# Counterfeit Detection Based on Unclonable Feature of Paper Using Mobile Camera

Chau-Wai Wong, *Member, IEEE,* and Min Wu, *Fellow, IEEE*

***Abstract*—This work studies the authentication problem of specific pieces of paper using mobile imaging devices. Prior work showing high matching accuracy has used the *normal vector field*, which serves as a unique, microscopic, physically unclonable feature of paper surfaces, estimated by consumer-grade scanners. Industrial cameras were also used to capture the appearance of the surface rendered after the normal vector field based on the laws of optics under a semi-controlled lighting condition. In comparison, past explorations based on mobile cameras were very limited and have not had substantial success in obtaining consistent appearance images due to the uncontrolled nature of the ambient light. We show in this work that images captured by mobile cameras can be used for authentication when the camera flash is exploited to create a semi-controlled lighting condition. We have proposed new algorithms to demonstrate that the normal vector field of the paper surface can be estimated by using multiple camera-captured images of different viewpoints. Perturbation analysis shows that the proposed method is robust to inaccurate estimates of camera locations, and a matching accuracy of $10^{-4}$ in equal error rate can be achieved using 6 to 8 images under a lab-controlled ambient light environment. Our findings can relax the restricted imaging setups and enable paper authentication under a more casual, ubiquitous setting with a mobile imaging device, which may facilitate duplicate detection of paper documents and counterfeit mitigation of merchandise packaging.**

***Index Terms*—Anti-Counterfeit, Paper Physically Unclonable Features (PUFs), Mobile Cameras, Photometric Stereo, Microstructure**

## I. Introduction

Merchandise packaging and valuable documents such as tickets and IDs are common targets for counterfeiting. Low-cost surface structures have been exploited for counterfeit detection by using their optical features. The randomness of the surface makes the structures physically unclonable or difficult to clone to deter duplications. Such surface structures can be extrinsic by adding ingredients such as fiber [2], [3], small plastic dots [2], air bubble [2], powders/glitters [4] that are foreign to the surface; and the surface structures can also be intrinsic by exploring the optical effect of the microscopic roughness of the surface, such as the paper surface formed by inter-twisted wood fibers [4]–[8]. The inherent randomness of the microscopic roughness quantified using the *normal vector field* has been used as a feature for the unique identification of a particular patch of a surface in [4], [5].

In this paper, we focus on the intrinsic property of the paper surface for counterfeit detection and deterrence, and seek to find a more casual, ubiquitous imaging setup using consumer-grade mobile cameras under commonly available lighting conditions. The previous work in [4]–[6] shows that the microscopic roughness of the paper surface can be optically captured by consumer-grade scanners and industrial cameras, both under controlled lighting conditions in the form of image appearance rendered according to the physical law of light reflection at the paper surface. The appearance images, and the subsequent normal vector field of the surface estimated from the appearance images, can achieve satisfactory authentication results. However, recent work in [4], [8] also showed that if the ambient lighting is not well controlled, the image appearance alone has not achieved satisfactory authentication results. Instead, features based on the intensity gradient of visually observable dots are less sensitive to the change of lighting and may be used for authentication at the cost of higher algorithm complexity and moderate discrimination capabilities [8].

Satisfying two requirements may facilitate paper authentication via mobile cameras. First, the mobile-captured images should be comparable in resolution and contrast to those captured by scanners. Second, lighting should be controlled to render a desirable appearance of the paper. The first requirement can be qualitatively checked by comparing the acquired images from scanners and mobile cameras. Images acquired in both ways do have similar, detailed intensity fluctuations when zoomed in. The second requirement can be fulfilled by activating the flash next to the camera lens on mobile devices. The desirable appearance of the surface can be reasonably expected from the geometric arrangement between the camera and the surface.

As we shall show in this paper, camera flash exploited for creating semi-controlled lighting conditions can significantly improve the performance of using appearance images as the authentication feature. More importantly, by exploiting the underlying rendering principle of the appearance of the surface, *i.e.*, the fully diffuse reflection model [5], [9], one can estimate the normal vector field of the surface without resorting to more restricted acquisition conditions. To the best of our knowledge, this paper together with its preliminary version [1] is the first set of work using mobile cameras to obtain an effective estimate of the normal vector field of the paper surface for authentication. In this journal version, additional

Originally published in IEEE T-IFS 2017. Updated to correct typographical errors; no substantive changes made.

Manuscript received September 27, 2016; revised January 6, 2017; accepted February 16, 2017. Date of publication April 17, 2017. The preliminary result of this work has been presented at WIFS 2015 [1]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chip-Hong Chang.

The authors are with the Department of Electrical and Computer Engineering and the Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 USA (e-mail: cwwong@terpmail.umd.edu; minwu@umd.edu).

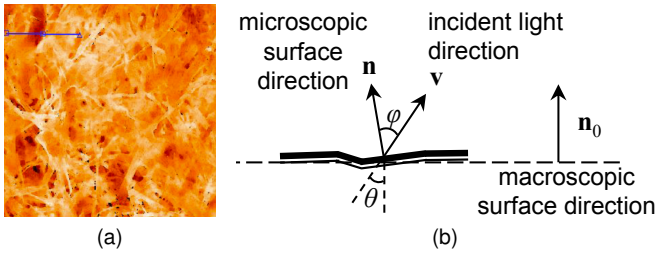


**Fig. 1:** (a) A topographic map of a 1mm-by-1mm region of a paper surface captured by a confocal microscope, reproduced from [10]. The pseudo-color represents the elevation of fibers in the $z$-direction. (b) Microscopic view of a particular spot on a paper surface. Note that $\varphi$ and $\theta$ are not co-planar in most cases. All vectors are unit vectors.

experimental results are presented for more practical capturing scenarios. Extended perturbation analyses of discrimination power on two factors, namely, the inaccuracy of estimated camera locations and the number of images used for normal vector field estimation, are conducted and explained using statistical methods. "Ground-truth" 3-D structure of a paper surface is obtained with confocal microscopy in order to quantitatively examine the linkage between the appearance and the physical structure of the paper surface.

The paper is organized as follows. In Section II, we review light reflection models, the method for paper surface registration, and the method for paper authentication. In Section III, we examine the authentication performances when restricted imaging setups are used, which serve as a performance baseline. In Sections IV and V, we propose methods working under a more flexible setup—mobile cameras with built-in flash, and compare the performances with prior work. In Section VI, we conduct perturbation analysis to demonstrate the practicality of the proposed mobile camera-based authentication method. In Section VII, we use confocal microscopic data from a physics aspect to elucidate a deeper understanding of the proposed work and also address some practical issues. In Section VIII, we conclude the paper.

## II. Background and Preliminaries

### A. Optical Imaging of Paper and Light Reflection Models

Seemingly smooth paper surfaces contain inherent microscopic 3-D structure due to overlapped and inter-twisted wood fibers. This microscopic structure is different from one paper to another and even from one location to another on the same paper, and therefore can serve as a unique identifier or fingerprint. One quantitative feature of such a 3-D structure, the surface direction, has been successfully exploited for authentication in [4]–[6].

Fig. 1(a) shows a topographic map of a 1 mm-by-1 mm region of a paper surface estimated from images captured by a confocal microscope [10]. The microscopic roughness due to fibers is clearly shown. The visual appearance of the surface follows the law of optics.

Geometrical light reflection models such as specular model and diffuse model have been widely used in computer vision/graphics applications, due to their good approximations to the law of optics and relatively simple analytical forms [9], [11], [12]. Under the *specular* reflection model, the perceived intensity is dependent on the direction of the reflected light and the direction of the eye/sensor. Under the *diffuse* reflection model, the perceived intensity is dependent on the direction of incident light and the normal direction of the microscopic surface. The appearances of most surfaces contain both reflection components.

Previous authentication work [4], [5] treating paper as a fully diffuse surface has led to satisfactory results. We follow a fully diffuse modeling assumption and provide an experimental justification in the discussion section that the strengths of the diffuse component versus the specular component is about six to one.

Fig. 1(b) shows the microscopic surface normal direction, $\mathbf{n}$, of a particular spot $\mathbf{p}$ in a microscopic view (which is often different from the macroscopic surface direction, $\mathbf{n}_0$), and an incident light direction, $\mathbf{v}$. The perceived reflected intensity $l_r$ of the fully diffuse reflectance model [5], [9] is

$$l_r(\mathbf{p}) = \lambda \cdot l(\mathbf{p}) \cdot \underbrace{\mathbf{n}(\mathbf{p})^T \mathbf{v}(\mathbf{p})}_{=\cos\varphi(\mathbf{p})}, \tag{1}$$

which depends on the angle $\varphi(\mathbf{p})$ between the normal direction of the surface at the microscopic level, $\mathbf{n} = (n_x, n_y, n_z)$, and the direction where the incident light is coming from, $\mathbf{v} = (v_x, v_y, v_z)$; the strength of the light at the current spot, $l(\mathbf{p})$; and the albedo, $\lambda$, characterizing the physical capability of reflecting the light [11], [12]. In our work, the assumption of $\lambda$ being constant over the whole paper patch is found to hold well for the purpose of authentication.

For an ideal point light source, the light strength $l(\mathbf{p})$ over a spatial field is modeled by considering the effect of energy fall-off due to the travel distance of the light [12]. In practice, a camera flash is not a perfect point source but has a finite dimension, *e.g.*, in a disc-like shape. When the flash is not perfectly oriented toward the paper surface, it can lead to a foreshortening effect reducing the strength of the light arriving at the projected point of the light on paper. Therefore, it is practically difficult to model $l(\mathbf{p})$ with a high precision for nonideal point sources. Instead, we estimate $l(\mathbf{p})$ by exploiting its spatial smoothness property. With the values of $l(\mathbf{p})$, the microscopic structure can then be determined in terms of normal vectors, $\mathbf{n}(\mathbf{p})$.

