

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220465288>

# Detecting image forgeries using metrology

Article in *Machine Vision and Applications* · March 2012

DOI: 10.1007/s00138-010-0296-6 · Source: DBLP

CITATIONS

3

READS

41

## 4 authors:



**Lin Wu**

University of Queensland

41 PUBLICATIONS 264 CITATIONS

[SEE PROFILE](#)



**Xiaochun Cao**

Chinese Academy of Sciences

154 PUBLICATIONS 1,279 CITATIONS

[SEE PROFILE](#)



**Wei Zhang**

The University of Manchester

142 PUBLICATIONS 2,231 CITATIONS

[SEE PROFILE](#)



**Yang Wang**

UNSW Sydney

33 PUBLICATIONS 213 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Object detection & tracking [View project](#)



structure optimisation [View project](#)

All content following this page was uploaded by [Yang Wang](#) on 14 January 2015.

The user has requested enhancement of the downloaded file.

# Detecting image forgeries using metrology

Lin Wu · Xiaochun Cao · Wei Zhang · Yang Wang

Received: 22 February 2010 / Revised: 15 August 2010 / Accepted: 17 August 2010 / Published online: 16 September 2010  
© Springer-Verlag 2010

**Abstract** Image forgery technology has become popular for tampering with digital photography. This paper presents a framework for detecting fake regions using single view metrology and enforcing geometric constraints from shadows. In particular, we describe how to (1) estimate the region of interest's 3D measurements from a single perspective view of a scene given only minimal geometric information determined from the image, (2) determine the fake region by exploring the imaged shadow relations that are modeled by the planar homology. We also show that image forgery on the vertical plane or arbitrary plane can be detected through the measurement on such plane. Our approach efficiently extracts geometric constraints from a single image and makes use of them for the digital forgery detection. Experimental results on both the synthetic data against noise and visually plausible images demonstrate the performance of the proposed method.

**Keywords** Digital forensics · Single view metrology · Planar homology

## 1 Introduction

Doctored photographs are appearing with a growing frequency and sophistication in tabloid magazines, mainstream media outlets, political campaigns, evidence in a courtroom, and cases involving scientific fraud [8]. With the rapid advancement in image editing softwares, photorealistic images will become increasingly easier to be generated [24]. Therefore, authenticating the integrity of digital image's content has become particularly important when images are used as critical evidence in journalism and security surveillance applications. Over the past several years, the field of digital forensics has emerged to authenticate digital images by enforcing several authentication methods.

Digital watermarking [6] has been proposed as a means to authenticate an image. However, a watermarking must be inserted at the time of recording, which would limit this approach to specially equipped digital cameras having no capabilities of adding a watermarking at the time of image capture.

Blind approaches for image authentication can be roughly grouped into five categories [8]. The first category is pixel-based technique that analyzes pixel-level correlations arising from tampering. Duplicated images detection has been elaborated by several pixel-based techniques [10, 25, 29]. In [25], the authors used the method based on approximate block matching to detect the copy–move forgery even when the copied area is retouched to merge it with the background and when the forged image is saved in a lossy format, such as JPEG. Other efficient algorithms based on pixels have also been proposed in [10, 29], in which copy–move forgery region can be detected using blur moment invariants and principal component analysis, respectively. The method aiming for detecting re-sampled images is described in [30], which requires re-sampling the original image onto a new

---

L. Wu · X. Cao (✉) · W. Zhang · Y. Wang  
School of Computer Science and Technology,  
Tianjin University, Tianjin, China  
e-mail: xcao@tju.edu.cn

L. Wu  
e-mail: wulin@tju.edu.cn; jolinwulin1986@gmail.com

W. Zhang  
e-mail: wzhang@tju.edu.cn

Y. Wang  
e-mail: wayag2000@yahoo.com.cn

sampling image, introducing specific periodic correlations between neighboring pixels. These correlations can be used to detect this specific manipulation because they are unlikely to occur naturally [30]. When the images are spliced, several methods [5, 26, 27] are efficient to detect the tampered region. In [26], the authors proposed a method based on 2D phase congruency and statistical moments of characteristic function. While in [5, 27], the tampered region is evaluated by the analysis of physical difference between images in terms of image generative process and higher order statistics, respectively. In addition, statistical properties in natural images are also utilized in [1]. As the authors in [1], the measurements between the original and processed images can be viewed as features in classifier design.

The second is format-based technique that detects tampering in lossy image compression: unique properties of lossy compression such as JPEG can be exploited for forensic analysis [22, 28]. The third is camera-based technique that exploits artifacts introduced by the camera lens, sensor or on-chip post-processing [33]. Models of chromatic aberration [18], color filter array [2], camera response [14, 21] and sensor noise [4, 12, 23] are estimated to infer the source digital cameras and reveal digitally altered images. Other works such as [33] that revealed altered images using the intrinsic traces in the entire in-camera and post-camera processing operations. The fourth is physically based technique that models and detects anomalies using physical rules. For example, three dimensional interaction between physical objects, light, and the camera can be used as the evidence of tampering [16, 17]. The fifth is geometric-based technique that makes use of geometric constraints that are recovered from perspective views [15, 19] to authenticate the integrity of an image.

The proposed approach falls into the fifth category, i.e. geometric techniques, since we detect forged images by enforcing geometric constraints. Several geometric-based techniques [15, 19, 35, 37] have been proposed for image forgery detection. The estimation of internal camera parameters including principal point [15] and skew [35] can be used as the evidence of tampering. In [15], the authors showed how translation in the image plane is equivalent to a shift of the principal point and differences in which can therefore be used as evidence of forgery. Wang and Farid [35] argued that the skew of the re-projected video is inconsistent with the expected parameter of an authentic video. However, there are also some drawbacks that it only applies to frames that contain a planar surface. Zhang et al. [37] described a technique for detecting image composites by enforcing two-view geometrical constraints. The approach can detect fake regions efficiently on pictures at the same scene but requires two images correlated with  $\mathbf{H}$  (planar homography) or  $\mathbf{F}$  (fundamental matrix) constraints.

In this work, we leverage single view geometric constraints to detect composited regions and enforce imaged shadow relations [13] when shadow composites are tampered in the image. The input of the proposed method is a single image to be authenticated, and our method is operated on verifying the region of interest's (ROI's) integrity. We have implemented the following respects: (1) the anomaly investigation on 3D object's height based on single view metrology technique; (2) fake region detection by exploring the imaged shadow relations; (3) metric measurements on vertical or arbitrary planes with respect to a reference plane.

