, 1964.

[11] H.G. Barrow and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images," Academic Press, 1978.

[12] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," Advances in Neural Information Processing Systems, 2003.

[13] G. Finlayson, M. Drew, and C. Lu, "Intrinsic images by entropy minimization," Proc. ECCV, 2004, pp. 582–595.