### B. Paper Patch Registration

We use a simple square-shaped registration container from our recent work [4] as shown in Fig. 2(b), and a tri-patch extension as shown in Fig. 6(b), to facilitate the precise registration in our experiments. Considering a printing resolution of 600 pixels per inch, each square container of size $\frac{2}{3}$-by-$\frac{2}{3}$ inch$^2$ (1.69-by-1.69 cm$^2$) corresponds to a box of 400-by-400 in pixel at a line width of 5 pixels, and there are four circles at corners of each square. A preliminary alignment based on four boundaries can be achieved using a Hough transform, and subpixel resolution refinement with perspective transform compensation is then carried out based on the circle markers. Lens location relative to the surface in the world

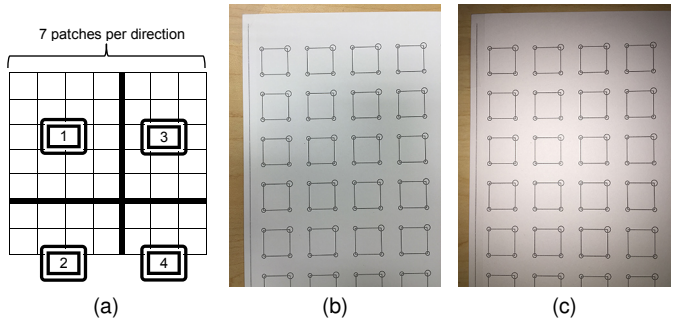


**Fig. 2:** (a) Four camera shots are needed for capturing 49 square patches located on a piece of paper. (b) Image captured at position#1 under ambient (fluorescent) light without flash (database 505), and (c) with flash (database 501). Capturing device: iPhone 6.

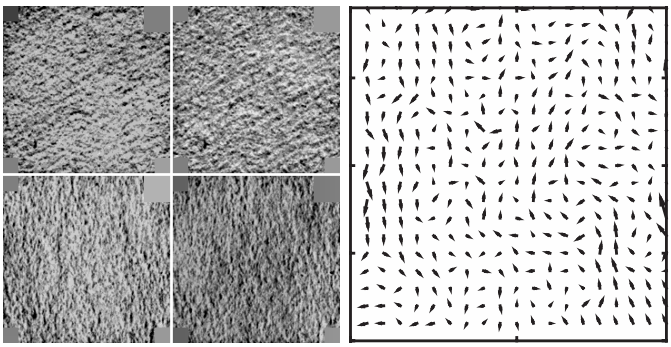

**Fig. 3:** Scanned images from four perpendicular orientations of a piece of $\frac{2}{3}$-by-$\frac{2}{3}$ inch$^2$ (1.69-by-1.69 cm$^2$) paper, and the resulting estimated norm map covering $1/100$ area of that paper.

coordinate system can be readily calculated from the estimated perspective transform matrix, and then the direction of incident light at every pixel location is known. Note that the world coordinate system is naturally defined to have the $xy$-plane located at the bottom plane of the paper surface and the $z$-axis pointed upwards. All camera-captured images were unwarped to remove the effect of lens distortion before being used if they were captured by a camera. This step improves the matching performance on average by $0.04$ in terms of correlation value in our experiments.

### C. Authentication via Hypothesis Testing

We approach the patch verification problem as a binary hypothesis testing problem [13] using discriminative features derived from images of the paper. The null hypothesis $H_0$ is that the test/query patch does not match with the patch from the reference database, whereas the alternative hypothesis $H_1$ is that the test/query patch matches with the reference patch. To quantify the degree of match, we use the normalized sample correlation $\hat{\rho}$ on the pair of extracted features, *e.g.*, pixel intensity in Section IV and surface normal vector in Section V. We estimate the probability density functions (PDFs) $f_{\hat{\rho}|H_0}(\hat{\rho})$ and $f_{\hat{\rho}|H_1}(\hat{\rho})$ that have very distinct mean values under unmatched and matched cases, and make a decision using the simple thresholding rule on an observed value of the random variable $\hat{\rho}$.

Under the simple thresholding rule with threshold $\tau$, the detection rate is defined to be $P_D(\tau) = \int_\tau^\infty f_{\hat{\rho}|H_1}(\xi)\, d\xi$ [or its complement, the miss rate, $P_M(\tau) = 1 - P_D(\tau)$] and the false-alarm rate is defined to be $P_F(\tau) = \int_\tau^\infty f_{\hat{\rho}|H_0}(\xi)\, d\xi$. The receiver operating characteristic (ROC) curve $\big(P_F(\tau), P_D(\tau)\big)$ can be drawn by varying the value of $\tau$ to reveal the discrimination capability of the system. Alternatively, the equal error rate (EER), $\{P_{\text{EE}} \,|\, P_{\text{EE}} = P_F(\tau) = P_M(\tau), \tau \in \mathbb{R}\}$, can be used as a compact, one-score indicator for the discrimination capability. For Gaussian and Laplacian distributions, it is not difficult to derive the analytical forms of EER to be $\Phi\left(\frac{\mu_0-\mu_1}{\sigma_0+\sigma_1}\right)$ and $\frac{1}{2}\exp\left[\frac{\lambda_0\lambda_1}{\lambda_0+\lambda_1}\,(\mu_0-\mu_1)\right]$, respectively [13], [14], where $\Phi(\cdot)$ is the cumulative density function for the standard Gaussian distribution. The theoretical quantities, including the mean $\mu_i$, standard deviation $\sigma_i$, and rate $\lambda_i$, can be replaced by their estimates from the real data. In this way, EER can be estimated even when there are a relatively limited number of data points and/or the PDFs are widely separated, in which the true tails of the PDFs may not be adequately revealed by simulated data. More discussions on using practical data for the theoretical model described above can be found in Section VII-E.

## III. Paper Authentication Using Scanners and Cameras

### A. Norm Maps by Scanners

The norm map as a physical feature of a paper surface has been found to have strong discrimination power. Clarkson *et al.* [5] used the fully diffuse reflectance model as described in Eq. (1) to estimate the projected normal directions at all integer-pixel locations of the surface. We refer to the collection of normal vectors for all pixels as the *normal vector field* (containing $x$-, $y$-, and $z$-components), and its projection onto surface plane as the *norm map* (containing $x$- and $y$-components only). A norm map can be estimated using images scanned from four different orientations of the paper: $0°$, $90°$, $180°$, and $270°$. Without knowing the exact direction of incident light, an estimate of one component of the norm map can be obtained as the difference between two scans in exactly opposite directions, canceling the effect of the unknown incident direction of the scanner light. The norm map containing randomly distributed vectors has been used as a feature for the unique identification of a particular patch of a surface in [4], [5]. In [5], a seeded hash is computed by random projection, and the Hamming distance of two hashes is used as the decision statistic. The sample statistics such as mean and variance measured from Fig. 8 of [5] reveal the EER to be between $10^{-130}$ and $10^{-15}$ (see Table VI for comparison) per our discussion on estimating the EER in Section II-C.

In order to provide more accurate norm map estimates as the reference data for our proposed method in Section V, we improve the norm map estimation algorithm over those in [4], [5] by removing the global bias for $x$- and $y$-components of the estimated norm map. Below we carry out experiments using the improved norm map estimator to provide a baseline for comparisons in later sections.

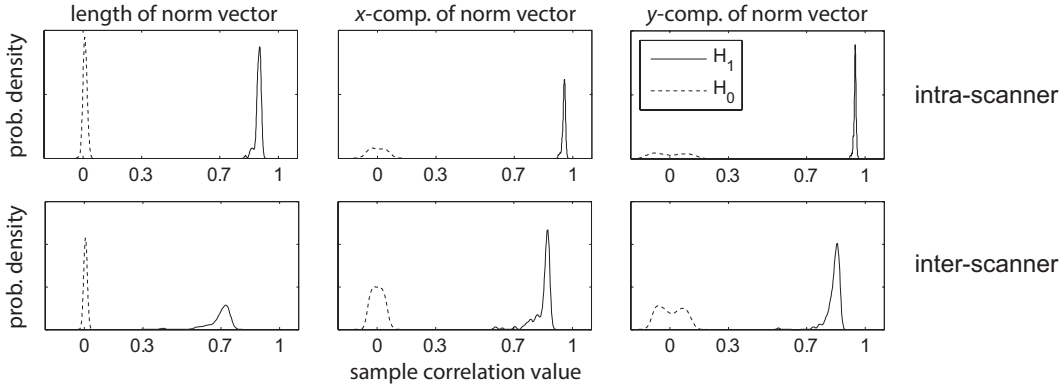


**Fig. 4:** Estimated PDFs of sample correlation coefficient $\hat{\rho}$ for unmatched ($H_0$) and matched ($H_1$) cases. First row (intra-scanner): Datasets #2–3 (test) vs. #1 (ref.) of scanner 2450, and second row (inter-scanner): Datasets #1–3 (test) of scanner GT vs. #1 (ref.) of scanner 2450. Features: length (column 1), $x$-component of normal vector (column 2), and $y$-component of normal vector (column 3).

We estimated norm maps for $49$ distinct square-shaped patches located on a piece of paper. The acquisition procedure was repeated using two Epson scanners: Perfection 2450 and GT-2500. Sample patches for scanner 2450 and the resulting norm map estimate of $1/100$ of the patch size are shown in Fig. 3. Authentication using the hypothesis testing described in Section II-C was carried out by correlating the test feature with the reference feature. Three features, namely, the normal vector's length, $x$- and $y$-components, were tested and the results are shown in the three columns of Fig. 4, respectively. Each plot contains two estimated PDFs of sample correlation coefficient $\hat{\rho}$: one for matched cases ($H_1$), and the other for unmatched cases ($H_0$). All six plots reveal that the distributions for the two hypotheses are far apart and have no overlap, thus a threshold can be set to have no false alarm and no miss detection, suggesting an excellent authentication performance. In addition, the performance for the intra-scanner case (*i.e.*, both test and reference data were obtained using the same scanner) shown in the first row of Fig. 4 is slightly better than that for the inter-scanner case (*i.e.*, test and reference data were obtained using different scanners) shown in the second row. They reveal that different acquisition devices can give slightly inconsistent norm map estimates but the inconsistency is not strong.

### *B. Appearance Images by Cameras*

Instead of using scanners to capture images with directional linear light and closely placed imaging sensors, Voloshynovskiy *et al.* [6], [7] examined the imaging setup of using two industrial cameras with a semi-controlled lighting condition—a fixed, ring-shaped light source. The resulting images have similar appearances during multiple capturing instances due to the semi-controlled lighting conditions. The ROC curves from [6] reveal that the EER is around $10^{-4}$.

Mobile cameras were used to test the authentication performance under uncontrolled ambient light in [4]. The uncontrolled light can lead to unpredictable surface appearances per the light reflection model in Eq. (1). Even using newer mobile cameras such as the iPhone 6 with improved acquisition quality over the older mobile devices, the authentication performances under uncontrolled ambient light are still limited, as revealed by Fig. 3 of [4]. One way to improve the authentication performance as shown by Diephuis *et al.* [8] is to use the intensity gradient-based features, *e.g.*, scale-invariant feature transform (SIFT), of high contrast spots that are less sensitive to the change of lighting, at the cost of increasing the design complexity of the authentication system. Extrapolating data points from Table 2 of [8] into the ROC plot of Fig. 8 of [8], we estimate the performance of the proposed SIFT-based method to be around $10^{-2}$ in terms of EER (see Table VI for comparison).