This paper is organized as follows. After reviewing the related works in Sect. 2, the proposed method is described in Sect. 3. Then we demonstrate the performance of our method in Sect. 4 to show that the proposed method can detect image forgeries based on anomaly investigation. The final discussions on the future work as well as the collaboration with other perspective methods are drawn in Sect. 5.

## 2 Related work and contributions

Metric measurements can be made from a planar surface after rectifying the image. In [19], the authors reviewed three techniques for the rectification of planar surfaces under perspective projection. They argued that knowledge of polygons of known shape, two or more vanishing points, and two or more coplanar circles can be used to recover the image to world transformation, thereby allowing metric measurements to be achieved on the plane. Each method in [19] requires only one single image but fails in measurements for objects out of the reference plane. Instead, we aim to retrieve the region of interest's (ROI) 3D measurements, even when the region is out of the reference plane.

Wang et al. [34] showed how to use the camera matrix and some available scene constraints to retrieve geometrical entities of the scene. The entities include height of an object on the reference plane, and measurements on a vertical or arbitrary plane with respect to the reference plane. However, they need to compute the camera projection matrix through the homography of the reference plane and its vertical vanishing point. Whereas, the vertical vanishing point is not required in our implementation and we employ minimal geometric constraints to get the planar homography. The looser requirement on inputs allows the proposed method to have wider applicability. In addition, we are focusing on the new application: the nonintrusive digital image forensic.

The single view metrology using geometric constraints has been addressed in [7]. The authors demonstrated that the affine 3D geometry of a scene may be measured from a single perspective image using the vanishing line of a reference plane and the vertical vanishing point. However, they are mainly concerned with measurements parallel planes and measurements on these planes.

In this paper, our novel contributions on image forgery authentication by enforcing geometric constraints are in the following three respects: (1) We apply new metrology methods to get 3D measurements on a vertical plane and an arbitrary plane rather than just metric measurements on a reference plane; (2) we need less scene information, e.g. no requirement of the vertical vanishing point, to obtain 3D measurements compared with the existing algorithms; (3) we also consider two geometric constraints of the shadow relationship for indicating a tampered image.

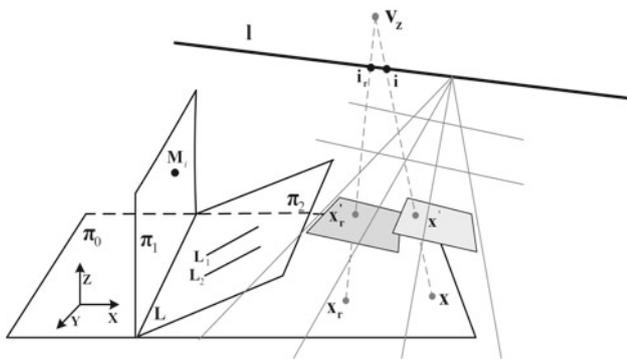
### 3 Metric measurements

In this section, we present the methods that recover 3D metric measurements on vertical and arbitrary planes. Then, we introduce the hypothesis testing to determine the ROI's integrity according to the estimated metric measurements. Finally, we model the shadow relationship in image forgeries utilizing the planar homology.

#### 3.1 Height measurement

We aim to measure the object's 3D height which can be treated as the distance between two parallel planes. The distance between scene planes is specified by a base point on the reference plane and top point in the scene [7]. The image containing such planes is illustrated in Fig. 1.

Suppose the base and the top points can be specified as  $\mathbf{X} = [X, Y, 0]^T$  and  $\mathbf{X}' = [X, Y, Z]^T$ , respectively, and their images are  $\mathbf{x}$  and  $\mathbf{x}'$ . Then, it is ready to show the metric value  $Z$  for target can be obtained by the following equation



**Fig. 1** Height measurement, measurement on vertical plane  $\pi_1$  and arbitrary plane  $\pi_2$ . Here  $\pi_0$  is the reference plane. The point  $\mathbf{x}_r$  and  $\mathbf{x}$  on the reference plane  $\pi_0$  correspond to the point  $\mathbf{x}'_r$  and  $\mathbf{x}'$ .  $\mathbf{l}$  is the horizon line, the vertical vanishing point  $\mathbf{v}_z$ ,  $\mathbf{i}_r$ ,  $\mathbf{x}'_r$  and  $\mathbf{x}_r$  can define a cross ratio:  $\{\mathbf{v}_z, \mathbf{i}_r; \mathbf{x}'_r, \mathbf{x}_r\}$ . The same situation holds true for  $\{\mathbf{v}_z, \mathbf{i}; \mathbf{x}', \mathbf{x}\}$ . The three planes intersect on line  $\mathbf{L}$ , and  $\mathbf{L}_1, \mathbf{L}_2$  are parallel lines on  $\pi_2$

in algebraic representation:

$$Z = \frac{\|\mathbf{x} \times \mathbf{x}'\|}{(\bar{\mathbf{l}} \cdot \mathbf{x}) \|\mathbf{v}_z \times \mathbf{x}'\|} Z_r \frac{(\bar{\mathbf{l}} \cdot \mathbf{x}_r) \|\mathbf{v}_z \times \mathbf{x}'_r\|}{\|\mathbf{x}_r \times \mathbf{x}'_r\|}, \tag{1}$$

where  $Z_r$  is the referred object's metric value,  $\mathbf{x}_r$  and  $\mathbf{x}'_r$  are images of referred object's base and top points,  $\bar{\mathbf{l}} = \mathbf{l} / \|\mathbf{l}\|$  and  $\mathbf{l}$  is the horizon. Let  $\mathbf{l} \sim [l_x, l_y, 1]^T$  and  $\mathbf{v}_z \sim [v_x, v_y, 1]^T$ , the vertical vanishing point  $\mathbf{v}_z$  can be automatically computed as follows:

$$\begin{pmatrix} l_x \\ l_y \\ 1 \end{pmatrix} \sim \begin{pmatrix} 1 & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ 1 \end{pmatrix} \tag{2}$$

where  $\omega_{ij}$  denotes the element in  $i$ th row and  $j$ th column of  $\omega$ . Here,  $\omega = (\mathbf{K}\mathbf{K}^T)^{-1}$  is the image of absolute conic (IAC) that may be derived from the first two columns of planar homography  $\mathbf{H}$  using the following relationship [13]:

$$\mathbf{h}_i^T \omega \mathbf{h}_i = 0 \tag{3}$$

where  $\mathbf{h}_i$  is the  $i$ th column of  $\mathbf{H}$ . Here,  $\mathbf{K}$  is a non-singular  $3 \times 3$  upper triangular matrix known as the camera calibration matrix [13].

