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 used the normal vector field, a unique, microscopic, physically unclonable feature of paper surfaces, estimated by consumer grade scanners. Industrial cameras was also used to capture the appearance of the surface rendered after the normal vector field per the law 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 are good for authentication when the camera flashlight is exploited for creating a semi-controlled lighting condition. We have proposed new algorithms to demonstrate that the normal vector field of 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 (EER) can be achieved using 6 to 8 images under a lab-controlled ambient light environment. Our findings can relax the restricted imaging setups to 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 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 flashlight 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 flashlights 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. Particularly in this journal version, more 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 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 the a more flexible setup—mobile cameras with built-in flashlights, and compare the performances with prior work. In Section VI, we conduct perturbation analysis to demonstrate the practicality of the proposed mobile cameras based authentication method. In Section VII, we use confocal microscopic data from a physics aspect to elucidate a deeper understanding in 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 one location to another on the same paper, and therefore can serve as a unique identifier or fingerprint. One of the quantifiers for such 3-D structure, namely, 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 eye/sensor. Under the diffuse reflection model, the perceived intensity is dependent on the direction of incident light and 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 stick to the fully diffuse modeling assumption, and provide an experimental justification in the discussion section that the strengths of 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 \( p \) 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

\[
l_r(p) = \lambda \cdot l(p) \cdot \mathbf{n}(p)^T \mathbf{v}(p),
\]

which depends on the angle \( \varphi(p) \) between 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, \( l(p) \); and the albedo, \( \lambda \), characterizing the physical capability of reflecting the light [11], [12]. In our work, assuming \( \lambda \) to be constant over the whole paper patch is found to hold for the purpose of authentication.

For an ideal point light source, light strength \( l(p) \) over a spatial field is modeled by considering the effect of energy fall-off due to travel distance of the light [12]. In practice, flashlight is not a perfect point source but has a finite dimension, e.g., in a disc-like shape. When the flashlight is not perfectly oriented towards 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 \( l(p) \) with a high precision for nonideal point sources. Instead, we estimate \( l(p) \) by exploiting its spatial smoothness property. With the values of \( l(p) \), the microscopic structure can then be exactly determined in terms of normal vectors, \( \mathbf{n}(p) \).

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 precise registration.
in experiments. Considering a printing resolution of 600 pixels per inch, each square container of size \( \frac{2}{3} \times \frac{2}{3} \text{inch}^2 \) (1.69-by-1.69 cm²) 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 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 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 powerful 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 \( \tilde{\rho} \) 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) \( f_{\tilde{\rho}|H_0}(\tilde{\rho}) \) and \( f_{\tilde{\rho}|H_1}(\tilde{\rho}) \) that have very distinct mean values under nonmatched and matched cases, and make a decision using the simple thresholding rule on an observed value of the random variable \( \tilde{\rho} \).

Under the simple thresholding rule with threshold \( \tau \), the detection rate is defined to be \( P_D(\tau) = \int_0^\tau f_{\tilde{\rho}|H_1}(\xi) d\xi \) [or its complement, the miss rate, \( P_M(\tau) = 1 - P_D(\tau) \)] and the false-alarm rate is defined to be \( P_F(\tau) = \int_0^\tau f_{\tilde{\rho}|H_0}(\xi) d\xi \). The ROC curve \( (P_F(\tau), P_D(\tau)) \) can be drawn by varying the value of \( \tau \) to reveal the discrimination capability of the system. Alternatively, the equal error rate (EER), \( (P_{\text{EE}}(\tau) = P_F(\tau) = P_M(\tau), \tau \in \mathbb{R}) \), 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 \( \Phi \left( \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right) + \frac{1}{2} \exp \left( \frac{-\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} (\mu_0 - \mu_1) \right) \), respectively [13], [14], where \( \Phi(\cdot) \) is the cumulative density function for the standard Gaussian distribution. The theoretical quantities, mean \( \mu_i \), standard deviation \( \sigma_i \), and rate \( \lambda_i \), can be replaced with their estimates from the real data. In this way, EER can be estimated even the data points are not that many and/or the PDFs are widely separated. 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, a physical feature of a paper surface, has been found to have great discrimination power. Clarkson et al. [5] use 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: \( 0^\circ \), \( 90^\circ \), \( 180^\circ \), and \( 270^\circ \). 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 read 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.