## IV. Proposed Paper Authentication Using Image Appearance Under the Camera Flash

Inspired by the success of the approaches discussed in Section III in which lighting for image acquisition is well controlled, we explore a semi-controlled lighting condition with the help of the built-in flash of mobile cameras for authentication. We achieve the semi-controlled lighting condition by exploiting the fact that relative positions among the light source, the lens, and the paper patch are known or can be estimated from a captured image.

The simplest case, presented in Section IV (this section), is to use the appearance of the patches when cameras are positioned at the same location relative to the physical patch so that the effect of lighting is the same for instances of capturing test and reference images. A more sophisticated case, presented in Section V (the next section), is to exploit the physics of lighting and to use multiple images for estimating the normal vector field as the feature for authentication.

### *A. Capturing Conditions and Proposed Method*

Patches were acquired by the built-in cameras of three mobile devices, iPhone 6, iPhone 5s, and iPhone 5, with and without a flash. The capturing process was done in a large room with 12 overhead fluorescent light arrays, and in a small room with 2 overhead fluorescent light arrays, respectively. The device was held by hand approximately in parallel with the surface of a piece of paper and at a height

of about $z = 15.5\,\text{cm}$. Detailed *capturing conditions* and the corresponding database ID that will be referred to in the remaining part of this section are shown in Table I.

We use a total of $49$ distinct square-shaped paper patches for the experiment. To acquire a *database* of a particular capturing condition, we captured three images for every patch, with slight camera rotation and panning among different capturing instances. Within each database, we refer to the $49$ patches for the $i$th capturing instance as Dataset #$i$, for $i = 1, 2, 3$. To speed up the capturing process, patches were acquired together with neighboring patches located on the same piece of paper. A total of four shots were needed to capture the whole region containing the patches, and the camera positions relative to the paper are shown in Fig. 2(a). Boundaries among different shots are separated by thick lines. Fig. 2(b) and (c) containing luminance non-uniformity were acquired without and with flash for the top-left $20$ patches on the layout of the paper.

The way of capturing the whole database of patches as laid out above ensures that any pair of matched test and reference images are captured at the same location relative to the physical patch. That is, the incident light for the test and reference images are the same, effectively controlling the acquisition conditions. In this way, the perceived intensities of patches are similar across different capturing instances per the fully diffuse light reflection model in Eq. (1). We shall examine in Section V how an authentication scheme should be designed when we do not constrain the relative locations of the cameras to the physical patches.

Each patch in the captured image was extracted, warped, and registered to a grid of 200-by-200 pixels using the registration procedure outlined in Section II-B. In this experiment, images captured at the height of about $z = 15.5\,\text{cm}$ contain around $300$ pixels along the edge/side for each patch in raw images, which are of enough resolution ($1.25\times$ more pixels than necessary) to generate the registered images. The collection of the $200 \times 200$ pixel values of the registered image is then considered as a feature for the paper patch, and normalized sample correlation $\hat{\rho}$ between features from test and reference patches can be calculated for authentication using the hypothesis testing described in Section II-C. In this experiment, Datasets #2 and #3 of a database are considered as the test data, and Dataset #1 of the same or a different database is considered as the reference data.

### B. Experimental Results

The contrast of the PDF plots between the first row and second row in Fig. 5 shows a significant improvement due to the use of the camera flash. The plots in the first row of Fig. 5 reveal that when the test patches with a flash are matched against reference patches without a flash, the authentication performances are limited. A representative plot such as Fig. 5(b) has an EER of around $10^{-1}$. The plots in the second row of Fig. 5 reveal that when both test and reference patches are captured under camera flash as we proposed, the authentication performances are good and the ambient lighting conditions do not have a major negative effect on the performances. A representative plot such as Fig. 5(f) has an EER of around $10^{-5}$ to $10^{-3}$ (see Table VI for comparison) per our discussion on estimating the EER with different probabilistic models in Section II-C.

Tables II–III present the comprehensive results of various combinations of test and reference databases. Table II reveals that the flash is the dominating factor affecting the authentication performance whereas the condition of the ambient light is not an important factor. Table III reveals that good authentication performance can be achieved across devices of similar imaging modules. The slightly lowered performances for the combinations of iPhone and Canon cameras can be attributed to the different imaging configurations for the two brands of cameras, such as the pattern of the flash, and the relative position of the flash module to the lens.

## V. Proposed Surface Norm Estimation for Paper Authentication Using Mobile Cameras

Although the authentication scheme using flash proposed in Section IV outperforms schemes that flash is not used, the requirement that the test and reference images must be captured at the same position relative to the physical patch is not practical. As mobile cameras have ever-increasing acquisition quality, we ask a research question: Is it possible to use mobile cameras to estimate the physical feature of the paper—the normal vector field—by using multiple images, while solving the issues of camera geometry and lighting? Photometric stereo approaches have long been used to reconstruct 3-D surfaces using photos of surfaces [12]. However, the challenge here is that the scale of the surface of interest in our problem is much smaller. We therefore need to carefully examine the physics of light reflection and arrive at a light reflection model with a proper level of sophistication, in order to obtain meaningful estimates of the normal vector field.

### A. Macroscopic Intensity Due to Camera Flash

Examining the images captured under the flash in Fig. 2(c), one can observe that there exists a mild spatial intensity change across the image. Examining the reflective intensity of a small region under a strong light both by human eyes and from the digital image, one can observe a high spatial frequency fluctuation in addition to the mild intensity change. Given that the light intensity arriving at the paper slowly varies spatially, this fluctuation of the reflective intensity is therefore attributed mainly to the inconsistent orientations of the paper surface at the microscopic level. To reveal the intensity change in fine detail, the mild change at the macroscopic level should be first removed. We define the image intensity at the macroscopic level as the macroscopic intensity, $l^{macro}$.

It can be shown as follows that the macroscopic intensity $l^{macro}$ is proportional to the light strength arriving at the surface, $l$, and cosine of the incident angle, $\theta$. We approximate the macroscopic intensity by the *averaged perceived intensity*

**TABLE I:** Capturing Conditions for Various Databases

| **Database ID** | 502 | 503 | 506 | 508 | 509 | 510 | 501 | 511 | 505 |
|---|---|---|---|---|---|---|---|---|---|
| **Lighting** | flash only | flash + ambient light | | | | | | | ambient light only |
| **Device** | iPhone 6 | | | iPhone 5s | | iPhone 5 | iPhone 6 | Canon SX230HS | iPhone 6 |
| **Room Size** | small | | large | small | large | small | | | large |

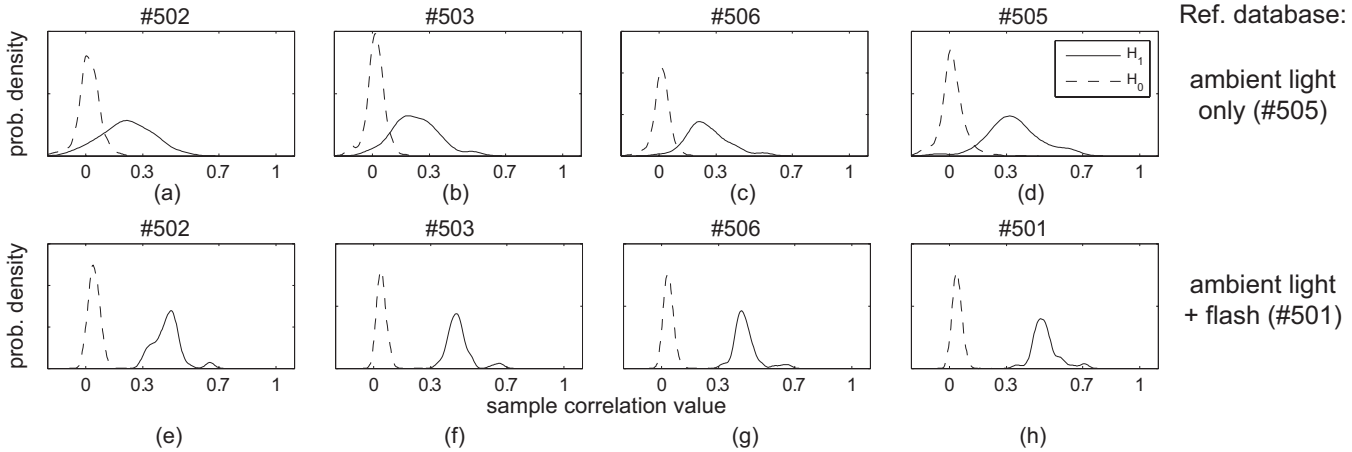


**Fig. 5:** First row: authentication performances of 4 test databases vs. reference database #505 (ambient light only). Second row: performances of 4 test databases vs. reference database #501 (flash + ambient light, proposed). Capturing device: iPhone 6.

**TABLE II:** Modes of estimated PDFs of correlation[1] for matched cases ($H_1$). Contrasting conditions: using flash or not, and small vs. large room size.

| Test \ Ref | With Camera's Built-in Flash | | | No flash |
|---|---|---|---|---|
| | Small Room | | Large Room | |
| | **501** | **502** | **503** | **506** | **505** |
| **501** | *0.48* | *0.44* | *0.43* | *0.40* | 0.21 |
| **502** | *0.45* | *0.45* | *0.47* | *0.40* | 0.22 |
| **503** | *0.43* | *0.47* | *0.48* | *0.41* | 0.19 |
| **506** | *0.42* | *0.45* | *0.42* | *0.43* | 0.21 |
| **507** | *0.40* | *0.41* | *0.40* | *0.40* | 0.23 |
| **505** | 0.21 | 0.24 | 0.24 | 0.24 | 0.32 |

[1] The italic numbers in this table and Table III correspond to the scenarios that estimated PDFs for matched ($H_1$) and unmatched ($H_0$) cases can be perfectly separated.