To determine  $\mathbf{H}$ , the transformation from 2D points in image to world coordinates up to similarity's ambiguity, we use the following relationship:

$$\mathbf{H} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \tag{4}$$

where  $\mathbf{l} = (l_1 \ l_2 \ l_3)^T$  is the vanishing line of the reference plane, the coefficients  $\alpha$  and  $\beta$  can be estimated from known angles, equal but unknown angles or length ratios [20].

In another example,  $\omega$  may be directly retrieved from the orthogonal relationship provided by two vanishing points that are perpendicular to each other if they are available from some scene.

$$\mathbf{v}_x^T \omega \mathbf{v}_y = 0 \tag{5}$$

where  $\mathbf{v}_x$  and  $\mathbf{v}_y$  are two vanishing points. Then,  $\mathbf{l} = \mathbf{v}_x \times \mathbf{v}_y$ , and  $\mathbf{H}$  can also be computed from such geometric constraints.

Consequently, our method aiming for single view metrology only requires two sets of parallel lines to obtain the horizon line and different combinations of geometric constraints to achieve metric rectification. These information is generally obtainable from images of structured scenes, as shown in Figs. 4, 5, 6.

#### 3.2 Measurement on vertical and arbitrary planes

In this section, we show that scene measurements on vertical and arbitrary planes can also be retrieved from the camera projection matrix and scene constraints.

Suppose  $\pi_0$  is the reference plane,  $\pi_1$  is the vertical plane perpendicular to  $\pi_0$  and intersects  $\pi_0$  at line  $\mathbf{L}$ , seeing Fig. 1. We denote  $\pi_1$  as  $\Pi_1 = [a_1, b_1, c_1, d_1]^T$ , and the point  $\mathbf{M}_i = [x_i, y_i, z_i]^T$  is in  $\pi_1$  if and only if  $\Pi_1^T \mathbf{M}_i = 0$ . Let  $\mathbf{l}$  be the corresponding image of  $\mathbf{L}$ , and  $\mathbf{H}$  be the planar homography, then  $a_1 = (\mathbf{H}^T \mathbf{l})_1, b_1 = (\mathbf{H}^T \mathbf{l})_2, c_1 = 0, d_1 = (\mathbf{H}^T \mathbf{l})_3$  [34]. For a point  $\mathbf{M}_i$  on the vertical plane  $\pi_1$ , its coordinates can be retrieved by the intersection of the back-projected ray of image point  $\mathbf{m}_i = [m_x, m_y, 1]^T$  and the plane  $\pi_1$  using the following linear equation:

$$\begin{pmatrix} p_{11} - p_{31}m_x & p_{12} - p_{32}m_x & p_{13} - p_{33}m_x \\ p_{21} - p_{31}m_y & p_{22} - p_{32}m_y & p_{23} - p_{33}m_y \\ a_1 & quadb_1 & c_1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} p_{34}m_x - p_{14} \\ p_{34}m_y - p_{24} \\ -d_1 \end{pmatrix}, \tag{6}$$

where  $p_{ij}$  is the element in  $i$ th row and  $j$ th column of camera matrix  $\mathbf{P}$  [13].

General methods [7,34] tried to compute the camera matrix  $\mathbf{P}$  induced by rigid constraints to achieve metric rectification. However, we use minimal geometric constraints to obtain the planar homography  $\mathbf{H}$ , as described in the Sect. 3.1 (Eq. 4 for details). The camera matrix  $\mathbf{P}$  can be retrieved from  $\mathbf{H}$  up to 3 degrees of freedom’s ambiguities since  $\mathbf{P}$  is 11 dof while  $\mathbf{H}$  is 8. Typically, the skew  $\gamma$  of the commercial-off-the-shelf (COTS) camera is zero, which provides one constraint. The remaining two ambiguities are relieved by the availability of vertical vanishing point which, however, needs restricted scene to provide the vertical direction. In our work, the principal point is used since it can provide two independent constraints on  $\mathbf{P}$ , and it is known to be approximately at the center of a natural image [3,15]. Therefore, in our implementation, the principal point is assumed to be the center of the image and  $\mathbf{P}$  is directly determined from  $\mathbf{H}$ . After  $\omega$  is determined from Eq. (3),  $\mathbf{K}$  may be computed by decomposing the IAC. As a result, the camera matrix  $\mathbf{P}$  is computed from  $\mathbf{H}$ , since  $\mathbf{h}_1 = \mathbf{K}\mathbf{r}_1, \mathbf{h}_2 = \mathbf{K}\mathbf{r}_2,$  and  $\mathbf{p}_3 = \mathbf{K}\mathbf{r}_3 = \mathbf{K}(\mathbf{r}_1 \times \mathbf{r}_2)$ .

For the measurement on an arbitrary plane, we assume that an arbitrary plane  $\pi_2$  intersect the reference plane at line  $\mathbf{L}$ . Then all the plane passing through  $\mathbf{L}$  can form a pencil which may be expressed as

$$\begin{cases} \Pi_0 = [0, 0, 1, 0]^T \\ \Pi_1 = [(\mathbf{H}^T \mathbf{l})_1, (\mathbf{H}^T \mathbf{l})_2, 0, (\mathbf{H}^T \mathbf{l})_3]^T \\ \Pi_2 = \Pi_1 + \lambda \Pi_0 = [(\mathbf{H}^T \mathbf{l})_1, (\mathbf{H}^T \mathbf{l})_2, \lambda, (\mathbf{H}^T \mathbf{l})_3]^T \end{cases} \tag{7}$$

where  $\Pi_0$  is the reference plane and  $\Pi_1$  is the vertical plane. Therefore, the arbitrary plane is defined up to a unknown parameter  $\lambda$  which can be determined from a pair of imaged parallel lines in the plane [34]. Once the plane  $\pi_2$  is determined, we can take measurements on the plane in a similar

way like the vertical plane, using the following equations.

$$\begin{cases} s_j \mathbf{n}_j = \mathbf{P}\mathbf{N}_j \\ \Pi_2^T \mathbf{N}_j = 0 \end{cases} \tag{8}$$

where  $s_j$  is a scalar factor,  $\mathbf{P}$  is the camera matrix,  $\mathbf{N}_j$  is a 3D point in  $\pi_2$  and  $\mathbf{n}_j$  is its corresponding image point.