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

Fig. 3: Scanned images from four perpendicular orientations of a piece of \( \frac{2}{3} \times \frac{2}{3} \text{inch}^2 \) (1.69-by-1.69 cm²) paper, and the resulting estimated norm map covering 1/100 area of that paper.
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 nonmatched cases (\( H_0 \)). All six plots reveal that the distributions for two hypotheses are far apart and have no overlap (i.e., no false alarm and no missed detection), suggesting a very good 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. 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. Extrapolated 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 lightings for image acquisition are well controlled, we explore a semi-controlled lighting condition with the help of the built-in flashlight 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 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 (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

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 were held by hand roughly in parallel with the surface of a piece of paper and at a height about \( z = 15.5 \) 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.
TABLE I: Capturing Conditions for Various Databases

<table>
<thead>
<tr>
<th>Database ID</th>
<th>502</th>
<th>503</th>
<th>506</th>
<th>508</th>
<th>509</th>
<th>510</th>
<th>501</th>
<th>511</th>
<th>505</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lighting</td>
<td>flash only</td>
<td>flash + ambient light</td>
<td>ambient light only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device</td>
<td>iPhone 6</td>
<td>iPhone 5s</td>
<td>iPhone 5</td>
<td>iPhone 6</td>
<td>Canon SX230HS</td>
<td>iPhone 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room Size</td>
<td>small</td>
<td>large</td>
<td>small</td>
<td>large</td>
<td>small</td>
<td>large</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 5](image_url) 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.

were needed to capture the whole region containing the patches, and the camera positions relative to the paper are shown on Fig. 2(a). Boundaries among different shots are separated by the thick lines. Figs. 2(b) and (c) containing luminance non-uniformity were acquired without and with flashlight for the top-left 20 patches on the layout of the paper.

B. Proposed Method

One should note that 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.

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 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.

C. 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 the 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 $10^{-3}$ to $10^{-3}$ (see Table VI for comparison) per our discussion on estimating the EER 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 to the authentication performance whereas the ambient light is not an important factor. Table III reveals that good authentication performance can be achieved across devices of similar imaging modules. Also, limited performance is within our expectation for all pairs of iPhone camera and Canon camera, because the relative position of the flash module to the lens and the pattern of the flashlight are different between these two brands of cameras.

V. PROPOSED SURFACE NORM ESTIMATION FOR PAPER AUTHENTICATION USING MOBILE CAMERAS

Although the authentication scheme using flashlight proposed in Section IV outperforms schemes that flashlight are not used, the requirement that the test and reference images must be captured at the same position relative to the physical
TABLE II: Mode of PDF of correlation values for $H_1$. Contrasting conditions: using flash or not, and room size.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Test</th>
<th>With Camera’s Built-in Flash</th>
<th>No flash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>501</td>
<td>502</td>
<td>503</td>
</tr>
<tr>
<td>501</td>
<td>0.48</td>
<td>0.44</td>
<td>0.43</td>
</tr>
<tr>
<td>502</td>
<td>0.45</td>
<td>0.45</td>
<td>0.47</td>
</tr>
<tr>
<td>503</td>
<td>0.43</td>
<td>0.47</td>
<td>0.48</td>
</tr>
<tr>
<td>504</td>
<td>0.42</td>
<td>0.45</td>
<td>0.42</td>
</tr>
<tr>
<td>505</td>
<td>0.21</td>
<td>0.24</td>
<td>0.24</td>
</tr>
</tbody>
</table>

1 Italic numbers in this table and Table III correspond to the cases that $H_0$ and $H_1$ PDFs can be perfectly separated.