**TABLE III:** Modes of estimated PDFs of correlation under for matched cases ($H_1$). Contrasting condition: camera model.

| Test \ Ref | iPhone 6 | | iPhone 5s | | iPhone 5 | Canon SX230 |
|---|---|---|---|---|---|---|
| | **503** | **506** | **508** | **509** | **510** | **511** |
| **503** | *0.48* | *0.41* | *0.47* | *0.44* | *0.36* | 0.31 |
| **506** | *0.42* | *0.43* | *0.44* | *0.42* | *0.36* | 0.29 |
| **508** | *0.46* | *0.42* | *0.52* | *0.47* | *0.38* | 0.33 |
| **509** | *0.44* | *0.41* | *0.47* | *0.50* | *0.37* | 0.32 |
| **510** | *0.37* | *0.35* | *0.38* | *0.37* | *0.35* | 0.24 |
| **511** | 0.28 | 0.26 | 0.32 | 0.30 | 0.23 | 0.26 |

$\overline{l_r}$ of background pixels over a small neighborhood $\mathcal{N}$ around a pixel location $\mathbf{p}$:

$$l^{macro}(\mathbf{p}) \approx \overline{l_r}(\mathbf{p}) \tag{2a}$$

$$= \frac{1}{|\mathcal{N}(\mathbf{p})|} \sum_{\mathbf{k}\in\mathcal{N}(\mathbf{p})} \lambda \cdot l(\mathbf{k}) \cdot \mathbf{n}(\mathbf{k})^T \mathbf{v}(\mathbf{k}) \tag{2b}$$

$$\overset{(a)}{\approx} \lambda \cdot l(\mathbf{p}) \cdot \left[\frac{1}{|\mathcal{N}(\mathbf{p})|} \sum \mathbf{n}(\mathbf{k})\right]^T \mathbf{v}(\mathbf{p}) \tag{2c}$$

$$\overset{(b)}{\approx} \lambda \cdot l(\mathbf{p}) \cdot \mathbb{E}\left[\mathbf{n}(\mathbf{p})\right]^T \mathbf{v}(\mathbf{p}) \tag{2d}$$

$$\overset{(c)}{=} \lambda \cdot l(\mathbf{p}) \cdot [0, 0, \mu_{n_z}]\, \mathbf{v}(\mathbf{p}) \tag{2e}$$

$$= \lambda \cdot l(\mathbf{p}) \cdot \mu_{n_z} \cdot \underbrace{v_z(\mathbf{p})}_{\cos\theta \text{ at } \mathbf{p}} \tag{2f}$$

where $|\mathcal{N}(\mathbf{p})|$ is the number of pixels in the small neighborhood of $\mathbf{p}$; step-a follows from the fact that $l(\mathbf{k})$ and $\mathbf{v}(\mathbf{k})$ are approximately constant over the small neighborhood; step-b follows from ergodicity; and step-c follows from the assumption that the normal vectors in the world coordinate system defined in Section II-B are on average pointing straight up, *i.e.*, $\mathbb{E}[n_x] = \mathbb{E}[n_y] = 0$ and $\mathbb{E}[n_z] = \mu_{n_z}$, where $\mu_{n_z}$ is a modeling constant between 0 and 1.

The smooth nature of the macroscopic intensity $l^{macro}$ over the spatial coordinates makes parametric surfaces promising candidate estimators. In this work, we fit a high-order polynomial surface directly to an image captured under flash using an iteratively reweighted least-squares method. The bisquare weights [15] were used to gradually lower the impact of outliers as iteration went on. The original image and its parametrically fitted version are shown in Fig. 6(b) and (c). As our objective is to obtain the macroscopic intensity due to the flash, the image pixels belonging to the registration container and the QR code are considered to be outliers for the surface

fitting purpose. The fitting was excellent with almost no bias. The sample standard deviation, about 2 out of 256 shades of gray, quantifies the magnitude of the fine details of the image appearance of the paper. Fig. 6(d) shows a representative row of pixels (with outliers) and its fitted curve.

One should note that even though a detrended patch image can be obtained by pixel-wise division of macroscopic intensity $l^{macro}$, the detrended patch image is not suitable directly for authentication via correlating with some reference image. After detrending, images for the same patch captured with light/camera located at different relative locations to the physical patch can have different visual appearances at a small scale. This is caused by the different incident light directions with respect to the microscopic surfaces. Fig. 6(e) shows four such detrended patch images when camera locations were at the four corners of the patch. They appear similar at a large scale after detrending but are very different at a small scale due to different incident light directions.

Fig. 7 shows the averaged correlations among the detrended patches as a function of the horizontal and vertical differences in patch locations $(\Delta N_x^{\mathrm{p}}, \Delta N_y^{\mathrm{p}})$, or equivalent in camera capturing locations $(\Delta N_x, \Delta N_y)$. The figure reveals that the farther the capturing distances of cameras for two patches are, the lower the correlation can be for the detrended patch images. This is reasonable as more change in the direction of the incident light leads to more change in the microscopic appearance of the paper surface. This implies that without the proper constraints of the relative position between the camera and the patch, it may not be sensible to verify a paper surface using its detrended image, as the correlation value can be unpredictable and not a single threshold can be selected to determine the authenticity. This observation further justifies the need to use normal vectors for authentication instead of using images directly.

### B. Estimating the Normal Vector Field

In order to solve for the normal vectors $\mathbf{n}(\mathbf{p})$, we combine Eq. (1) characterizing the pixel-wise intensity and Eq. (2) characterizing the macroscopic intensity via canceling their common term $\lambda \cdot l(\mathbf{p})$. One can arrive at the following equality by grouping constants and known terms to the left-hand side:

$$\zeta(\mathbf{p}) \approx \mathbf{n}(\mathbf{p})^T \mathbf{v}(\mathbf{p}) \tag{3}$$

where $\zeta(\mathbf{p}) = \mu_{n_z} v_z(\mathbf{p}) \cdot l_r(\mathbf{p}) \,/\, l^{macro}(\mathbf{p})$ is defined as the normalized intensity and contains the unknown modeling constant $\mu_{n_z}$, the image acquired under flash $l_r$, and the already estimated terms $l^{macro}$ and $v_z$. On the right-hand side, normal vector $\mathbf{n}(\mathbf{p})$ is yet to be solved, and incident light direction $\mathbf{v}(\mathbf{p})$ is known from the previous estimation. The inference problem of the normal vector field can therefore be restructured into a linear regression problem when Eq. (3) is overdetermined.

More specifically, we estimate the normal vectors independently at every pixel location for a total of $200 \times 200$ pixels. For each pixel location $\mathbf{p}$, we set up a system of linear equations using $M = 20$ acquired images, where $M$ is far greater than 4, the number of unknowns:

$$\underbrace{\begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_M \end{bmatrix}}_{\zeta} = \underbrace{\begin{bmatrix} \mathbf{v}_1^T & 1 \\ \mathbf{v}_2^T & 1 \\ \vdots & \vdots \\ \mathbf{v}_M^T & 1 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} n_x \\ n_y \\ n_z \\ b \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}}_{\mathbf{e}}. \tag{4}$$

The unknown parameter $\beta$ contains the normal vector and an intercept $b$ capturing any offset at location $\mathbf{p}$ such as the one indirectly due to ambient light. The observation vector $\zeta$ consists of normalized intensity values at the collocated position $\mathbf{p}$ from images #1 to #$M$. The data matrix $\mathbf{X}$ is composed of vectors of incident directions, and the noise from measurement and/or modeling is modeled by a zero-mean error vector $\mathbf{e}$.

### C. Proposed Method

Fig. 8 is a block diagram for the proposed authentication system. To authenticate a given test surface patch, $M > 4$ photos should be taken under flash. Each photo is processed to extract, warp, and register the captured patch to a grid of 200-by-200 pixels using the registration procedure outlined in Section II-B. The resulting $M$ registered patches with luminance nonuniformity are then processed by the diffuse reflection-based estimator proposed in Section V-B. An estimated normal vector field is therefore obtained and is used as an authentication feature for the surface patch. Its $x$- or $y$-component can be correlated with a reference to determine the authenticity using the hypothesis testing described in Section II-C.

We treat the estimated norm maps from scanners as the reference, since they are reliable as discussed in Section III-A and relatively easy to obtain. More precise estimates of the norm maps can be obtained using microscopes. However, the benefit brought by the microscope with a much more controlled acquisition condition is marginal, and we will keep using norm maps from scanners as a reference.

### D. Experimental Conditions and Results

Fig. 9 illustrates the experimental setup for capturing paper patches for estimating normal vector field using a mobile device. The mobile device was placed on a tripod and adjusted to be in parallel with the surface. The photos captured at the height of about $z = 11.4\,\mathrm{cm}$ contain around 500 pixels along the edge/side for each patch in raw images, which are of enough resolution ($5.25\times$ more pixels than necessary) to generate the registered images of size 200-by-200. Detailed lighting conditions and models of mobile cameras are described individually for each capturing session.

One should note that the exact parallel configuration is not a required condition for our proposed method, and in this exploratory work, the parallel configuration was designed to avoid perspective-related complications. We later conducted additional experiments in which mobile cameras were held by hand and not exactly in parallel with the surface, and the results showed no degradation in authentication performance.

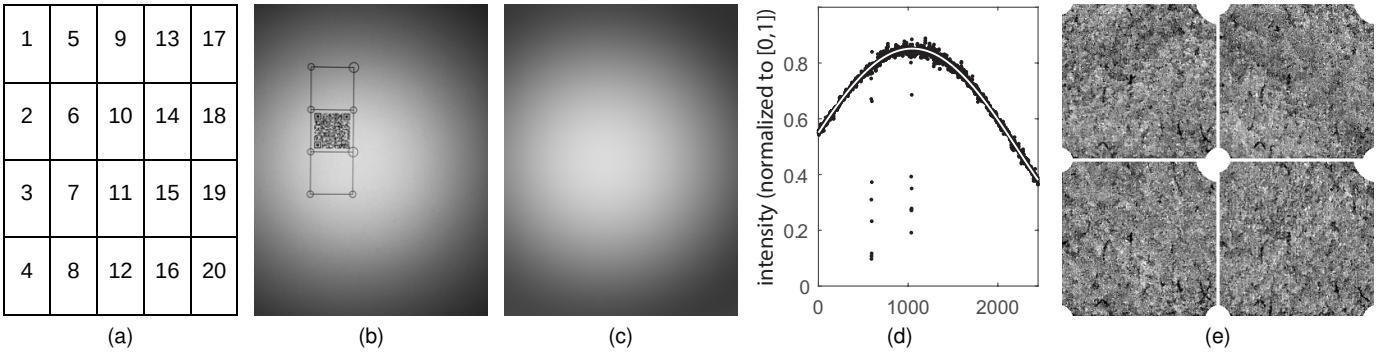


**Fig. 6:** (a) A total of $M = 20$ indexed locations for captured patches in images, (b) a paper patch at location #6 of Session 6, paper #920, (c) its estimated macroscopic intensity image $l^{macro}$ obtained by fitting an order-$(5, 5)$ polynomial surface, (d) a row of intensity values from the middle of the image in (b) and the fitted curve, and (e) detrended images (with contrast enhancement) of Session 6 for the paper patch #920 at locations #1, #17, #20, and #4, respectively, shown clockwise.