### 3.3 Hypothesis testing

We model the metric value  $Z$  for 3D metric measurement to obey the ordinary Gaussian distribution with the density function  $Z \sim N(\mu, \sigma^2)$ :

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

and the cumulative distribution function  $G(x)$ , where  $\mu$  and  $\sigma$  are mean value and standard deviation, respectively. For example, in our implementation on person height measurement for fake region detection, we set it to be  $\mu = 180$  cm,  $\sigma = 15$  cm, and thereby, the probability is

$$G\{\mu - 2\sigma < Z < \mu + 2\sigma\} = 95.44\% \tag{10}$$

Using this model, the probability that an ordinary Gaussian random variable with the estimated distribution will attain the value  $Z$ , otherwise, larger or smaller is

$$p = 1 - G(Z). \tag{11}$$

This  $p$  value is taken as the measure of statistical significance to evaluate the authenticity of the ROI. In other words, we determine that the ROI is tampered if  $p > \alpha = 4.56\%$  and not tampered otherwise.

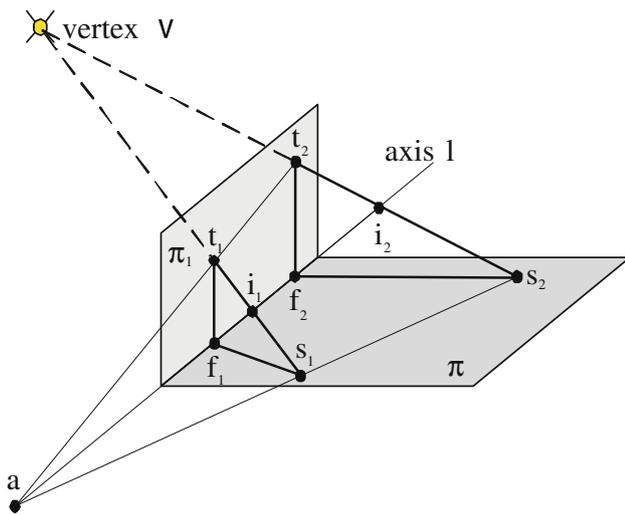
We use hypothesis testing to determine the ROI is tampered or not. First, we propose a hypothesis:  $\mathbf{H}_0 : \hat{Z} = \mu$ . We would like to use the following equation to weigh the deviation between the estimated value  $\hat{Z}$  from ROI and the mean value  $\mu$ . Then we define

$$P\{if \mathbf{H}_0 \text{ is true, reject } \mathbf{H}_0\} = P_\mu \left\{ \left| \frac{\hat{Z} - \mu}{\sigma} \right| \geq k \right\} = \alpha, \tag{12}$$

where  $\alpha = 0.05$  is the significance level. When  $\mathbf{H}_0$  is true,  $Z = \frac{\hat{Z} - \mu}{\sigma} \sim N(0,1)$ , i.e. standard normal distribution, we obtain the critical value  $k = z_{\alpha/2}$ . Consequently, we deny  $\mathbf{H}_0$  on condition that the  $\hat{Z}$  satisfies the following relationship:

$$|z| = \left| \frac{\hat{Z} - \mu}{\sigma} \right| \geq k = z_{\alpha/2}, \tag{13}$$

and regard it as tampered. Otherwise, we take in the estimated result of ROI and make no doubt of it.



**Fig. 2** Geometry of a planar homology. A plane,  $\pi_1$ , and its shadow, illuminated by a point light source  $\mathbf{v}$  and cast on a ground plane  $\pi$ , are related by a planar homology

### 3.4 Shadow geometry

Shadow compositing must be taken into account in some target scenes when there presents shadow projection from the sun. In this section, we utilize the planar homology [11,32,36] 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 homology, 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,  $\mathbf{H}$ ,

$$\mathbf{H} = \mathbf{I} + (\mu - 1) \frac{\mathbf{v}\mathbf{l}^T}{\mathbf{v}\mathbf{T}\mathbf{l}}, \tag{14}$$

where  $\mu$  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  $\pi_1$  and  $\pi$ . 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,  $\mu$ , 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

$$((\mathbf{t}_2 \times \mathbf{t}_1) \times (\mathbf{s}_2 \times \mathbf{s}_1)) \cdot (\mathbf{f}_2 \times \mathbf{f}_1) = 0, \tag{15}$$

$$\{\mathbf{v}, \mathbf{t}_1; \mathbf{s}_1, \mathbf{i}_1\} = \{\mathbf{v}, \mathbf{t}_2; \mathbf{s}_2, \mathbf{i}_2\}, \tag{16}$$

where

$$\mathbf{v} = (\mathbf{t}_2 \times \mathbf{s}_2) \times (\mathbf{t}_1 \times \mathbf{s}_1). \tag{17}$$

Therefore, we use this two constraints to detect composites 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  $\pi_1$  and  $\pi$ . 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.

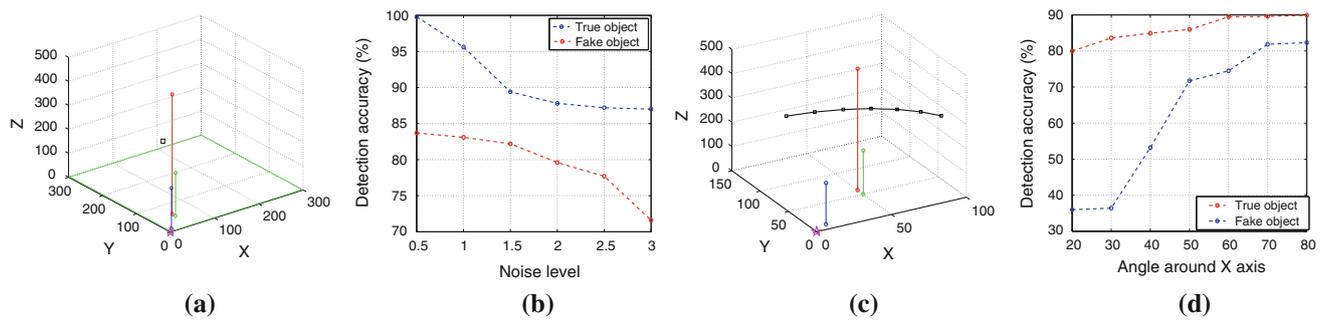
Hence, if the geometric constraints from two regions don't satisfy the Eqs. (15) and (16), i.e. the corresponding lines are not converging to a point on the axis and the cross ratios for the two regions are not equal, we may conclude that one of the region is tampered and further authenticating process is needed. Generally, three regions are required and the region which is inconsistent with the others may be identified as the fake one.

