Enabling Paper-Based Surface Authentication via Digital Twin and Experimental Verification

Prasun Datta
NC State University
pdatta2@ncsu.edu

Chau-Wai Wong
NC State University
chauwai.wong@ncsu.edu

Min Wu
University of Maryland, College Park
minwu@umd.edu

***Abstract*—Paper surfaces can be used for anticounterfeiting due to their inherent and physically unclonable irregularities. Prior work used mobile cameras to capture paper's microstructure with the help of camera flash. However, prolonged exposure to flash in the workplace may harm the eyes of workers involved in the authentication process. This work proposes an authentication method that exploits indoor lighting without the need for a camera flash. Indoor lighting has a lower strength and leads to interference due to secondary reflections, making it challenging to achieve a good authentication performance. To this end, we create a digital twin (DT) replication of a real world in which paper patches are captured under multiple lights, taking account of key physics and optical laws. From simulations of DT, we identify important factors to the authentication performance and design an authentication method for an office setup. We have experimented with three different types of paper and showed that the DT-guided authentication method can achieve satisfactory authentication performance without using active light sources.**

***Index Terms*—Digital twin, camera flash, light reflection model, normal vector, authentication, anticounterfeiting, physically unclonable function (PUF).**

I. Introduction

Counterfeiting becomes more prevalent amid the restructuring of the global supply chains, which affects everything from consumer rights to public health, and to national security. Paper-based surface authentication offers a robust, economical solution to the counterfeiting problem, leveraging the unique random microstructures of paper created by wood filaments [1]–[6]. This randomness at the microscopic level serves as a natural fingerprint, making it ideal for verifying valuable documents and expensive products. Clarkson et al. [3] considered paper surfaces to be fully diffuse and proposed the use of flatbed scanners to estimate surface normals as the authentication feature. This method's widespread use is constrained by scanners' lack of portability and the need for specialized operating knowledge. To address the limitations, Wong and Wu [5], [6] introduced a mobile camera-based authentication method for estimating surface normals, offering a more accessible solution for businesses, academic institutions, and government agencies with document or product authentication needs. However, this method's reliance on camera flash poses a potential risk, as prolonged exposure to flash could cause irreversible injuries to the retinas of workers [7], [8] involved in authentication processes.

This work is supported in part by the US National Science Foundation (award numbers ECCS-2227499 and ECCS-2227261).

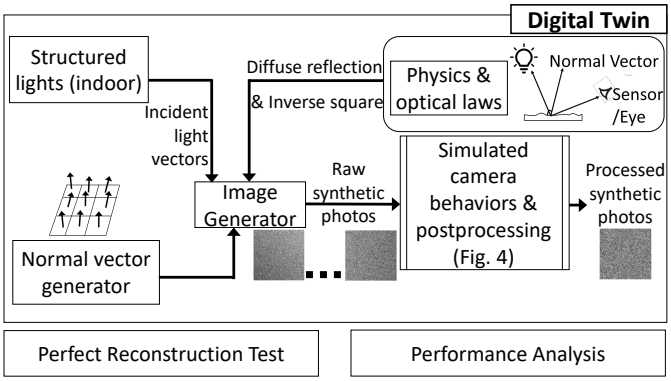


Fig. 1: Proposed digital twin uses an image generator to synthesize photos of paper patches by applying physics and optical laws. The synthetic process was verified through a perfect reconstruction test and thoroughly analyzed. This DT simulation, as discussed in Section V-A, guided the design of the proposed authentication method for a challenging office setup without active light sources.

In this work, we propose an authentication approach that eliminates the use of camera flash and relies on indoor lighting to assure worker safety. At first glance, indoor lighting has two challenges compared to using active light sources such as camera flash. First, the arriving incident light may be weaker, lowering the accuracy of estimated surface normals. Second, shadows and secondary reflections by walls and other reflective objects indoors may alter the visual appearance of paper patches. Both will make adapting paper-surface-based authentication methods from [3] or [5], [6] difficult, if not impossible. To overcome the aforementioned challenges, we introduce a digital twin (DT) to guide the design of a physical authentication method under indoor lighting. The proposed DT, as shown in Fig. 1, consists of an image generator that renders camera photos from surface normals by taking into account physics and optical laws, indoor lighting environment, camera behaviors, and postprocessing steps. DT simulates real-world camera operations and subsequent processing, thereby generating synthetic photos that closely resemble real ones. The postprocessing includes spatial blurring, detrending, and histogram matching, which are seen in the verification algorithms. A separate *perfect reconstruction test* box at the

bottom of Fig. 1 is used to ensure the correctness of DT in its simplest configuration. The *performance analysis* block is used for revealing important factors to facilitate the design of a real-world authentication method under indoor lighting. With insights learned from DT, we design an authentication method for an office setup in the physical world and verify its effectiveness with paper patches made from resume paper, copy paper, and cardstock paper. The authentication results suggest that verification without a camera flash is feasible. The contributions of this paper are threefold:

- We build a digital twin to assist in developing a physical authentication method that can verify a product through photos captured under indoor lighting without the need for a camera flash.
- We experimentally verify the effectiveness of our proposed authentication method on real-world patches captured in an office setup.
- We mathematically prove that the capturing condition of turning off one light is better than leaving one light on.