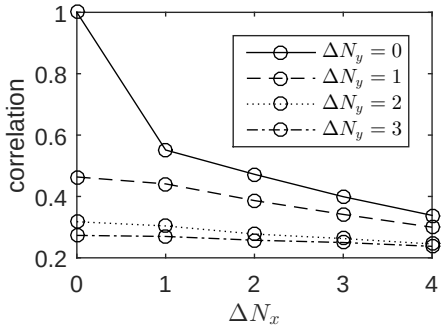


**Fig. 7:** Averaged correlation values of detrended images as a function of the distance between camera capturing locations, namely, $(\Delta N_x, \Delta N_y)$.

**TABLE IV:** Statistics for correlation values of matched cases ($H_1$) for images captured in a completely dark environment (EXP. 1).

| Norm Vector | session 4 | | session 5 | |
|---|---|---|---|---|
| | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\mu}$ | $\hat{\sigma}$ |
| **x-component** | 0.534 | 0.011 | 0.554 | 0.013 |
| **y-component** | 0.523 | 0.012 | 0.493 | 0.015 |

*1) Completely Dark Environment:* Two sessions, namely, Session 4 and Session 5 (*aka* **EXP. 1**), were independently captured using iPhone 6 at the same paper patch in a completely dark environment. Each session contains 20 camera-captured images for the paper patch at 20 different locations indexed in Fig. 6(a).

For each norm map component and each session, we correlate the estimate from the mobile camera with the six estimates from two scanners (three slightly different norm maps for each scanner), and a set of six scores is obtained. A $t$-test is carried out over the group of scores to check if the correlation is significantly greater than $0$.

The results in terms of the sample mean and sample standard deviation are shown in Table IV. It is revealed that either $x$- or $y$-component of Sessions 4 and 5 has a correlation around $0.5$, and the $t$-tests show that all correlation values obtained are statistically significantly ($p$-value $< 10^{-9}$).

*2) Environment with Ambient Light:* We relax the completely dark assumption by investigating more realistic scenarios with the addition of ambient light. Sessions 6–10 (*aka* **EXP. 2**) and Sessions 11–15 (*aka* **EXP. 3**) were captured in a low-strength diffuse ambient light environment using iPhone 6 and iPhone 6s, respectively. Further, Sessions 21–25 (*aka* **EXP. 4**) were captured in an environment with ambient light at the strength of indoor offices using iPhone 6s.

In addition to the result obtained in EXP. 1 that the correlation achieved using norm maps from mobile camera is significantly greater than 0, we would like to further measure quantitatively the discrimination capability that can be achieved in terms of the ROC curve $\big(P_F(\tau), P_M(\tau)\big)$ and/or more compactly, the equal error rate (EER), as outlined in Section II-C.

For the rest of this paper, each session will generate only one correlation value in each normal vector component, and the value is calculated by averaging over the six scores that can be computed from correlating with the slightly different versions of the reference norm maps. This approach is an effort towards reducing the effect of the inaccurate norm map estimates used as references at the service provider side, without adding burden to users during the verification process.

Fig. 10(a) shows the estimated PDFs for the matched ($H_1$) and unmatched ($H_0$) cases for EXP. 2. Under the acquisition condition for EXP. 2, the correlation values do not contain outliers and are distributed around a certain value. We therefore can consider they are sample points drawn from some probability distribution, and we use this modeling assumption to help extrapolate the tails of PDFs and ROC curve. We select Gaussian and Laplace distributions to model the cases in which the true distribution has light versus heavy tails, respectively. Detailed discussion on this modeling can be found in Section VII-E. Using the simple thresholding rule, we draw the ROC curves in Fig. 10(b) and (c). The discrimination capability measured in EER is $10^{-156}$ by assuming the correlation is Gaussian distributed and $10^{-16}$ by assuming the correlation is Laplacian distributed.

We list in Table V the discrimination capability for EXP. 2–EXP. 4 measured in EERs, and the detailed statistics that the EERs are calculated from, namely, the sample mean, $\hat{\mu}_i$,

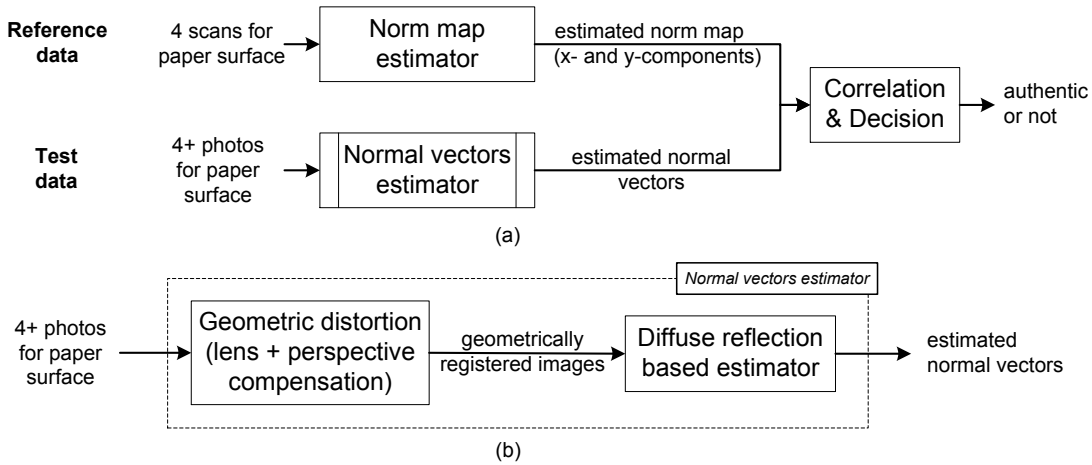


**Fig. 8:** (a) Block diagram for the proposed authentication system using mobile camera "captured" surface normal vectors. (b) Sub-diagram for the *normal vectors estimator* in (a).

**Fig. 9:** Setup for experiments conducted in Section V.

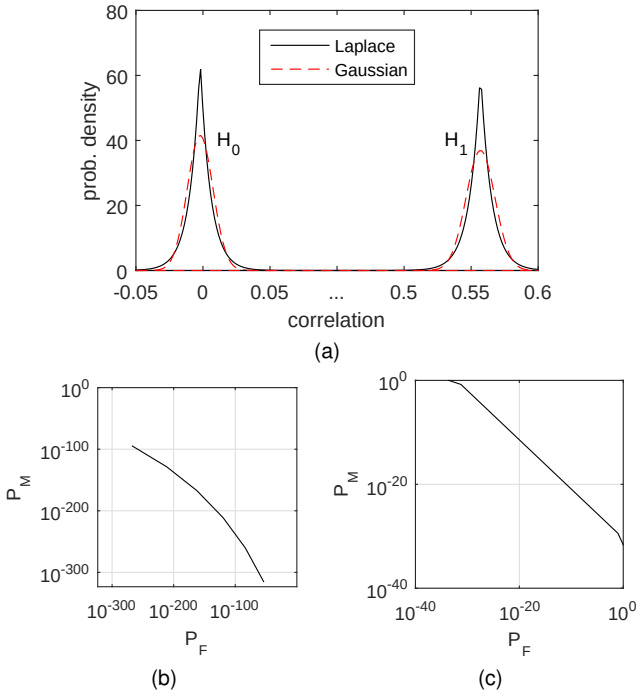


**Fig. 10:** (a) Estimated PDFs of correlation for EXP. 2 (Sessions 6–10 & iPhone 6), (b) ROC curve by assuming correlation is Gaussian distributed, (c) ROC curve by assuming correlation is Laplace distributed.

**TABLE V:** Discrimination capability in EERs and corresponding statistics for images captured in environments with ambient light.

| | Match ($H_1$) | | | No match ($H_0$) | | | EER | |
|---|---|---|---|---|---|---|---|---|
| | $\hat{\mu}_1$ | $\hat{\sigma}_1$ | $\hat{\lambda}_1$ | $\hat{\mu}_0$ | $\hat{\sigma}_0$ | $\hat{\lambda}_0$ | Gau. | Lap. |
| **EXP. 2** | 0.557 | 0.011 | 122.8 | −0.002 | 0.010 | 129.5 | $10^{-156}$ | $10^{-16}$ |
| **EXP. 3** | 0.532 | 0.015 | 97.4 | −0.004 | 0.011 | 113.4 | $10^{-94}$ | $10^{-13}$ |
| **EXP. 4** | 0.528 | 0.012 | 106.4 | −0.004 | 0.012 | 103.9 | $10^{-109}$ | $10^{-13}$ |

the sample standard deviation, $\hat{\sigma}_i$, the maximum-likelihood estimates for the rate parameter of the Laplacian distribution, $\hat{\lambda}_i$. As revealed by Table V, the authentication performances are similar with different strength levels in ambient lighting and with different capturing devices (iPhones 6 and 6s). The high authentication accuracy and the flexible image acquisition procedure make the proposed method a promising technology to be deployed in a practical working environment. In addition to the above authentication performances that are measured for one acquisition condition per experiment, it is also beneficial in future work to measure the performance in a single experiment containing a variety of practical acquisition conditions.

*3) Mobile Cameras of Other Brands:* The investigation in this section has mainly used the cameras of the iPhone series for exploring the possibility of estimating the normal vector field of a paper surface. We also carried out experiments using mobile cameras of other brands such as Samsung Galaxy Alpha. After obtaining the estimated normal vector field, we correlated the $x$- or $y$-component with the reference norm maps provided by scanners. The sample mean of correlation for matched cases ($H_1$) is around $0.23$ with similar sample variance as in the experiments for iPhones. The smaller mean value compared to that of the iPhone cameras, $0.53$, may be due to the fact that the flash of Samsung Galaxy Alpha is not so bright as those of iPhone cameras. The authentication performance measured in EER ranges from $10^{-22}$ to $10^{-6}$. The EER results suggest satisfactory performances by the proposed method, and an effective decision strategy is to adjust the decision thresholds differently for Samsung Galaxy Alpha and for iPhone cameras considering their different PDFs of correlation under $H_1$.

### E. Comparison with Prior Work

In Table VI, we summarize the performances of the proposed methods and prior art as discussed in the previous sections of the paper. Our proposed image-based method using a mobile camera with flash has similar performance as the work in [6] that uses an industrial camera and a semi-controlled light, as we created a semi-controlled light using flash and captured the test and reference images at the same position relative to the physical patch. The proposed image-based method outperforms the method in [8] that uses robust point features, suggesting that a favorable lighting condition is a more important factor than a robust image processing technique.