## 4 Experimental results

### 4.1 Computer simulation

The simulated camera has the focal length of  $f = 2,000$  (in pixels)<sup>1</sup>, the aspect ratio of  $\lambda = 1$ , the skew of  $\gamma = 0$ , and the principal point at  $u_0 = 640$  and  $v_0 = 320$ . The camera look-at position is at the world origin as shown as magenta star in Fig. 3a, c. In Fig. 3a, the camera position is  $[50, 88, 280]^T$ . We use the orthogonal constraints (green lines) to calibrate the camera. Note that all the lines are sampled into 10 equally distributed points. The referenced object is the vertical green segment, which is 170 unit high. We have two test objects in blue (authentic) and red (faked). We perform the detection on true and fake objects, and the detection accuracy for true object is defined as:  $\text{Detection accuracy} = \frac{\# \text{True detection}}{\# \text{Total}}$ . The same situation holds true for fake object detection. As shown in Fig. 3b, the detection accuracy decreases as the noise increases. For each noise level (the unit is one pixel), 1,000 independent trials are performed. In the evaluating process, we regard the target as a true object if its height locates between 150 and 210 units.

<sup>1</sup> Once we know the focal length (in millimeter) and the CCD width (in millimeter), the focal length in pixels can be easily converted using the following equation and vice versa.  $\text{focal length in pixels} = (\text{image width in pixels}) \times (\text{focal length in millimeter}) / (\text{CCD width in millimeter})$ .



**Fig. 3** Performance on 3D height measurement with respect to noises and camera orientations. **a** The referenced, authentic and faked objects are marked in *blue*, *green* and *red*, respectively. The camera is placed at the *black square* looking at the *magenta star*. **b** The detected rates

with respect to noises ranging from 0.5 to 3 pixels. **c** Seven cameras labeled with *black squares* form a trajectory to view the objects with different viewing directions. **d** Detected rate on actual and fake objects with respect to camera location (color figure online)



**Fig. 4** *Left* Two natural images with visually plausible composites marked in *red*. The reference and authentic objects are labeled in *green* and *yellow*, respectively. Two set of *parallel lines* labeled with *cyan*

and *blue* are perpendicular to each other. *Right* Rectified images using the estimated planar homography of the reference plane (color figure online)

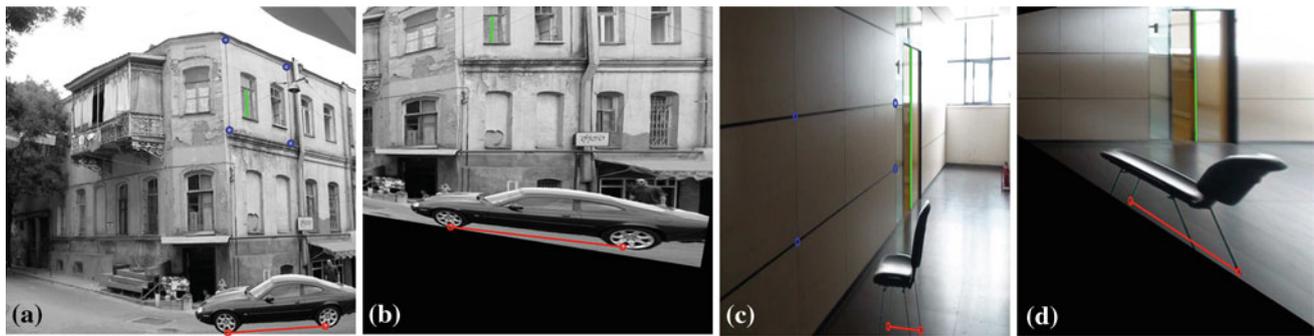
The second simulation is to evaluate the performance of fake region detection with respect to camera locations. In this experiment, camera internal parameters and the look at position are the same as the previous simulation. As shown in Fig. 3c, the camera location moves along a curve by rotating around the X axis from  $20^\circ$  to  $80^\circ$  (the unit of angle is  $10^\circ$  in degree). The camera height is 280. The reference, authentic and the faked objects are the same as the first simulation. To simulate the real scenario, a Gaussian noise of 1 pixel is added to the sampled points. A true object detection responds more robustly than fake region detection with respect to the various angles in the set up of camera. The reason is that it is difficult for a counterfeiter to insert a fake object which is consistent with the geometric constraints. Figure 3d shows the detection accuracy with respect to camera's viewing direction.

#### 4.2 Verification of a selected ROI

In this section, we verify the integrity of a selected ROI using real images. Note that the ROI is selected manually by a user as an area whose integrity is in question.

#### 4.2.1 Height measurement

Figure 4 shows two sets of images with composites which are visually plausible. Figure 4a, c is captured by a COTS camera with resolution  $1,024 \times 768$  and  $768 \times 1,024$ , respectively. The fake objects are inserted from unknown images. The referenced objects and suspicious targets are labeled in green and red segments. Note that in our test we obtain the referenced objects' 3D heights by hand. We define that  $\alpha = 0.05$ ,  $k = z_{\alpha/2} = z_{0.025} = 1.96$  throughout this section. The  $\mu$  and  $\sigma$  are set as (180, 15 cm) and (12, 3 cm), respectively. To calibrate the cameras, we use the constraints provided by two perpendicular vanishing points (the intersections of two sets of parallel lines labeled with blue and cyan in Fig. 4a, c). The rectified images using the estimated planar homography of the reference plane are shown in Fig. 4b, d. In Fig. 4a, the estimated height is 216.8 cm which is out of the human height probability distribution, since  $|z| = \left| \frac{216.8 - 180}{15} \right| = 2.45 > k$ . In Fig. 4c, the suspicious magic cube's estimated height is 24 cm from the referenced height (60 cm)



**Fig. 5** Composite detection based on measurement on a vertical plane with respect to the reference plane. *Left* Four blue circles on the walls (the reference plane) are used to calibrate the camera. The reference height is the green line segment. The red line is the distance we aim to

measure, which lies on the ground plane that is vertical to the reference plane. *Right* Rectified images using the estimated planar homography of the reference plane (color figure online)

of the chest, which diverges from the expected 12 cm as  $|z| = |\frac{24-12}{3}| = 4 > k$ . In addition, we estimated the vacuum bottle’s height (a true object labeled in yellow in Fig. 4c) to demonstrate that the proposed method doesn’t give false alarms. The estimated height for the vacuum bottle is 38.78 cm, yielding an error of 0.28 cm since its actual height is 38.5 cm.