II. Background and Related Work

Authentication techniques that capture unique features of paper surfaces may be divided into two categories: visual [1], [2], [9] and physical feature approaches [3]–[6], [10], [11].

Authentication via Visual Features. The visual approach authenticates based on pixel intensities or handcrafted visual features of the paper. Buchanan et al. [1] employed a laser scanner to characterize paper surfaces and used cross-correlation of intensity measurements for authentication purposes. Beekhof et al. [2] leveraged mobile phones equipped with macro-lenses to capture paper surfaces, applying minimum reference distance decoding for identification. Toreini et al. [9] examined the patterns formed by light passing through paper, capturing the visual rendering of intrinsic texture using a consumer-grade camera. All these methods for detecting paper surfaces rely on visual features, whereas the literature reports that physical features are in general more reliable for authentication [3], [6].

Authentication via Physical Features. The microstructures of paper surfaces have random and intrinsic characteristics and may, therefore, be used for unique identification. The microstructures can be quantified through a norm map, which is a collection of three-dimensional (3-d) surface normal vectors projected onto the horizontal/xy plane. Assuming light reflection is entirely diffuse, Clarkson et al. [3] proposed using a flatbed scanner to acquire paper patches in opposite scanning directions to estimate a scaled norm map. Wong and Wu [4]–[6] avoided the use of scanners and proposed a method to estimate the normal vector field from patches captured by a mobile device's built-in camera flash. Liu et al. [10] improved reflection modeling by incorporating such factors as ambient lighting and camera contrast adjustment. Liu and Wong [11] analytically demonstrated that in the flatbed scanner setup, specular reflection is not an important factor in estimating norm maps. In this work, we demonstrate for the first time that indoor lighting alone is sufficient for conducting norm map-based authentication.

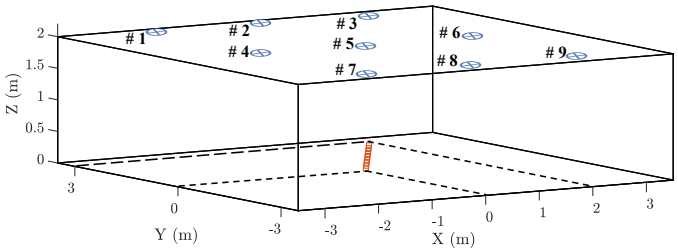


Fig. 2: A 3-d view of the capturing setup in the digital twin. The blue-circled crosses indicate equally spaced light sources. The red squares along a diagonal of the projected light matrix were patch locations used in Section IV-C for assessing the authentication performance within the digital twin. The light sources are 2 m high above the patches.

Digital Twin. A typical digital twin (DT) architecture includes three key components: the physical world, the virtual world, and the interconnection between the two worlds [12]. In the physical world of DT, cameras and/or sensors are used to capture images with different resolutions, depths, and thermal signatures, resulting in unique data properties [13]. In the virtual world of DT, game engines such as Unity Perception Package [14], Microsoft AirSim [15], and UnrealROX [16] can generate a far greater number of high-resolution synthetic images than can be captured in real life. Models trained on these synthetic images can be fine-tuned with limited real-world images for improved and more robust performance in physical world inferences [17]. Traditional game engines, while adept at rendering larger objects/scenes, struggle with generating finer, microscopic details [18]. In this work, our proposed digital twin is capable of synthesizing the subtle intensity variations due to microstructures.

III. Proposed Digital Twin for Paper Patch Generation, Rendering, and Imaging

In this section, we propose a digital twin to model the generation process of camera-captured photos of paper patches in the real world. We also incorporate several known processing steps from verification algorithms.

Synthetic Paper Patches. In the proposed digital twin, we model for each patch of a paper surface the only quantity that is relevant for authentication purposes, namely, a matrix of normal vectors. Since normal vectors typically point close to straight up [6], we sample each normal vector in a spherical coordinate from a narrow range of 5° in polar angle around the vertical axis. These normal vectors will be used to produce synthetic images using the fully diffuse reflection law.

Lighting Condition in a Simulated Room. We simulate a structured indoor lighting environment that replicates an office setup as illustrated in Fig. 2. Instead of using linear, rectangular, or circular sources, we opt for point sources as they can simplify the synthesis and analysis of patch photos.