Inspired by the success of various methods designed to use controlled lighting (such as the method proposed in Section IV and the work in [6]) or inherently used controlled lighting [5], we have explored a semi-controlled lighting condition with the help of the flash of mobile cameras. The proposed norm map-based method significantly outperforms all image appearance-based methods. Although it performs slightly worse than the case using the scanner as the acquisition device [5], the flexibility of the mobile device modality makes the proposed method more practical for ubiquitous deployment such as counterfeiting detection by end consumers.

## VI. Perturbation Analysis on Discriminability

This section analyzes the performance of the method proposed in the previous section under perturbations. We do not consider controllable factors that can potentially be taken care of at the service provider side, such as the number of norm maps used as references, and whether or not lens distortion should be compensated on the query images. Instead, we focus on the factors that are uncontrollable, such as the inaccuracy of the estimated camera locations, and the factors that may increase the burden on the users in the verification process, such as the number of flash images that users need to shoot in each session of verification.

### A. Precision of Estimated Lens Location

The incident light direction $\mathbf{v}_i$ in Eq. (4) is an essential quantity for estimating the normal vector field. Its value is directly related to the camera location that may be imprecisely estimated. In this part, we first quantify the inaccuracy of the camera location estimate, and then perturb the camera location when calculating $\mathbf{v}_i$ to examine how the authentication performance will be affected.

**Inaccurate Camera Location** The standard deviation of location offset indicates how far away the estimated camera locations are from the true locations in a statistical sense. The true locations of the camera/lens were manually recorded while Sessions 4–10 were captured, each containing 20 images. Together with estimated locations calculated from the projection matrix (connecting the world and image coordinate systems), the 3-D location offset was obtained. For each image, the location offset is a vector containing quantities in $x$-, $y$-, and $z$-directions. The standard deviation for $x$-, $y$-, and $z$-directions were $1.86\,\text{mm}$, $2.16\,\text{mm}$, and $0.84\,\text{mm}$, respectively, when the camera was placed at the height of about $z = 11.4\,\text{cm}$.

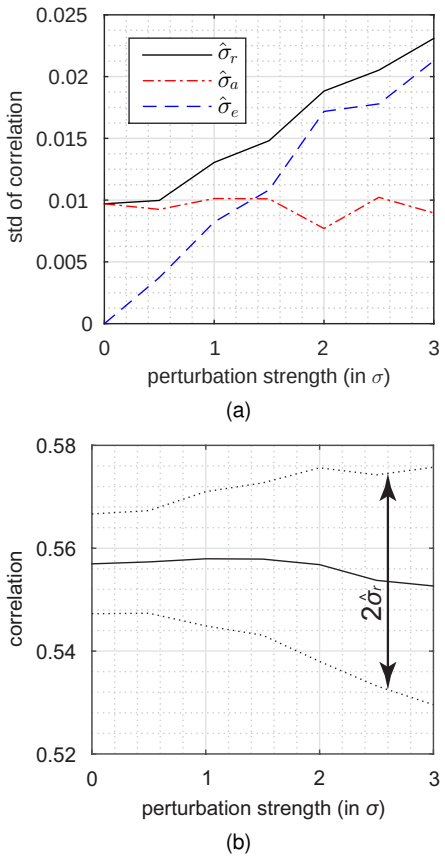


**Fig. 11:** (a) Estimated standard deviation of the correlation $\hat{\sigma}_r$ and its decomposition $\hat{\sigma}_a$ and $\hat{\sigma}_e$ for correct matches as a function of inaccuracy of lens location estimate, and (b) estimated mean correlation with two-sigma wide performance region, $(\hat{\mu}-\hat{\sigma}_r, \hat{\mu}+\hat{\sigma}_r)$.

**Performance Drop Under Perturbation** With the knowledge of the amount of inaccuracy of camera location estimates, we can examine how authentication performance will be affected by adding a reasonable amount of perturbation. We chose $\sigma_x = 2\,\text{mm}$, $\sigma_y = 2\,\text{mm}$, and $\sigma_z = 0.9\,\text{mm}$ as the unit standard deviation in each direction, and scaled them by a list of scalars $[0, 0.5, 1, 1.5, 2, 2.5, 3]$, in which "1" corresponds to the nominal strength we obtained above. The larger the scalar is, the stronger the perturbation will be added.

For each perturbation level, a self-contained sub-experiment was carried out. The sub-experiment was carried out using the images from Sessions 6–10. For each session, the estimated camera location would be biased for 20 times by different location offset vectors that were independently drawn from the distribution of the current perturbation level. The resulting $5 \times 20$ correlation values are expected to have an increased variance due to the additional perturbation.

We analyzed the results of the sub-experiments using a *random effect model* [16] in order to reveal quantitatively the effect on the correlation value due to the additional perturbation and different sessions. The correlation $r_{ij}$ obtained in

**TABLE VI:** Authentication Performances of the Proposed Methods and Prior Art

| Feature | | Modality | Lighting | Flexibility | Performance |
|---|---|---|---|---|---|
| **Type** | **Detail** | | | | **EER** |
| Image | pixel value | Industrial camera, Voloshynovskiy *et al.* [6] | semi-controlled | no | $10^{-4}$ |
| | SIFT descriptor | Mobile camera, Diephuis *et al.* [8] | uncontrolled | yes | $10^{-2}$ |
| | **pixel value** | **Mobile camera** (proposed in Section IV) | **semi-controlled** | **no** | $\mathbf{10^{-5}}$ **to** $\mathbf{10^{-3}}$ |
| Norm map | seeded hash | Scanner, Clarkson *et al.* [5] | fully controlled | no | $10^{-130}$ to $10^{-15}$ |
| | **surface normal direction** | **Mobile camera** (proposed Section V) | **semi-controlled** | **yes** | $\mathbf{10^{-109}}$ **to** $\mathbf{10^{-13}}$ |

the $j$th random trial of the $i$th session of the sub-experiment is assumed to be a summation of the mean correlation value $\mu$ (an unknown but fixed parameter), the zero-mean random effect $a_i$ of the $i$th session, and the remaining error $e_{ij}$ at the perturbation strength level of the current sub-experiment:

$$r_{ij} = \mu + a_i + e_{ij}, \qquad \begin{matrix} i = 6, \cdots, 10, \\ j = 1, \cdots, 20, \end{matrix} \tag{5}$$

where $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, and they are assumed to be jointly independent. Note that the variance of $r_{ij}$ is composed of those of $a_i$ and $e_{ij}$, namely, $\sigma_r^2 = \sigma_a^2 + \sigma_e^2$. We are interested in the values of the modeling parameters $\mu$, $\sigma_a$, and $\sigma_e$, and the analytical expressions of the maximum-likelihood estimators are discussed in textbooks on the random effect model [16].

We obtained a distinct set of parameter estimates for each sub-experiment corresponding to a certain perturbation level, and plotted them with respect to the perturbation level accordingly. Fig. 11(a) shows a near-constant effect (quantified by $\hat{\sigma}_a$) for the session, and an increased effect of perturbation (quantified by $\hat{\sigma}_e$) as the perturbation level increases. Fig. 11(b) shows the mean correlation against the perturbation level with a two-sigma-wide performance region. Both figures reveal that the resulting increase of perturbation is small compared to the mean correlation value, with $\hat{\sigma}_r < 0.014$ when the perturbation strength is at the nominal level, and $\hat{\sigma}_r < 0.024$ even when the perturbation strength is 3 times of the nominal level.

### *B. Number of Images for Normal Vector Field Estimation*

We now consider the effect of the number of available images on the estimation of the normal vector. Recall in the main experiment, we used $M = 20$ images to estimate the normal vector field, which may not be very user-friendly with a large number of light flashes in a short period of time. We varied the number of images from $M = 20$ down to $M = 4$, and carried out a self-contained sub-experiment similar to those in the last subsection using the images from Sessions 6–10. For each session, 20 subsets of the available images were selected and their correlation values were examined at the current perturbation level. The resulting $5 \times 20$ correlation values are expected to have an increased variance due to fewer images used.

Regarding the selection of the subsets of images, one should identify whether the available image set contains extremely "bad" ones towards the estimation of the normal vector and correlation, which should be carefully handled in the selection process. We tried to identify "bad" images using the following two criteria: i) the fitting error in the model of Eq. (4), and ii) the correlation improvement when excluding an image. No "bad" image was identified out of the 20 available images, and we therefore constructed subsets of images by uniformly random selections from the indices $1, \cdots, 20$.

We analyzed the results of the sub-experiments using the same random effect model as in the last subsection to reveal quantitatively the effect of having fewer images for the normal field estimation. We obtained a distinct set of parameter estimates for each sub-experiment corresponding to a certain number of images, and plotted them accordingly. Fig. 12(a) shows, as expected, a constant effect for the session, and an

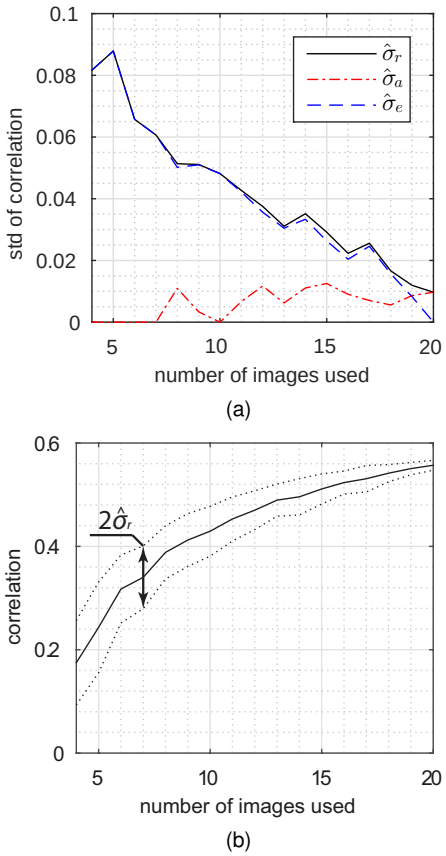


**Fig. 12:** (a) Estimated standard deviation of the correlation $\hat{\sigma}_r$ and its decomposition $\hat{\sigma}_a$ and $\hat{\sigma}_e$ for correct matches as a function of number of images for norm map estimation, and (b) estimated mean correlation with two-sigma wide performance region, $(\hat{\mu}-\hat{\sigma}_r, \hat{\mu}+\hat{\sigma}_r)$.

increased effect of perturbation as fewer images were used. The value of $\hat{\sigma}_r$ reaches almost $0.1$ when the number of images used is reduced to $4$. Fig. 12(b) shows that the mean correlation can drop to below $0.2$ and the rate of the drop accelerates as the number of images reduces. Both figures reveal that the number of images used for norm map estimation can significantly affect the correlation.