#### 4.2.2 Measurement on vertical plane

The old building in Fig. 5a was captured by an unknown camera with the resolution  $1,000 \times 940$ . The car was copied from another unknown image. Figure 5c is a corridor image captured by a SONY camera. It is inserted with a chair from another image captured by the same camera. The mean value and standard deviation  $(\mu, \sigma)$  for Fig. 5a, c are (270, 20 cm) and (35, 5 cm), respectively. Here the car is designated as the ROI and we wish to estimate the distance between the car’s frontal and rear tyres’ centers. The reference plane is the building’s side face containing four corners labeled in blue circles. The estimation of distance between tyres can be seen as the point measurement (red circles in Fig. 5a) on the vertical plane. The window’s height in Fig. 5a is about 90 cm from our common knowledge, and thereby, the estimated distance is 365.7 cm. From  $|z| = |\frac{365.7-270}{20}| = 4.785 > k$ , we have sufficient evidence that the car is tampered. In Fig. 5d, the image rectification and metric measurement processes are the same as (b). The referenced door’s height is about 230 cm in world coordinate. The estimated distance between the chair’s legs is 14 cm compared with the true value 35 cm, and  $|z| = |\frac{14-35}{5}| = 4.2 > k$ .

#### 4.2.3 Measurement on arbitrary plane

At last, we show the detection process on image forgeries on an arbitrary plane, seeing Fig. 6. The image captured by an

unknown camera in Fig. 6a is tampered with a dollar bill cut from another unknown image. Figure 6d is a real scene with an insertion of letter “J” on the stands. The mean value and standard deviation  $(\mu, \sigma)$  for (a) is (15, 3 cm). After metric rectification, we use the reference distance (the green segment) to remove the scale ambiguity. The estimated width of the dollar is 44 cm which is an improper value since it has to be 15.6 cm, i.e.  $|z| = |\frac{44-15.6}{3}| = 9.47 > k$ .

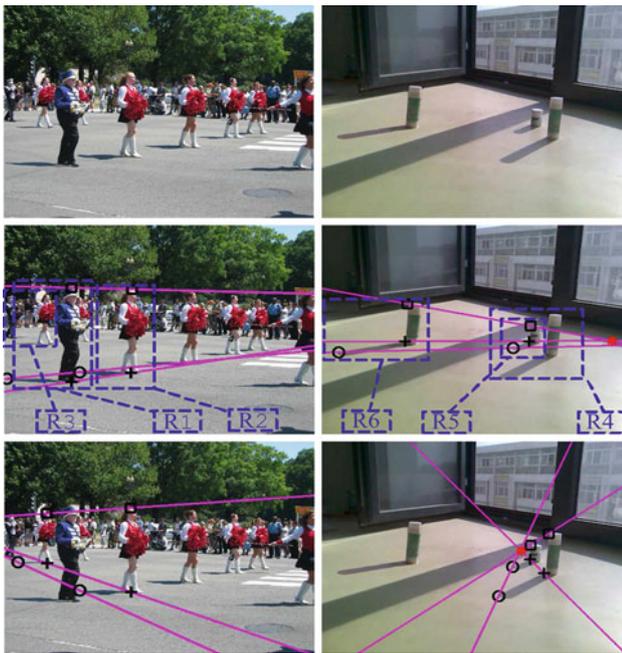
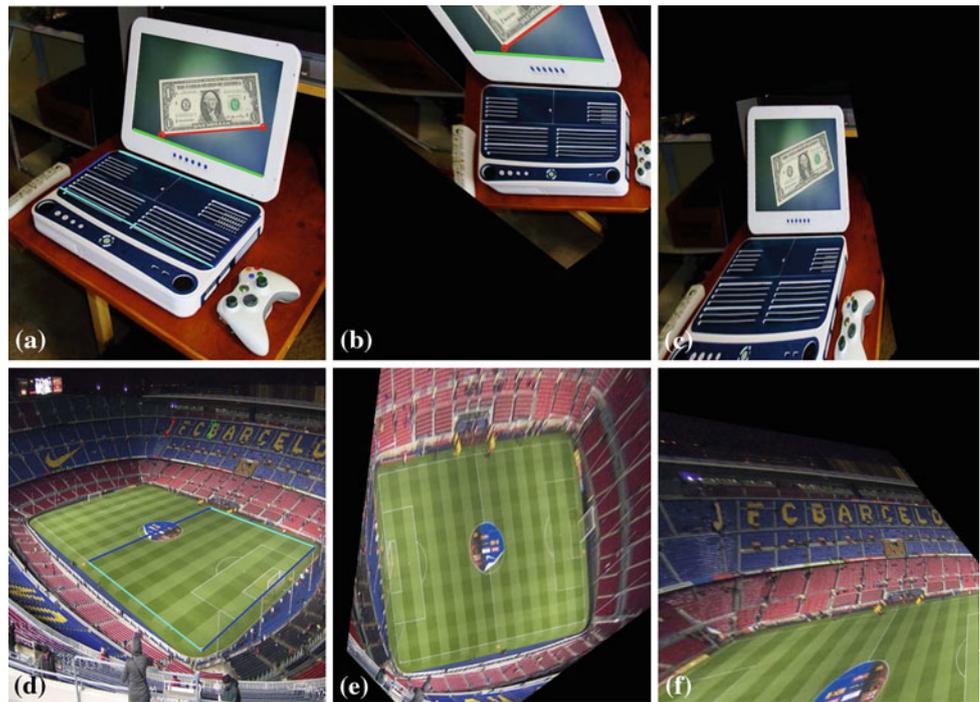
In Fig. 6d, we insert a letter “J” to the left of the authentic letter “F” on the stands. We have no knowledge about how high the letter “J” should be due to the uncontrolled image source. Therefore, we use the ratio of  $h_J, h_B$  (the 3D heights for the letters “J” and “B”) which is expected to be 1. Applying the proposed method in Sect. 3.2, the estimated ratio  $r$  from Fig. 6 is 1.3 which can inform us the suspicion of alphabet “J”. However, it is worthy noting that the stands is trapezoidal, not coplanar.

#### 4.3 Image forgery detection based on shadow geometry

Shadow composites can be found in sophisticated forgeries. Generally, it is still difficult for current image compositing technology to satisfy geometric constraints exactly. Therefore, it is possible to detect tampered regions by enforcing the geometric constraint modeled by the planar homology.