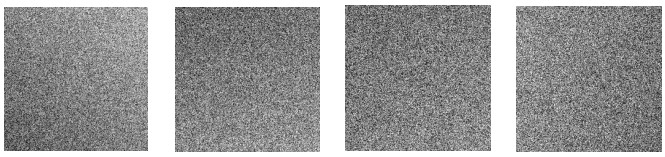


Fig. 3: Raw synthetic photos of a paper patch generated in the digital twin when experimented with a subset of lights #2, #3, #5, and #6 as indicated in Fig. 2. The photos were obtained under the image acquisition protocol mentioned in Section IV-A, which sequentially turns off lights #5, #6, #3, and #2, respectively, while keeping the rest three lights on. (The contrast of the displayed images was enhanced for visualization purposes.)

A matrix of nine lights is spaced 2 m apart horizontally and 3 m vertically and positioned at a height of 2 m above a flat surface where the paper patch is placed. Our experimental design for the digital twin's validation and testing utilizes merely lights #2, #3, #5, and #6. The remaining five lights are used in the subsequent exploration of the effects of increasing the number of light sources detailed in Section IV-C. The walls of real indoor rooms are assumed to be smooth and diffuse reflective surfaces, which may serve as secondary light sources. The satisfactory physical-world experimental results in Section V-C reveal that the digital twin does not need to model this phenomenon in this initial work.

Laws of Light Reflection and Inverse Square. To generate synthetic photos of paper patches, we utilize the fully diffuse reflection model as in [3], [6], [10]. We ignore the specular reflection component because the specular component is much weaker than the diffuse component [6]. The pixel intensity l_{r} at a pixel location $\mathbf{p}$ can be written as follows:

$$l_{\mathrm{r}} = \lambda \cdot l_0 \cdot \mathbf{n}^\top \mathbf{v} / \|\mathbf{v}\|_2^3, \tag{1}$$

where unit vector $\mathbf{n} = [n_x, n_y, n_z]^\top$ is the microscopic normal direction at pixel location $\mathbf{p}$ of the paper surface; $\mathbf{v}/\|\mathbf{v}\|_2$ is the incident light direction pointing from location $\mathbf{p}$ to the light source; albedo λ characterizes the physical capability of the micro-surface for reflecting the light; $l = l_0/\|\mathbf{v}\|_2^2$ is the arriving light intensity at location $\mathbf{p}$ due to the inverse square law, where l_0 is the intensity at the light source.

Patch Photos Generation. We generate *raw synthetic photos* of patches by individually computing the pixel intensity for each pixel location using the diffuse and decaying model (1). For example, we can choose a set of three lights, e.g., lights #2, #3, and #6, to obtain a raw synthetic photo. For normal vectors of the paper patch, we use the synthetic ones generated in the first subsection. For incident light vectors, each pixel location has a combined vector from the three light sources. (See Section IV-A for details.) A resulting raw synthetic patch of 200-by-200 pixels is shown as the left-most image in Fig. 3.

Simulated Camera Behaviors & Postprocessing. Capturing images of paper patches with a camera and passing them through a verification system may lead to additional spatial or distributional processing of the pixel intensities. In this digital twin, we design a pipeline illustrated in Fig. 4 aiming to capture needed processing traces like the real ones have. The pipeline comprises three modules, namely, spatial blurring filter, detrending, and histogram matching. We introduce them as follows.

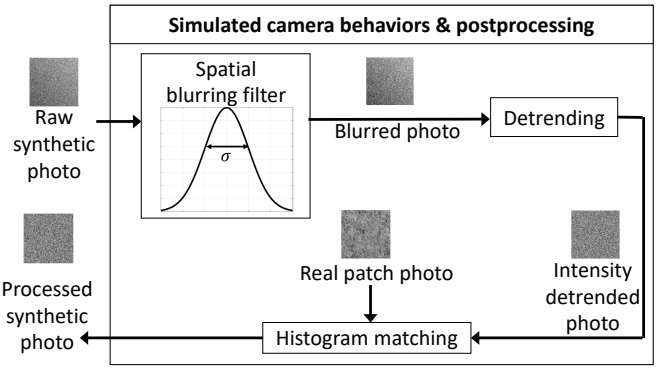
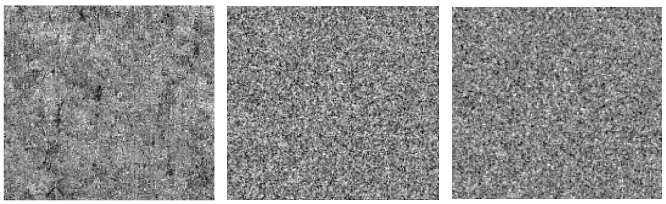


Fig. 4: A detailed view of the spatial or distributional processing steps applied to a raw synthetic photo of a patch to be verified, supplementing the generalized block in Fig. 1. The spatial blurring filter models camera blur and the effect of light diffusion in the patch. Detrending removes the spatial trend not used in verification. Histogram matching generates a final, processed synthetic patch that resembles the high-level appearance of a real patch.