<table>
<thead>
<tr>
<th>Ref</th>
<th>Test</th>
<th>iPhone 6</th>
<th>iPhone 5s</th>
<th>iPhone 5</th>
<th>Canon SX230</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>503</td>
<td>506</td>
<td>508</td>
<td>509</td>
<td>510</td>
</tr>
<tr>
<td>503</td>
<td>0.48</td>
<td>0.41</td>
<td>0.47</td>
<td>0.44</td>
<td>0.36</td>
</tr>
<tr>
<td>506</td>
<td>0.42</td>
<td>0.43</td>
<td>0.44</td>
<td>0.42</td>
<td>0.36</td>
</tr>
<tr>
<td>508</td>
<td>0.46</td>
<td>0.42</td>
<td>0.52</td>
<td>0.47</td>
<td>0.38</td>
</tr>
<tr>
<td>509</td>
<td>0.44</td>
<td>0.41</td>
<td>0.47</td>
<td>0.50</td>
<td>0.37</td>
</tr>
<tr>
<td>510</td>
<td>0.37</td>
<td>0.35</td>
<td>0.38</td>
<td>0.37</td>
<td>0.35</td>
</tr>
<tr>
<td>511</td>
<td>0.28</td>
<td>0.26</td>
<td>0.32</td>
<td>0.30</td>
<td>0.23</td>
</tr>
</tbody>
</table>

A. Macroscopic Intensity Due to Flashlight

Examining the images captured under the flashlight in Fig. 2 (c), we observe that there exists a mild spatial intensity change at the scale of the image dimension. 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 details, the mild change at the macroscopic level should be first removed. We define the image intensity at the macroscopic level as the macroscopic intensity, $I_{\text{macro}}$.

It can be shown as follows that the macroscopic intensity $I_{\text{macro}}$ 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 $I_{\text{e}}$ of background pixels over a small neighborhood $N$ around a pixel location $p$:

$$I_{\text{macro}}(p) \approx I_{\text{e}}(p)$$

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

(a) $\approx \lambda \cdot l(p) \cdot \frac{1}{|N(p)|} \sum_{k \in N(p)} n(k)^T v(p)$

(b) $\approx \lambda \cdot l(p) \cdot \mathbb{E}[n(p)]^T v(p)$

(c) $\approx \lambda \cdot l(p) \cdot \mu_{n_z} \cdot v_z(p)$

where $|N(p)|$ is number of pixels in the small neighborhood of $p$; step (a) follows from the fact that $l(k)$ and $v(k)$ 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 $I_{\text{macro}}$ over the spatial coordinates makes parametric surfaces good candidate estimators. In this work, we fit a high-order polynomial surface directly to an image captured under flashlight using an iteratively reweighted least-squares method. The bisquare weights [15] were used to gradually lower the impact of outliers as iteration goes on. The original image and its parametrically fitted version are shown in Figs. 6 (b) and (c). As our objective is to obtain the macroscopic intensity due to the flashlight, the image pixels belonging to the tri-patch 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 fitting curve.

One should note that even though a detrended patch image can be obtained by pixel-wise division of macroscopic intensity $I_{\text{macro}}$, the detrended patch image is not suitable for authentication via correlating with some reference image. After detrending, images captured with flashlight/camera located at different relative locations to the physical patch can have similar visual appearances at a large scale, but differ in small details. This is caused by the different directions of incident light during the capturing with respect to the macroscopic surfaces. Fig. 6(e) shows four such detrended patch images when camera locations were at the four corners to the patch. They appear similar at a large scale due to detrending, but are very different at a small scale. Fig. 7 shows the averaged correlations among the detrended patches as a function of the differences in camera capturing locations, $(\Delta N_x, \Delta N_y)$. The figure reveals that the farther the capturing
distance between two patches is, the lower the correlation can be for the detrended patch images. This implies that it is not possible 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 of using 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:

\[
\zeta(\mathbf{p}) \approx \mathbf{n}(\mathbf{p})^T \mathbf{v}(\mathbf{p})
\]

where \( \zeta(\mathbf{p}) = \mu_{n\times} v_z(\mathbf{p}) \cdot l_x(\mathbf{p}) / l^{\text{macro}}(\mathbf{p}) \), defined as the normalized intensity, contains the unknown modeling constant \( \mu_{n\times} \), the image acquired under flashlight \( l_x \), and the already estimated terms \( l^{\text{macro}} \) and \( v_z \). On the right-hand side, normal vectors \( \mathbf{n}(\mathbf{p}) \) is yet to be solved, and incident light direction \( \mathbf{v}(\mathbf{p}) \) is known from previous estimation. The inference problem of 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 200×200 pixels. For each pixel location \( \mathbf{p} \), we setup a system of linear equations using \( M = 20 \) acquired images, where \( M \) is far greater than 4, the number of unknowns:

\[
\begin{bmatrix}
\zeta_1 \\
\zeta_2 \\
\vdots \\
\zeta_M
\end{bmatrix}
= 
\begin{bmatrix}
\mathbf{v}_1^T & 1 \\
\mathbf{v}_2^T & 1 \\
\vdots & \vdots \\
\mathbf{v}_M^T & 1
\end{bmatrix}
\begin{bmatrix}
\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \\
\beta \\
\end{bmatrix}
+ 
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_M
\end{bmatrix}
\]

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 \( \epsilon \).

### 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 flashlight. 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, as 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 the reference in our work.
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 \( z = 11.4 \text{ cm} \) contain around 500 pixels along the edge for each patch in raw images, which are of enough resolution (5.25× 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 complications due to camera perspective. 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.

1) Totally 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 totally 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 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 are 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 totally dark assumption by investigating more realistic scenarios with the addition of the ambient light. Sessions 6–10 (aka Exp. 2) and Sessions 11–15 (aka Exp. 3) and were captured in a low-strength diffuse ambient light environment using iPhone 6 and iPhone 6s, respectively. 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 \( (P_F(\tau), P_M(\tau)) \) 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

TABLE IV: Statistics for correlation values of matched cases (\( H_1 \)) for images captured in a totally dark environment (Exp. 1).

<table>
<thead>
<tr>
<th>Norm Vector</th>
<th>session 4</th>
<th>session 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\mu} )</td>
<td>0.534</td>
<td>0.554</td>
</tr>
<tr>
<td>( \bar{\sigma} )</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td>( x )-component</td>
<td>0.523</td>
<td>0.493</td>
</tr>
<tr>
<td>( y )-component</td>
<td>0.012</td>
<td>0.015</td>
</tr>
</tbody>
</table>
burden to users during the verification process.

Fig. 10(a) shows the estimated PDFs for the matched ($H_1$) and nonmatched ($H_0$) 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 that have light and heavy tails, respectively. Detailed explanation can be found in Section VII-E.

The EER results are calculated from, namely, the sample mean, $\hat{\mu}_i$, 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 good authentication performance 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 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 used iPhone series camera modules for exploring the possibility of estimating the normal vector field of 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 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 due to the fact that the flashlight 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 customize the decision thresholds differently for Samsung Galaxy Alpha and for iPhone cameras considering their different PDFs under $H_1$.

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

<table>
<thead>
<tr>
<th>EER</th>
<th>Match ($H_1$)</th>
<th>No match ($H_0$)</th>
<th>Gaussian</th>
<th>Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 2</td>
<td>$0.557 \pm 0.011$</td>
<td>$122.8$</td>
<td>$-0.002 \pm 0.010$</td>
<td>$129.5$</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>$0.532 \pm 0.015$</td>
<td>$97.4$</td>
<td>$-0.004 \pm 0.011$</td>
<td>$113.4$</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>$0.528 \pm 0.012$</td>
<td>$106.4$</td>
<td>$-0.004 \pm 0.012$</td>
<td>$103.9$</td>
</tr>
</tbody>
</table>

E. Comparison with Prior Work

In Table VI, we summarize the performances of the proposed methods and prior work as discussed in the previous sections of the paper. Our proposed image based method has similar performance as [6] using scanner as the acquisition device, slightly worse than [5] using industrial camera and a semi-controlled light, as we created a semi-controlled light using flashlight and by requiring the test and reference images to be captured at the same position relative to the physical patch. The proposed image based method outperforms [8] that uses a robust feature, implying 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 (i.e., [6], and the method proposed in Section IV) or inherently used (i.e., [5]) controlled lighting, we explored a semi-controlled lighting condition with the help of the flashlight of mobile cameras. The proposed norm map based method significantly outperforms all image based methods. Although it performs slightly worse than [5] using scanner as the acquisition device, the flexibility due to the mobile device modality makes the proposed method more practical for ubiquitous deployment such as counterfeiting detection by end consumers.