The estimation of the planar homology constraints requires three pairs of points. These six points are manually selected as shown in Fig. 7 in black squares, crosses and circles. First, we check the constraint in Eq. 15). The two corresponding lines of the two shadows in Fig. 7 (Row 2) don’t intersect on the axis, while those in Row 3 do. Take Fig. 7 (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. 7. Without surprise, the cross ratios that involve fake regions 1 and 6 are different (around 21% on average), while the cross ratios that involve

**Fig. 6** Composite detection based on measurement on an arbitrary plane with respect to the reference plane. **a, d** Blue and cyan lines on the reference plane are used to calibrate the camera. The reference height is the green segment. The red segment is the distance we aim to measure, which lies on an arbitrary plane with respect to the referenced plane. **b, e** Rectified images using the estimated planar homography of the reference plane. **c, f** Rectified images using the estimated planar homography of the arbitrary plane (color figure online)



**Fig. 7** Composite detection based on shadow geometry. *Top row* Two nature images with composited regions. *Middle row* The corresponding lines that involve composited regions ( $R1, R6$ ) do not intersect on the axis, i.e. disobeying the constraint in Eq. (15). In addition, they dissatisfy the characteristic invariance constraint in Eq. (16) (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  $t, f$  and  $s$  in Fig. 2, respectively (color figure online)

all authentic regions (regions 2 ~ 5) are within the relative difference of 1%. Note that cross ratios of the same shadow are not necessarily equal in various shadow pairs, as a new

**Table 1** Cross ratios of planar homologies in Fig. 7

Region A	Region B	$\mu_A$	$\mu_B$	Diff ratio (%)
R1	R2	0.1741	0.1231	29.26
R2	R3	0.1587	0.1573	0.8794
R1	R3	0.4454	0.4966	11.51
R4	R5	0.6298	0.6352	0.8647
R4	R6	0.4473	0.3384	24.35
R5	R6	0.3237	0.2625	18.92

pair that defines a different plane  $\pi_1$  in Fig. 2, and thus results in a different location of  $\mathbf{i}$ .

## 5 Discussion

### 5.1 Conclusion and future work

This paper has presented a new framework for detecting image forgery based on single view metrology and shadow geometry constraints. Algorithms have been introduced to obtain different kinds of measurements for fake region detection: 3D height measurement requiring a reference distance; measurements on a vertical or arbitrary plane with respect to the reference plane. The proposed method is especially useful when the scene is a wide area and the required calibration objects are not measurable. The experimental results have demonstrated that this method was efficient and able to be applied to a variety of target scenes. As a pragmatic and flexible framework, it is also simple and easy to implement.

However, it is evident that the problem of detecting digital image forgeries is a complicated one with no universally applicable solution, and what we need is a large set of methods based on different principles that can be applied to all tampered images. Then such accumulative evidence may provide a convincing proof that each individual method cannot carry out.

Digital forensic techniques have been developed on detection which can be categorized as: region duplication, inconsistencies in camera response function, inconsistencies in lighting and inconsistencies in sensor noise [8]. The proposed method concentrates on the inconsistencies in metric value and geometric constraints. The authenticating results provides a reliable clue for further determination since it can be combined with other methods to improve the detecting robustness. For example, the suspicious object alarmed by our method facilitates the method in [31], which requires the selection of small fixed size image blocks to yield a reduced dimension representation. The principal component analysis in [31] may detect duplicated objects more accurately provided that our method provides skeptical object to be authenticated.

At the same time, our method deploying the shadow inconsistency of planar homology may lend credence for other methods based on shadow geometry such as [9] to produce more reliable and stable results. It would seem from what we mentioned that the proposed method can be used to combined with [9] to detect inconsistencies in cast shadows, in which the method in [9] stands on the ground that light travels in a straight line, a point on the shadow, its corresponding point on the object, and the light source all lie on a single line.

## 5.2 Limitations

First of all, the limitation underlying in the height measurement and measurements on vertical and arbitrary plane are dependent on particular geometric relationship to recover the camera calibration. Admittedly, the metric value cannot be retrieved without the first step of the availability of parallel structures in the image. Provided that the scene fails to offer required geometric constraints, the detection on tampered region cannot be figured out by our method.

Second, the shadow geometry authentication, as a matter of fact, is limited in the scene with the availability of two shadow regions. Since the method evaluates the consistency of cross ratio and convergence of the axis between two shadow objects, it fails in the situation where no shadow or only one casting shadow is available. On the other hand, some requirements that must be taken to select appropriately matched points on the object and its shadow press limitations to this geometric analysis of homology characters and light position.

**Table 2** Process cost for experiments (in second)

	Figure 4c	Figure 5c	Figure 6a	Figure 6d	Figure 7
Processing cost	0.01563	3.359	2.938	3.828	1.688

Finally, the authentication process requires manual interference. After the achievement of calibration, the semi-automatic method requires the manual selected features to determine the suspicious region to be authenticated. Besides, the inputs such as the point measurement for skeptical object may not be determined automatically, which need manual resolve.

Though our method is not desirable for batch processing, the propose method is effective and easy implemented. The experiments are processed under the computer environmental parameters of Intel Pentium processor 1.73 GHZ (single core) and 1 GB memory. The process cost measured by CPU processing time for experiments are listed in the Table 2 (in the unit of second).

**Acknowledgments** This work was supported by National Natural Science Foundation of China (No. 60905019, 50735003), Scientific Research Foundation for the Returned Overseas Chinese Scholars, Open Projects Program of National Laboratory of Pattern Recognition, Tianjin University 985 research fund, and State Key Laboratory of Precision Measuring Technology and Instruments open fund.

## References

1. Bayram, S., Memon, N., Ramkumar, M., Sankur, B.: A classifier design for detecting image manipulations. *IEEE Int. Conf. Image Process.* **4**, 2645–2648 (2004)
2. Bayram, S., Sencar, H.T., Memon, N.: Source camera identification based on cfa interpolation. In: *IEEE International Conference on Image Processing*, Genoa, Italy, vol. 3, pp. 69–72 (2005)
3. Cao, X., Foroosh, H.: Camera calibration using symmetric objects. *IEEE Trans. Image Process.* **15**(11), 3614–3619 (2006)
4. Chen, M., Fridrich, J., Goljan, M., Lukas, J.: Determining image origin and integrity using sensor noise. *IEEE Trans. Inf. Forensics Secur.* **3**(1), 74–90 (2008)
5. Chen, W., Shi, Y., Su, W.: Image splicing detection using 2d phase congruency and statistical moments of characteristic function. In: *Proceedings of SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents*, p. 6505 (2007)
6. Cox, I.J., Miller, M.L., Bloom, J.A.: *Digital Watermarking*. Morgan Kaufmann, Menlo Park (2002)
7. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. In: *International Conference on Computer Vision*, pp. 434–441 (1999)
8. Farid, H.: A survey of image forgery detection. *IEEE Signal Process. Mag.* **26**(2), 16–25 (2009)
9. Farid, H., Bravo, M.J.: Image forensic analyses that elude the human visual system. In: *SPIE Symposium on Electronic Imaging*, San Jose, CA (2010)
10. Fridrich, J., Soukal, D., Lukas, J.: Detection of copy–move forgery in digital images. In: *Digital Forensic Research Workshop*, Cleveland, OH (2003)