The spatial blurring filter models the spatial intensity mixing/spreading caused by (*i*) the diffusion effect of the light within the fiber structure [19] and (*ii*) the point spread function of the camera [20]. In this first effort toward building a digital twin, we capture the intensity spreading using a Gaussian kernel, $G(x, y) = (2\pi\sigma^2)^{-1} \exp\left[-(x^2 + y^2)/2\sigma^2\right]$, where x and y are the pixel coordinates, σ is the standard deviation quantifying the spatial spread. We aim to search for the best σ to enhance the similarity between synthetic and real images. We parameterized real and synthetic images as first-order autoregressive (AR) processes [21] and formulated a loss function based on their respective AR parameters. We used iterative search to find the best σ: first in a search range of [0.3, 1.30] in increments of 0.1 and then narrowed the range to [0.4, 0.7] in more precise increments of 0.01. We determined that $\sigma = 0.5$ minimized the loss function. We provide a detailed analysis in Section IX of the supplemental document.

The detrending process eliminates the slowly varying spatial trends from raw synthetic images. Each spatial trend is induced by a specific lighting setup. Given that previous work [6] removed spatial trends in verification systems, they are not modeled in the digital twin.

Histogram matching aims to match the pixels' intensity distribution of a detrended blurred synthetic image with that of a real paper patch image. This process ensures that a final, processed synthetic image, as shown in Fig. 5(c) mimics the high-level appearance of a real patch, as shown in Fig. 5(a). The presence of the dark and bright spots in the real image



(a) (b) (c)

Fig. 5: Images of (a) a real paper patch; and a synthetic patch (b) after spatial blurring and detrending and before histogram matching, and (c) after spatial blurring, detrending, and histogram matching.

indicates regional, non-flat paper imperfections. We decided not to model these slowly varying spatial trends in the digital twin as they are removed in a practical authentication system that uses high-frequency subbands of reconstructed heightmap as the authentication feature [10], [11].

IV. Simulations with Digital Twin

In this section, we provide details for testing the digital twin and the design of a normal vector estimator. We also conduct various tests in the digital twin to evaluate the impact of lights on designing a real-world authentication method.

A. Lighting Setup and Normal Vectors Estimator Design

To simulate in the digital twin, we employ the light configuration in Fig. 2. They are point sources and arranged in a rectangular grid spaced horizontally at 2 m and vertically at 3 m apart. We choose the point source instead of other common shapes for light sources, including rectangular and linear, to reduce the design complexity of the normal vectors estimator. The paper patch is, by default, placed at a coordinate of (0.5 m, 0.5 m, 0 m). Since one needs $N \geq 3$ photos to estimate the three unknowns n_x, n_y, and n_z of a normal vector, we use four lights to generate four different lighting conditions/photos for the patch. Each photo is captured by turning off one light at a time and leaving the other three lights on. Since our scenario involves multiple turned-on lights except one, we can create an equivalent incident light vector by summing the individual incident light vector from the turned-on lights. This scenario of multiple turned-on lights is better than that of a single turned-on light as described in Section VII of the supplemental document. We analytically proved this claim in Section VIII of the supplemental document.

Next, we design an estimator to estimate normal vectors from synthetic photos of a patch generated in the digital twin, using the pixel intensity of these photos and the corresponding incident light vectors from all turned-on lights as inputs. We adapt the linear regression approach from [6] to create the estimator. The estimator works independently for each pixel and estimates normal vectors using a system of four equations. For example, when turning off light #1, we obtain intensity contribution $l_{\mathrm{r}}^{(k)} = \lambda l^{(k)} \cdot \mathbf{n}^\top \mathbf{v}^{(k)}$ from each light $k \in \{2, 3, 4\}$. Since the estimator is per pixel-based, we ignore location $\mathbf{p}$ for clarity. Dividing by the scalar term $\lambda l^{(k)}$ on both sides and summing up all three equations, we obtain

$$\zeta^{(1)} = \sum_{k=2}^{4} l_{\mathrm{r}}^{(k)} \big/ \big[\lambda l^{(k)}\big] \tag{2a}$$

$$= \mathbf{n}^\top \Big(\sum_{k=2}^{4} \mathbf{v}^{(k)} \Big) \stackrel{\mathrm{def}}{=} \mathbf{n}^\top \mathbf{v}^{(-1)}, \tag{2b}$$