VI. Perturbation Analysis on Discrimination Power

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 to the users in the verification process, such as the number of flash images 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 strength of inaccuracy of the camera location estimate, and then perturb the camera location when calculating $\mathbf{v}_i$ to examine how 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 mm, 2.16 mm, and 0.84 mm, respectively, when the camera was placed at the height of about $z = 11.4$ cm.

Performance Drop Under Perturbation With the knowledge of the strength of inaccuracy of camera location estimates, we can examine how authentication performance will be affected by adding reasonable amount of perturbation. We chose $\sigma_x = 2$ mm, $\sigma_y = 2$ mm, and $\sigma_z = 0.9$ 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 current perturbation level. The resulting $5 \times 20$ correlation values are expected to have an increased variance due to the additional perturbation.

We analyze 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 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, $2\sigma_r$. This distribution of current perturbation level. The resulting $5 \times 20$ correlation values are expected to have an increased variance due to the additional perturbation.

We analyze 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 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, $2\sigma_r$. This
where \( a_i \sim N(0, \sigma_a^2) \) and \( e_{ij} \sim N(0, \sigma_e^2) \). 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 their analytical expressions of the maximum-likelihood estimators can be found 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) for \( \hat{\sigma}_r, \hat{\sigma}_a, \) and \( \hat{\sigma}_e \) shows a constant estimated random effect for the session, and an increasing estimated effect of perturbation as the perturbation level increases. Fig. 11(b) shows the mean correlation against the perturbation level with 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 this amount of light flashing 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 images from Sessions 6–10. For each session, 20 subsets of the available images were selected to reveal the correlation 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 has to identify whether the available image set contains extremely “bad” ones towards the estimation of the normal vector and correlation, which should be carefully treated in the selection process. We had 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 of 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 analyze 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 estimated random effect for session, and an increasing estimated effect of perturbation as images become fewer. The value of \( \hat{\sigma}_r \) reaches almost 0.1 when images used are 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, over other factors such as the capturing session 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, 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. 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 deems to follow a light-tailed Gaussian distribution. In contrast, if the correlation deems to follow

\[
r_{ij} = \mu + a_i + e_{ij}, \quad i = 6, \cdots, 10,
\]

\[
\hat{\sigma}_r, \hat{\sigma}_a, \hat{\sigma}_e
\]
a heavy-tailed Laplacian distribution, one should on average acquire at least 8 flash images.

VII. DISCUSSIONS

A. Interpretation of Norm Map Obtained From Low Resolution Images

When 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 linearized 2-D index 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 \( n_i^T \cdot v_i \) of the high-resolution image, where \( v_i \) and \( n_i \) are the directions of the incident light and normal vector at location with index \( i \), respectively. Hence,

\[
    u = \sum_{i=1}^N w_i \cdot n_i^T v_i \approx \left( \sum_{i=1}^N w_i \cdot n_i \right)^T v = \bar{n}^T v \tag{6}
\]

where \( v \) 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., \( v \approx v_i \). 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 \( \bar{n} = \sum_{i=1}^N w_i n_i \). 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 that 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 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 that 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 for lowering the authentication performance, one can consider applying blind deconvolution in the first place for deblurring before using them for authentication purpose.

C. Understanding the Physics of Paper Surface Reflection

In this section, 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 over 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 the alternative estimate. These two estimates for the surface direction agree with 0.98 correlation, implying that the physical normal vectors are not sensitive to different definitions of direction and estimation algorithms, and are therefore reliable.
We then regress vector, \( \mathbf{v} \), here, \( \mathbf{n} \) including the common effect of the light strength at location \( r \) by the light strength, non-negative weights for diffuse and specular images scaled \( \mathbf{M} \).

We first calculate a synthesized diffuse image with the help of the physical normal vectors from a confocal microscope. We examine the correlation of norm maps obtained from mobile camera and scanner with respect to the ground truth. The results are shown in Table VII. The non-zero 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 implies that the fully diffuse reflection model is not precise enough for the purpose of estimating the physical norm map.