### C. Perturbation Factors Combined

The perturbation analyses in the above two subsections reveal that the number of images used for norm map estimation dominates the correlation value, compared to other factors such as the capturing session setup and the accuracy of the estimated camera location.

We now evaluate the discrimination capability in terms of EER by considering all possible factors investigated above. The EER will be plotted against the dominant factor, namely, the number of images. Such other remaining factors as the session setup, and the inaccuracy of the estimated camera location will be taken into consideration by boosting the overall variance in their respective amount estimated earlier in this paper.

The two plots in Fig. 13 show the EER as decreasing functions of the number of images under Gaussian and Laplacian models, respectively. The results show that in order to obtain an EER of $10^{-4}$, one should on average acquire at least $6$ flash images if the correlation follows a light-tailed Gaussian distribution. In contrast, if the correlation follows a heavy-tailed Laplacian distribution, one should on average acquire at least $8$ flash images. More discussions on modeling the PDFs of correlation using the Gaussian versus the Laplacian can be found in Section VII-E.

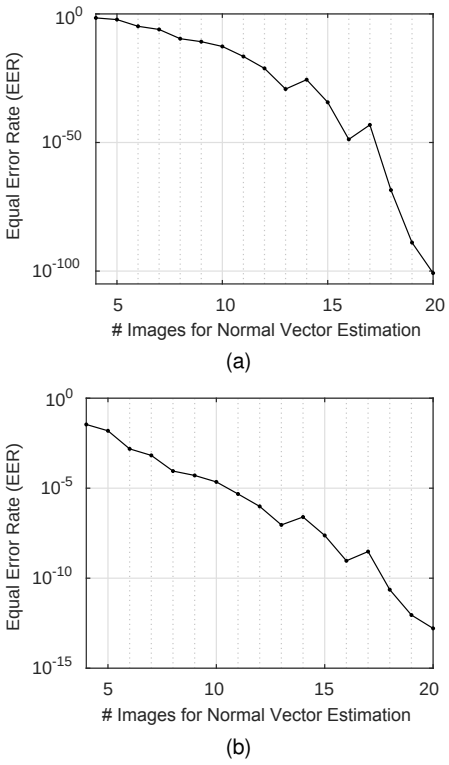


**Fig. 13:** Discrimination capability in terms of EER taken into consideration of all factors as a function of number of images for (a) Gaussian, and (b) Laplacian distributed correlation values.

## VII. Discussions

### A. Interpretation of Norm Map Obtained From Low Resolution Images

When a camera's capturing resolution is high enough, the area covered by each pixel is relatively flat, and the normal vector assigned to the pixel represents the physical surface direction of the area. The collection of the normal vectors therefore serves as a fingerprint for the paper surface.

When the resolution is lower than the aforementioned scenario, however, is the normal vector still a meaningful quantity? Let us relate a high-resolution image and its low-resolution version by a virtual 2-D low-pass filter with coefficients $\{w_i > 0 \,|\, \sum_{i=1}^{N} w_i = 1\}$, where $i$ is a location index linearized from a 2-D index pair and $N$ is the number of pixels covered by the filter. A pixel value $u$ in the low-resolution image is therefore the weighted sum of $N$ pixels each with intensity $\mathbf{n}_i^T \mathbf{v}_i$ of the high-resolution image, where $\mathbf{v}_i$ and $\mathbf{n}_i$ are the directions of the incident light and normal vector at the location with index $i$, respectively. Hence,

$$u = \sum_{i=1}^{N} w_i \cdot \mathbf{n}_i^T \mathbf{v}_i \approx \left( \sum_{i=1}^{N} w_i \mathbf{n}_i \right)^T \mathbf{v} = \bar{\mathbf{n}}^T \mathbf{v} \qquad (6)$$

where $\mathbf{v}$ is the direction of the incident light for the pixel in the low-resolution image. The approximation can be justified because, in a small neighborhood, the direction of incident light is almost constant, *i.e.*, $\mathbf{v} \approx \mathbf{v}_i$. The term enclosed in the parentheses immediately on the RHS of the approximation sign can be regarded as the reflected intensity in a larger area with an averaged direction $\bar{\mathbf{n}} = \sum_{i=1}^{N} w_i \mathbf{n}_i$. That is, the norm maps estimated from low-resolution images can be considered as a downsampled norm map using the virtual filter $\{w_i\}$ that relates the high and low-resolution images.

### B. Effect of Motion Blur

Slight panning motion during the capturing process often results in blurred images. The effect of the panning motion can be modeled by a linear-spatial invariant filter. In a special case where the motion blur is the same for all images captured, the normal vector field will be blurred by the same filter of the motion blur per the propagation property discussed above in Section VII-A. A blurred normal vector field may lead to a lower verification rate. It is interesting to study how fast the authentication performance will drop as the strength of the motion blur increases. In a general case where the motion blur is not consistent for all images captured, the lowpass filtering effect does not directly propagate to the normal vector field. In this case, a study on how the motion blur will change the normal vector field and its ultimate impact can be carried out. If motion blur turns out to be a major factor in lowering the authentication performance, one can consider applying blind deconvolution in the first place for deblurring the images before using them for authentication purposes.

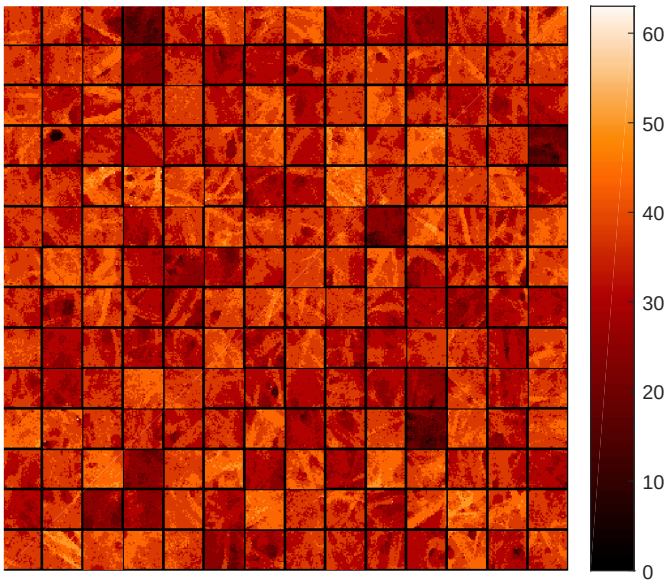


**Fig. 14:** Sample collection of topographic blocks from copy paper of size $85\,\mu$m-by-$85\,\mu$m by confocal microscope. The depth unit on color bar is also $\mu$m.

### C. Understanding the Physics of Paper Surface Reflection

In this subsection, we use a confocal microscope to obtain the 3-D structure of a paper surface as a topographic map. This "ground-truth" map helps us examine the linkage between the reflected image appearance and the physical structure of the paper surface.

**Normal Vectors From Confocal Microscope** We use a Leica confocal microscope (under the reflection imaging mode using 488 nm laser light) to obtain a topographic map with a per-sample resolution of $3\,\mu$m, $3\,\mu$m, and $5.7\,\mu$m in $x$-, $y$- and $z$-directions, respectively. For the square surface patches of edge length $2/3$ inch digitized to 200 pixels (*aka* working pixels), the area covered by each pixel contains about 796 pixels in the topographic map (*aka* confocal pixels).

We estimate the normal direction for each working pixel described as follows. Fig. 14 is a sample collection of topographic blocks with each showing the area covered by one working pixel. Our examination of the whole set reveals that most blocks were not flat because the scale of fibers is smaller than the area of a working pixel. Using the result from Section VII-A, we calculate a surface direction for each working pixel area by weighted averaging over the directions of all confocal pixels. Alternatively, we fit a plane to all confocal pixel locations and use the direction of the plane as an alternative estimate. These two estimates for the surface direction agree with each other with a correlation of 0.98, implying that the physical normal vectors are not sensitive to different definitions of direction and estimation algorithms, and are therefore reliable. Hence, the plane estimates are sufficiently good estimates for normal directions, and are considered as the physical ground truth in the experiments followed.

We examine the correlation of norm maps obtained from mobile cameras and scanners with respect to the ground truth. The results are shown in Table VII. The nonzero correlation values imply that norm maps estimated by the scanner and mobile camera are indeed related to the ground truth, *i.e.*, the physical norm map obtained by the confocal microscope. However, the correlation values with the ground truth are low, around 0.2–0.3. This suggests that although the fully diffuse reflection model provides a surface norm estimation that is sufficiently discriminative for authentication purposes, the estimation may not be highly precise at the accuracy level of confocal microscopes.

**TABLE VII:** Correlation of Norm Maps with Ground Truth

| Pair of Quantities | Correlation |
|---|---|
| Mobile camera *vs.* confocal (ground truth) | 0.19 |
| Scanner *vs.* confocal (ground truth) | 0.28 |
| [Reference]: Mobile camera *vs.* scanner | 0.59 |

**Dominant Reflection Type** In this part, we study the relative contributions of diffuse and specular components, with the help of the physical normal vectors from a confocal microscope. We first calculate a synthesized diffuse image $\mathrm{Im_d}(\mathbf{p}) = \max\{0, \mathbf{n}(\mathbf{p})^T\mathbf{v}_i(\mathbf{p})\}$ and specular image $\mathrm{Im_s}(\mathbf{p}) = \max\{0, \mathbf{v}_c(\mathbf{p})^T\mathbf{v}_r(\mathbf{p})\}$ using known quantities, without including the common effect of the light strength at location $\mathbf{p}$. Here, $\mathbf{n}$ is the surface normal vector, $\mathbf{v}_i$ is the incident light vector, $\mathbf{v}_c$ is the camera direction vector, and $\mathbf{v}_r$ is the specular reflection vector that can be represented as $\mathbf{v}_r = (2\mathbf{n}\mathbf{n}^T - \mathbf{I})\mathbf{v}_i$. We then regress 20 camera captured images $l_r(\mathbf{p})$ against the diffuse image and specular image, in order to obtain non-negative weights for diffuse and specular images scaled by the light strength, $\{w_d \cdot l^{(k)}(\mathbf{p}),\ k = 1, \cdots, 20,\ \forall \mathbf{p}\}$ and $\{w_s \cdot l^{(k)}(\mathbf{p}),\ k = 1, \cdots, 20,\ \forall \mathbf{p}\}$, respectively. Using non-negative matrix factorization with rank 1, we obtain estimates for $w_d$ and $w_s$ up to a multiplicative scalar. The ratio between the contributions of diffuse and specular components, $w_d/w_s$, is 5.82. This high ratio of nearly six-to-one reaffirms the diffuse reflection model in this paper, and explains the excellent authentication performance in prior work [4], [5].