where $\zeta^{(1)}$ represents the aggregated pixel intensity due to all three turned-on lights except light #1, and $\mathbf{v}^{(-1)}$ is defined as the combined incident light vector except light #1. When blocking the second, third, and fourth lights, we obtain three more equations alike, respectively. Hence, we can set up a system of four equations $\boldsymbol{\zeta} = \mathbf{V}\mathbf{n}$ to solve for the normal vector $\mathbf{n}$ for every pixel location, where $\boldsymbol{\zeta} = [\zeta^{(1)},\ \zeta^{(2)},\ \zeta^{(3)},\ \zeta^{(4)}]^\top$ and $\mathbf{V} = [\mathbf{v}^{(-1)},\ \mathbf{v}^{(-2)},\ \mathbf{v}^{(-3)},\ \mathbf{v}^{(-4)}]^\top$.

B. Perfect Reconstruction Test

Similar to multirate signal processing/filterbank research [22], we conduct perfect reconstruction (PR) tests for our proposed DT and normal vector estimator to ensure the correctness of our models. Four lights were used in the PR tests. Synthetically generated normal vectors were used as ground truth. We captured four distinct synthetic photos to estimate the normal vectors. Given the pixel intensities of the captured images, the arriving light intensity, and the incident light direction vector, the estimator should reconstruct the normal vectors with no error. To measure the reconstruction accuracy, we adopt cosine similarity $S_C(\mathbf{n}_{\text{ref}}, \mathbf{n}_{\text{est}}) = \mathbf{n}_{\text{ref}}^\top \mathbf{n}_{\text{est}} / (\|\mathbf{n}_{\text{ref}}\|_2 \|\mathbf{n}_{\text{est}}\|_2)$, and ℓ_2 distance $d(\mathbf{n}_{\text{ref}}, \mathbf{n}_{\text{est}}) = \|\mathbf{n}_{\text{ref}} - \mathbf{n}_{\text{est}}\|_2$ as performance metrics, where $\mathbf{n}_{\text{ref}}$ and $\mathbf{n}_{\text{est}}$ denote the reference and estimated normal vectors, respectively. Perfect reconstruction should result in a cosine similarity of one and an ℓ_2 distance of zero. Our proposed estimator achieves an average cosine similarity of 1 and an average ℓ_2 distance of 10^{-16}, indicating perfect reconstruction of normal vectors. The estimator operates pixel-wise for 200-by-200-pixel synthetic patches, and the average was taken over 40,000 pixels.

C. Performance Analysis

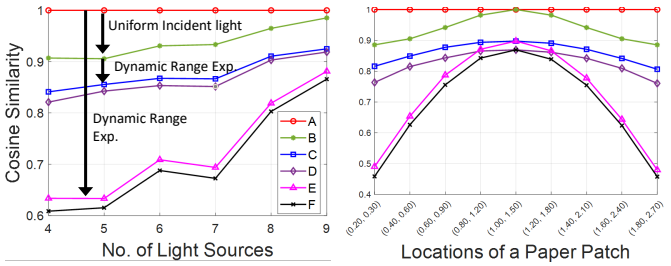
We explore the performance of the proposed digital twin under various configurations. We also investigate the impact of the number of light sources and the location of paper patches.

Simulated Configurations. We study various DT configurations and their differentiating factors' impact on reconstruction quality and present the results in TABLE I. Configuration A represents the baseline perfect reconstruction test setup detailed in Section IV-B. Configurations B to F deviate from perfect reconstruction by gradually including more factors. The factors examined include uniform incident light assumption, dynamic range expansion, spatial blurring, detrending, and histogram matching. We discuss the following key result that is relevant to the real-world authentication method design in Section V-B: Configuration B assuming uniform incident light

TABLE I: Various Configurations in Digital Twin and Reconstruction Quality.

	Configuration	Cos-sim ↑	ℓ_2 dist ↓
(A)	Raw synthetic photo (PR test)	1	10^{-16}
(B)	Uniform incident light + (A)	0.90	0.43
(C)	Dynamic range expansion + (B)	0.84	0.51
(D)	Postprocessing[1] + (C)	0.80	0.57
(E)	Dynamic range expansion + (A)	0.63	0.82
(F)	Postprocessing[1] + (E)	0.60	0.86

[1] The postprocessing includes spatial blurring, detrending, and histogram matching.