11. Gool, L.V., Proesmans, M., Zisserman, M.: Planar homologies as a basis for grouping and recognition. *Image Vis. Comput.* **16**(1), 21–26 (1998)
12. Gou, H., Swaminathan, A., Wu, M.: Noise features for image tampering detection and steganalysis. In: *IEEE International Conference on Image Processing*, San Antonio, pp. 97–100 (2007)
13. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, London (2004)
14. Hsu, Y.F., Chang S.F.: Image splicing detection using camera response function consistency and automatic segmentation. In: *IEEE International Conference on Multimedia and Expo* (2007)
15. Johnson, M., Farid, H.: Detecting photographic composites of people. In: *6th International Workshop on Digital Watermarking* (2007)
16. Johnson, M., Farid, H.: Exposing digital forgeries in complex lighting environments. *IEEE Trans. Inform. Forensics Secur.* **2**, 450–461 (2007)
17. Johnson, M.K., Farid, H.: Exposing digital forgeries by detecting inconsistencies in lighting. In: *ACM Multimedia and Security Workshop*, New York, NY, pp. 1–9 (2005)
18. Johnson, M.K., Farid, H.: Exposing digital forgeries through chromatic aberration. In: *ACM Multimedia and Security Workshop*, pp. 48–55 (2006)
19. Johnson, M.K., Farid, H.: Metric measurements on a plane from a single image. In *Technical Report*, TR2006-579 (2006)
20. Liebowitz, D., Zisserman, A.: Metric rectification for perspective images of planes. In: *Computer Vision and Pattern Recognition* (1998)
21. Lin, Z., Wang, R., Tang, X., Shum, H.-Y.: Detecting doctored images using camera response normality and consistency. In: *IEEE Computer Vision and Pattern Recognition*, pp. 1087–1092 (2005)
22. Lukas, J., Fridrich, J.: Estimation of primary quantization matrix in double compressed jpeg images. In: *Digital Forensic Research Workshop* (2003)
23. Lukas, J., Fridrich, J., Goljan, M.: Determining digital image origin using sensor imperfections. *Proc. SPIE Electron. Imaging Image Video Commun. Process.* **5685**(2), 249–260 (2005)
24. Lyu, S., Farid, H.: How realistic is photorealistic? *IEEE Trans. Signal Process.* **53**(2), 845–850 (2005)
25. Mahdian, B., Saic, S.: Detection of copy-move forgery using a method based on blur moment invariants. *Forensic Sci. Int.* **171**(2), 180–189 (2007)
26. Ng, T.-T., Chang, S.-F.: A model for image splicing. In: *IEEE International Conference on Image Processing*, pp. 1169–1172 (2004)
27. Ng, T.-T., Chang, S.-F., Sun, Q.: Blind detection of photomontage using higher order statistics. In: *Proceedings of the 2004 International Symposium on Circuits and Systems* **5**, 688–691 (2004)
28. Pevny, T., Fridrich, J.: Detection of double-compression in jpeg images for applications in steganography. *IEEE Trans. Inf. Forensics Secur.* **3**(2), 247–258 (2008)
29. Popescu, A., Farid, H.: Exposing digital forgeries by detecting duplicated image regions. In: *Technical Report TR2004-515*, Department of Computer Science, Dartmouth College (2004)
30. Popescu, A.C., Farid, H.: Exposing digital forgeries by detecting traces of re-sampling. *IEEE Signal Proc. Mag.* **53**(2), 758–767 (2005)
31. Popescu, A.C., Farid, H.: Exposing digital forgeries by detecting duplicated image regions. In: *Technical Report*, TR2004-515, Dartmouth College, Computer Science (2010)
32. Shen, Y., Foroosh, H.: View-invariant action recognition from point triplets. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(10), 1898–1905 (2009)
33. Swaminathan, A., Wu, M., Liu K.J., Ray: Digital image forensics via intrinsic fingerprints. *IEEE Trans. Inf. Forensics Secur.* **3**(1), 101–117 (2008)
34. Wang, G., Hu, Z., Wu, F., Tsui, H.T.: Single view metrology from scene constraints. *Image Vis. Comput.* **23**(9), 831–840 (2005)
35. Wang, W., Farid, H.: Detecting re-projected video. In: *International Workshop on Information Hiding* (2008)
36. Yacoob, Y., Davis, L.: Segmentation of planar objects and their shadows in motion sequences. *Int. J. Comput. Vis.* **67**(1), 53–69 (2006)
37. Zhang, W., Cao, X., Feng, Z., Zhang, J., Wang, P.: Detecting photographic composites using two-view geometrical. In: *IEEE International Conference on Multimedia and Expo* (2009)

### Author biographies

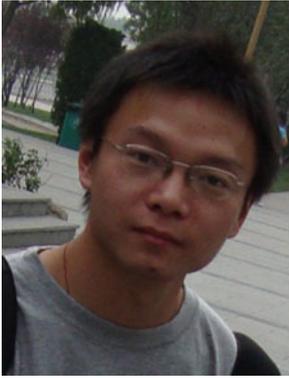


**Lin Wu** received her B.E. in software engineering from the Dalian University of Technology, Dalian, China, with excellent undergraduate awards in 2008. Since September 2008, she has been a graduate in the School of Computer Science and Technology, Tianjin University. Her interests include camera calibration and multi-view geometry.



**Xiaochun Cao** received his B.E. and M.E. degrees, both in computer science, from Beihang University (BUAA), Beijing, China in 1999 and 2002, respectively. He received his Ph.D. degree in computer science from the University of Central Florida, Orlando, FL in 2006, with his dissertation nominated for the university-level Outstanding Dissertation Award. After graduation, he spent about two-and-a-half years at the ObjectVideo Inc. as a research scientist. Since August 2008, he has

been with the Tianjin University, where he is currently professor of computer science and adjunct professor of computer software.



**Wei Zhang** received his B.E. degree in software engineering from the School of Computer Software, Tianjin University, Tianjin, China, in 2008. He is currently a second-year graduate student in the School of Computer Science and Technology, Tianjin University. His research interests lie in the computer vision field, covering mainly image processing, multimedia, forensic and pattern recognition.



**Yang Wang** is a first year master's candidate in the School of Computer Science and Technology of Tianjin University, China. Before joining the Tianjin University, he had obtained his bachelor's degree from the Dalian University of Technology, China with excellent undergraduate awards in 2009. His research area includes machine learning, image processing, data mining and information retrieval.