**Generative Modeling** Using the relative weights of diffuse and specular components in the reflection model of paper, we synthesize reflection images and examine their relationship with the camera-captured images. We again use correlation as the similarity measure, and carefully remove the macroscopic trend of spatial intensity in order to avoid correlation inflation.

For 20 pairs of synthetic and camera-captured images, we observe a statistically significant correlation of 0.13. This result from generative modeling and the result of Table VII from discriminative modeling show that it is possible to connect the physical normal vectors to the surface appearance. Possible future directions to improve such connection include examining the role of diffraction, as well as the roles of transmitted and re-emitted light, as paper is not always fully opaque.

### D. Robustness of the Norm Map

Practical deployment of the proposed scheme requires understanding the robustness of the norm map under various conditions, and designing adaptive authentication algorithms when necessary. Clarkson *et al.* [5] conducted tests on scribbling with a pen and printing single-spaced text on around $10\%$ of the test regions. They also tested the water treatment by using paper dried and ironed after being submerged. Their experiments demonstrated that the norm map is a robust feature under these conditions.

More investigations should be carried out on the resilience against tampering of the paper surface, such as scratching, folding, and crumpling, and on the reliability of the physical structure over time. Large-scale tests for papers from different manufacturers are beneficial to understanding issues that may arise in the deployment of the proposed technology. Below we discuss how to achieve resilience against the folding operation.

**Resilience Against Folding** Paper can be easily folded, resulting in a change of directions of those surfaces around the fold lines. In order to maintain a high correlation for true matches, the following strategies can be applied. The first strategy masks in correlation calculation those pixels whose surface directions are affected by folding. This method is intuitive but relies on the detection and segmentation of folded regions. As the distortion to the norm map field due to folding can be viewed as the addition of a slowly spatially varying trending surface, the second strategy is to apply detrending methods before calculating the correlation. For example, highpass filtering can be applied to remove the global trend. Such a highpass filter should be designed to properly reject the frequency components of the trending surface. Alternatively, parametric surfaces can be fitted to estimate the trending surface, and the resulting residue can be used to perform correlation. A practical challenge lies in the selection of a parametric surface that neither overfits nor underfits.

### E. Considerations for Using Statistical Methods for Inference

In practice, the theoretical PDFs, $f_{\hat{\rho}|H_0}(\hat{\rho})$ and $f_{\hat{\rho}|H_1}(\hat{\rho})$, as well as the performance metrics, $P_D(\tau)$, $P_F(\tau)$, ROC, and EER derived from them, are not known *a priori*, and need to be estimated from the practical data obtained from experiments. One can construct normalized histograms or empirical probability mass functions (PMFs) for $H_0$ and $H_1$ as estimates for the true PDFs, and use the resulting PMFs to calculate the performance metrics. This approach using the real data can give estimates that are close to the true PDFs especially when the sample size is large. However, using PMFs as the estimates for PDFs leads to piecewise ROC curves and imprecise EER estimates, and the lack of distribution data in tails requires an extremely large sample size to reveal the true performances around the two tail regions of the ROC curve. For example, for a $50$-image dataset containing $\binom{50}{2} = 1,225$ possible pairs of images for verification, the smallest possible estimates for $P_M$ and $P_F$ are $1/50$ and $1/1225$, respectively, which may not precisely reflect the performance of the system if the achievable rates are much smaller than $1/1225$.

To alleviate the limitations of using the empirical PMFs for inferences, we can incorporate more modeling flavors by assuming the theoretical PDFs $f_{\hat{\rho}|H_0}(\hat{\rho})$ and $f_{\hat{\rho}|H_1}(\hat{\rho})$ follow some commonly seen distributions such as Gaussian and Laplacian. Adding this additional assumption has the advantage that the tails of PDFs and ROC can be better extrapolated, and EER can be calculated as a deterministic function of moments such as the mean and the variance. One should note that the accuracy of the extrapolated tails depends heavily on the assumption that the data would match with the assumed distribution. As sample points for tails are usually lacking for samples of small size, it is reasonable to try a heavy-tailed distribution (such as Laplacian) to infer the lower performance bound, and to try a light-tailed distribution (such as Gaussian) to infer the upper performance bound. As can be recalled, we have calculated in Section V-D such performance bounds for both ROC curves and EERs.

## VIII. Conclusion

In this paper, we have investigated intrinsic microscopic features of the paper surface for authentication purposes. We have shown that it is possible to use the cameras and built-in flash of mobile devices to estimate the normal vector field of paper surfaces. Perturbation analysis shows that the proposed method is robust against inaccurate estimates of camera locations, and using $6$ to $8$ images can achieve a matching accuracy of $10^{-4}$ in EER under a lab-controlled ambient light environment. This finding can relax the restricted imaging setup in prior art, and enable paper authentication under a more casual, ubiquitous setting with a mobile imaging device. The proposed technique may facilitate duplicate detection of important and/or valuable documents such as IDs, and facilitate counterfeit mitigation of merchandise via detection of duplicated labels and packages.

**Acknowledgment** The work was carried out in collaboration with AiDiXing Inc., Beijing, China. The authors thank Ms. Amy Beaven from Imaging Core of the Department of Cell Biology and Molecular Genetics of the University of Maryland for scanning the paper surfaces using the Leica SP5 X confocal microscope, Dr. King Lam Hui for providing expertise in microscopic imaging and taking photos for paper samples using a DSLR camera, and Mr. Zhili Yang for taking photos for paper samples using a confocal camera. We thank the reviewers for their constructive comments.

## References

[1] C.-W. Wong and M. Wu, "Counterfeit detection using paper PUF and mobile cameras," in *Proc. IEEE International Workshop on Information Forensics and Security (WIFS)*, Rome, Italy, Nov. 2015.

[2] Product Overview on BubbleTag™, Ramdot™, FiberTag™, *Prooftag SAS*, Retrieved Jan. 2015. http://www.prooftag.net/

[3] Kinde Anti-Counterfeiting Labels, *Guangdong Zhengdi (Kinde) Network Technology Co., Ltd.*, Retrieved Jan. 2015. http://www.kd315.com/

[4] C.-W. Wong and M. Wu, "A study on PUF characteristics for counterfeiting detection," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Quebec City, Canada, Sep. 2015, pp. 1643–1647.

[5] W. Clarkson, T. Weyrich, A. Finkelstein, N. Heninger, J. Halderman, and E. Felten, "Fingerprinting blank paper using commodity scanners," in *Proc. IEEE Symposium on Security and Privacy*, Berkeley, CA, May 2009, pp. 301–314.
[6] S. Voloshynovskiy, M. Diephuis, F. Beekhof, O. Koval, and B. Keel, "Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (FAMOS)," in *Proc. IEEE International Workshop on Information Forensics and Security (WIFS)*, Tenerife, Spain, Dec. 2012, pp. 43–48.
[7] M. Diephuis and S. Voloshynovskiy, "Physical object identification based on FAMOS microstructure fingerprinting: Comparison of templates versus invariant features," in *Proc. International Symposium on Image and Signal Processing and Analysis (ISPA)*, Trieste, Italy, Sep. 2013, pp. 119–123.
[8] M. Diephuis, S. Voloshynovskiy, T. Holotyak, N. Stendardo, and B. Keel, "A framework for fast and secure packaging identification on mobile phones," in *Proc. SPIE, Media Watermarking, Security, and Forensics*, San Francisco, CA, Feb. 2014, p. 90280T.
[9] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 19, no. 4, pp. 210–218, 1986.
[10] "High resolution surface topography FRT MicroProf chromatic aberration sensor," Aug. 2012, a product sheet by Innventia AB.
[11] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: Physical and geometrical perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 611–634, Jul. 1991.
[12] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.
[13] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008.
[14] N. Poh and S. Bengio, "How do correlation and variance of base-experts affect fusion in biometric authentication tasks?" *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4384–4396, Nov. 2005.
[15] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiley, 2009.
[16] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*. Wiley, 2001.

**Chau-Wai Wong** (S'05–M'16) received the B.Eng. degree with first class honors in 2008, and the M.Phil. degree in 2010, both in electronic and information engineering from The Hong Kong Polytechnic University (PolyU). He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering at the University of Maryland, College Park. His research interests include multimedia forensics, statistical signal processing, data analytics, and video coding.

Mr. Wong was the recipient of multiple scholarships and awards, including Top 4 Student Paper Award, Future Faculty Fellowship, HSBC Scholarship, Hitachi Scholarship, Chinese Government Award for Outstanding Students Abroad, two-year and four-year merit-based full scholarships for M.Phil. and B.Eng. studies, respectively. He is a member of the IEEE and the APSIPA, and he was the general secretary of the IEEE PolyU Student Branch from 2006 to 2007. He was involved in organizing the third edition of the IEEE Signal Processing Cup in 2016 on electric network frequency forensics.

**Min Wu** (S'95–M'01–SM'06–F'11) received the B.E. degree (Highest Honors) in electrical engineering and the B.A. degree (Highest Honors) in economics from Tsinghua University, Beijing, China, in 1996, and the Ph.D. degree in electrical engineering from Princeton University in 2001.

Since 2001, she has been with the University of Maryland, College Park, where she is currently a Professor and a University Distinguished Scholar-Teacher. She leads the Media and Security Team (MAST), University of Maryland, where she is involved in information security and forensics and multimedia signal processing. She has coauthored two books and holds nine U.S. patents on multimedia security and communications.

Dr. Wu coauthored several papers that won awards from the IEEE, ACM, and EURASIP, respectively. She also received an NSF CAREER award in 2002, a TR100 Young Innovator Award from the MIT Technology Review Magazine in 2004, an ONR Young Investigator Award in 2005, a ComputerWorld "40 Under 40 IT" Innovator Award in 2007, an IEEE Mac Van Valkenburg Early Career Teaching Award in 2009, and University of Maryland Invention of the Year Award in 2012 and 2015. She has served as Vice President-Finance of the IEEE Signal Processing Society from 2010 to 2012, and Chair of the IEEE Technical Committee on Information Forensics and Security from 2012 to 2013, and an IEEE Distinguished Lecturer in 2015 to 2016. She is currently Editor-in-Chief of the *IEEE Signal Processing Magazine*. She is an IEEE Fellow for contributions to multimedia security and forensics.