(a) (b)

Fig. 6: Reconstruction quality in terms of average cosine similarity for all configurations of TABLE I w.r.t. (a) the number of light sources, (b) the locations of a paper patch. The plots reveal that more lights or being closer to the geometric center of lights improve paper patches' authentication performance.

reduces cosine similarity by 0.10. In the capturing setup of Fig. 2, light sources are relatively far away when considering the physical dimension of the paper patches. It is reasonable to make a simplified assumption that all pixels are receiving light of the same strength and from the same direction. Due to its limited negative impact on performance, we adopt this uniform lighting assumption in Section V-B's design of the physical authentication method.

Number of Light Sources. For the most part of this work, we evaluate the estimator's performance using $N_{\mathrm{L}} = 4$ lights. We now investigate the impact of the number of light sources on the estimator's performance using $N_{\mathrm{L}} \in \{4, \ldots, 9\}$. Fig. 6(a) reveals that as the light count increases, the performance of the estimator improves almost linearly. This is reasonable as the least-squares estimate improves with more equations. We note that for designing a real-world authentication method in Section V, using a baseline of four lights can simplify the lighting setup while not incurring penalties proportional to the geometric function of the light count.

Location of the Paper Patch. We investigated how the location of a paper patch in the digital twin environment affects the estimator's performance. As the paper patch moves, the arrival light intensity and the incident light direction for every pixel of the patch change. We varied the location of the paper patch along the diagonal joining (0 m, 0 m, 0 m) and (2 m, 3 m, 0 m) in Fig. 2. Nine different locations listed in Fig. 6(b) were experimented. Fig. 6(b) reveals that as the patch moves from

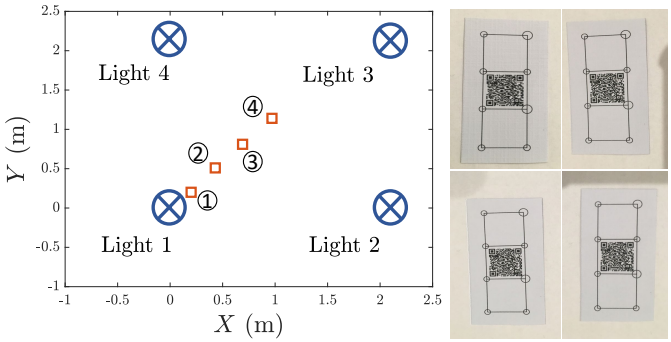


Fig. 7: The blue crosses indicate the lights. The red squares along the diagonal are patch test locations that were chosen with the help of the results of DT in Fig. 6(b). The patch areas in each photo were lit with only three lights due to the blockage of one of the four available lights, with shadows near the edges resulting from the obstruction of an additional light.

the corner toward the geometric center (1.00 m, 1.50 m) of the four lights, the performance improves. The best performance of the estimator is achieved at the geometric center for every configuration of TABLE I. This is due to symmetry, and we plan to provide analytic proof in our extended version. This result led us to pick candidate locations for real-world patches in Section V-B from only one of four quadrants shown in Fig. 7.

V. Proposed Real-World Authentication Method

A. Digital Twin Guidance

Designing a paper surface-based physical authentication method without active light sources such as camera flash could be challenging due to the lower strength of arriving light at the patch, resulting in a reduced signal-to-noise ratio (SNR) for the captured images. This reduction in SNR can make the precise estimation of the 3-d surface structure of paper patches nontrivial, leading to an enlarged design space of a real-world authentication method. We narrowed down the design space of a real-world authentication method utilizing the digital twin. First, the configuration of indoor lighting, which includes factors like the shape of light sources and the number of lights, is crucial to the authentication method design. We chose circular-shaped light sources that closely approximate point sources used in the digital twin in Section IV-A. Simulation results from Section IV-C/Fig. 6(a) led us to adopt a baseline of four lights for simplicity and with limited penalties for the real-world setup. Second, the placement of the paper patch is crucial for achieving satisfactory authentication performance since the illumination contributions from various sources can vary significantly at different locations. Fig. 6(b) aided us in selecting probable locations of paper patches for performing real-world image-capturing experiments. Third, through DT experiments, we found that the proposed capturing condition of turning off one light was better than leaving one light on, and our subsequent analytic analysis in Section VIII of the

supplemental document supports this. We took a similar capturing approach in real-world authentication. To summarize, digital twin guided our design of a real authentication method by identifying optimal configurations, thus saving time, effort, and financial resources that would otherwise be spent on extensive real-world experiments.

B. Experimental Design in Real World

As shown in Fig. 7, we selected an office setup where the indoor lights were arranged in a grid and the lights were circular-shaped. Three paper types, namely, copy paper, resume paper, and cardstock paper, were used in experiments. The right panel of Fig. 7 shows four photos of the same piece of resume paper with a printed tripatch alignment pattern [6]. Images were acquired with an iPhone 6s. Since the estimator's performance in DT is symmetric around the geometric location of the grid as revealed in Fig. 6(b), we varied the locations of paper patches along the diagonal line extending up to the geometric center in the office setup, as shown in Fig. 7.