**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 \( \mathbf{I}_d(\mathbf{p}) = \max \{ 0, \mathbf{n}(\mathbf{p})^T \mathbf{v}_i(\mathbf{p}) \} \) and a specular image \( \mathbf{I}_s(\mathbf{p}) = \max \{ 0, \mathbf{v}_r^T \mathbf{v}_r \} \) 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 \( \mathbf{v}_r = (2\mathbf{n}^T - \mathbf{I})\mathbf{v}_i \).

We then regress 20 camera captured images \( \mathbf{l}_c(\mathbf{p}) \) against the diffuse image and specular image, in order to obtain non-negative weights for diffuse and specular images scaled by the light strength, \( \{ w_d \cdot I_d^{(k)}(\mathbf{p}), k = 1, \cdots , 20, \forall \mathbf{p} \} \) and \( \{ w_s \cdot I_s^{(k)}(\mathbf{p}), k = 1, \cdots , 20, \forall \mathbf{p} \} \), 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 paper reflection model, 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 a generative modeling and the result of Table VII from a 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 on the robustness of the norm map under various conditions, and tailored remedy methods 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. All their experiments demonstrated that the norm map is a robust feature over these conditions.

More investigations should be carried out on the resilience against tampering of the paper surface, such as scratching, folding, crumpling, and on the reliability of the physical structure of over the time. Large scale tests for papers from different manufactures are beneficial to understand issues may arise in the deployment of the proposed technology. Below we discuss how to prevent the folding operation from being used as a denial of service (DoS) attack.

**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 trend 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

---

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

<table>
<thead>
<tr>
<th>Pair of Quantities</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile camera vs. confocal (ground truth)</td>
<td>0.19</td>
</tr>
<tr>
<td>Scanner vs. confocal (ground truth)</td>
<td>0.28</td>
</tr>
<tr>
<td>[Reference]: Mobile camera vs. scanner</td>
<td>0.59</td>
</tr>
</tbody>
</table>

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 camera and scanner with respect to the ground truth. The results are shown in Table VII. The non-zero 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 implies that the fully diffuse reflection model is not precise enough for the purpose of estimating the physical norm map.

**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 trend 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 \( f_{\hat{\beta}|H_0}(\hat{\rho}) \) and \( f_{\hat{\beta}|H_1}(\hat{\rho}) \), as well as the performance metrics \( P_D(\tau) \), \( P_F(\tau) \), ROC, and EER derived from them, are not \textit{a priori} known and need to be estimated from the practical data obtained from experiments. One can construct 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 has the advantage that it sticks to the real data especially when 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 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 \( \binom{50}{2} = 1225 \) 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 system if the achievable rates are much smaller than 1/1225.

To alleviate the drawbacks of using the empirical PMFs for inferences, we can incorporate more modeling flavors by assuming the theoretical PDFs \( f_{\hat{\beta}|H_0}(\hat{\rho}) \) and \( f_{\hat{\beta}|H_1}(\hat{\rho}) \) 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 extrapolated, and EER can be calculated as a deterministic function of moments such as the mean and the second moment. One should note that the correctness of extrapolated tails heavily depends on the assumption that data matches 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. In Section V-D, we drew/calculated the performance bounds for both ROC curves and EERs.

VIII. Conclusion

In this paper, we have investigated intrinsic microscopic feature of the paper surface for authentication purpose. We have shown that it is possible to use the cameras and built-in flashlights of mobile devices to estimate the normal vector field of paper surfaces. Perturbation analysis shows that the proposed method is robust to 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

We 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. We thank Dr. King Lam Hui for providing expertise in microscopic imaging and taking photos for paper samples using a DSLR camera. We thank Mr. Zhili Yang for taking photos for paper samples using a confocal camera. We thank the reviewers for the constructive comments.

REFERENCES

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, two-year and four-year merit-based full scholarships for M.Phil. and B.Eng. degrees, respectively. He is members 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 a University of Maryland Invention of the Year Award in 2012 and in 2015. She has served as the Vice President-Finance of the IEEE Signal Processing Society from 2010 to 2012 and the Chair of the IEEE Technical Committee on Information Forensics and Security from 2012 to 2013. She is currently the Editor-in-Chief of the IEEE Signal Processing Magazine and an IEEE Distinguished Lecturer. She is an IEEE Fellow for contributions to multimedia security and forensics.