We captured photos of real patches using four lights, which is the smallest-scale real-world lighting setup. Since the distance between a light and a patch is much greater than the physical dimension of the patch, this lighting setup closely approximates the uniform illumination condition analyzed in Section IV-C. We opted to take images of paper patches with multiple lights turned on, as this approach is superior to using just a single turned-on light, as discussed in Section V-A and proved in Section VIII of the supplemental document. The photographer used his body to block one of the four lights for each image captured. This casts a single shadow onto the paper patch, effectively making the patch "perceive" only three lights. While shadows are often viewed as detrimental in image processing and computer vision, in our design, they serve as a tactical capture method that eliminates the need to turn off any lights. This approach can smoothly fit into real-world workplace environments without disturbing other coworkers' workflow.

C. Authentication Performance of Real Patches

The authentication performance for three paper types placed in four locations is summarized in TABLE II. We managed to obtain meaningful authentication results in terms of nonzero correlation coefficients[1] for each location. This suggests that our digital twin successfully guided the design of a real-world authentication system in one shot, even though we did not explicitly model secondary reflections in our estimator design. We note that the resume paper performs the best with an average median correlation of 0.43 for the x-component and 0.46 for the y-component. In comparison, cardstock paper exhibits the lowest performance among the three paper types. When compared to copy paper, cardstock paper shows an average decrease of 0.11 in median correlation for the x-component and 0.06 for the y-component. We note that the surface of the resume paper is the most textured, whereas the surface of the cardstock paper is the smoothest. Our experimental results imply that rougher surfaces are easier to authenticate. We also notice a general trend of improved performance as the patch moves toward the geometric center of lights, which is consistent with the digital twin simulation in Section IV-C/Fig. 6(b).

[1]For this preliminary study, we skipped examining the correlations of unmatched cases as they tend to be tightly concentrated around zero per prior studies [4]–[6], [10], [11].

TABLE II: Median Correlation for Real-World Captured Patches Without Camera Flash.

Location Index	Resume Paper		Copy Paper		Cardstock	
	x	y	x	y	x	y
①	0.26	0.44	0.16	0.14	0.09	0.14
②	0.42	0.53	0.15	0.09	0.07	0.07
③	0.47	0.49	0.23	0.18	0.10	0.04
④	0.56	0.40	0.27	0.11	0.12	0.02
Average	0.43	0.46	0.20	0.13	0.09	0.07

Finally, we compare our copy paper's results (see the mid-column of TABLE II) without camera flash to those reported in the literature that leveraged flash. Wong and Wu [6] achieved a correlation of 0.52–0.56 for copy papers using the x- or y-component as the authentication feature. We note that the indoor lighting setup, given its challenging nature, leads to lower performance than the camera flash setup.

VI. Conclusion and Future Work

We have proposed a paper-based authentication method that leverages indoor lighting without the need for active light sources. We have proposed a digital twin to aid the design of the proposed method. In the digital twin, we have simulated synthetic patches and analyzed the performance under various configurations. With various important design factors revealed by the digital twin, we have successfully developed an authentication method for an office setup utilizing indoor lighting in only one shot. The nonzero correlation results across three different types of real-world paper surfaces suggest such an authentication system without a camera flash is feasible. In our expanded work, we plan to conduct more comprehensive real-world verifications using more paper patches and lighting setups. We also plan to include more theoretical analyses for key enablers of the proposed method.

References

[1] J. D. Buchanan, R. P. Cowburn, A.-V. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D. A. Allwood, and M. T. Bryan, "'Fingerprinting' documents and packaging," *Nature*, vol. 436, no. 7050, pp. 475–475, 2005.

[2] F. Beekhof, S. Voloshynovskiy, O. Koval, R. Villán, and T. Pun, "Secure surface identification codes," in *SPIE Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, 2008, pp. 142–153.

[3] W. Clarkson, T. Weyrich, A. Finkelstein, N. Heninger, J. A. Halderman, and E. W. Felten, "Fingerprinting blank paper using commodity scanners," in *IEEE Symposium on Security and Privacy*, 2009, pp. 301–314.

[4] C.-W. Wong and M. Wu, "A study on PUF characteristics for counterfeit detection," in *IEEE International Conference on Image Processing*, 2015, pp. 1643–1647.

[5] C.-W. Wong and M. Wu, "Counterfeit detection using paper PUF and mobile cameras," in *IEEE International Workshop on Information Forensics and Security*, 2015, pp. 1–6.

[6] C.-W. Wong and M. Wu, "Counterfeit detection based on unclonable feature of paper using mobile camera," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1885–1899, 2017.

[7] S. M. Michaelson, "Human exposure to nonionizing radiant energy–Potential hazards and safety standards," *Proceedings of the IEEE*, vol. 60, no. 4, pp. 389–421, 1972.

[8] D. H. Sliney and B. C. Freasier, "Evaluation of optical radiation hazards," *Applied Optics*, vol. 12, no. 1, pp. 1–24, 1973.

[9] E. Toreini, S. F. Shahandashti, and F. Hao, "Texture to the rescue: Practical paper fingerprinting based on texture patterns," *ACM Transactions on Privacy and Security*, vol. 20, no. 3, pp. 1–29, 2017.

[10] R. Liu, C.-W. Wong, and M. Wu, "Enhanced geometric reflection models for paper surface based authentication," in *IEEE International Workshop on Information Forensics and Security*, 2018, pp. 1–7.

[11] R. Liu and C.-W. Wong, "On microstructure estimation using flatbed scanners for paper surface-based authentication," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3039–3053, 2021.

[12] A. E. Campos-Ferreira, J. d. J. Lozoya-Santos, A. Vargas-Martínez, R. Mendoza, and R. Morales-Menéndez, "Digital twin applications: A review," in *Memorias del Congreso Nacional de Control Automático*. Asociación de México de Control Automático Puebla, Mexico, 2019, pp. 606–611.

[13] M. Ruman Islam, M. Subramaniam, and P.-C. Huang, "Image-based deep learning for smart digital twins: A review," *arXiv e-prints*, pp. arXiv–2401, 2024.

[14] Unity Technologies, "Unity perception package," https://github.com/Unity-Technologies/com.unity.perception, 2020.

[15] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and Service Robotics: Results of the 11th International Conference*. Springer, 2018, pp. 621–635.

[16] P. Martinez-Gonzalez, S. Oprea, A. Garcia-Garcia, A. Jover-Alvarez, S. Orts-Escolano, and J. Garcia-Rodriguez, "UnrealROX: An extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation," *Virtual Reality*, vol. 24, pp. 271–288, 2020.

[17] A. Alkhateeb, S. Jiang, and G. Charan, "Real-time digital twins: Vision and research directions for 6G and beyond," *IEEE Communications Magazine*, 2023.

[18] L.-Q. Yan, "Realistic rendering in 'details'," *IEEE Computer Graphics and Applications*, vol. 41, no. 4, pp. 20–26, 2021.

[19] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.

[20] K. Rossmann, "Point spread-function, line spread-function, and modulation transfer function: Tools for the study of imaging systems," *Radiology*, vol. 93, no. 2, pp. 257–272, 1969.

[21] A. K. Jain, *Image Representation by Stochastic Models*. Prentice-Hall, Inc., 1989, ch. 6, pp. 189–230.

[22] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Pearson Education India, 2006.

SUPPLEMENTAL DOCUMENT

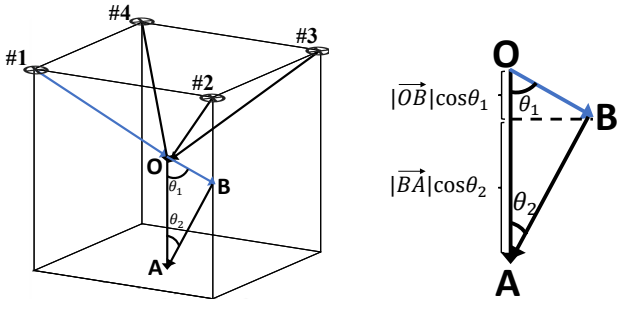
VII. ALTERNATIVE LIGHTING SETUP & CORRESPONDING NORMAL VECTOR ESTIMATOR DESIGN

Alternative Lighting Setup. We use four lights, i.e., lights #2, #3, #5, and #6, from Fig. 2 of the main paper to generate four distinct photos of the patch. Each photo is captured by turning on only one light at a time and leaving the rest of the lights off.

Estimator Design. Utilizing the pixel intensities from four photos of a synthetic patch and their corresponding incident light vectors, the normal vector is estimated for each pixel independently. For instance, when light #1 is turned on, the intensity is $l_{\mathrm{r}}^{(1)} = \lambda l^{(1)} \cdot \mathbf{n}^\top \mathbf{v}^{(1)}$. Dividing by the scalar term $\lambda l^{(1)}$, we obtain $\zeta^{(1)} = l_{\mathrm{r}}^{(1)} / \left[\lambda l^{(1)}\right] = \mathbf{n}^\top \mathbf{v}^{(1)}$, where $\zeta^{(1)}$ is the normalized pixel intensity due to light #1 and $\mathbf{v}^{(1)}$ is the incident light vector for light #1. Similarly, by turning on the second, third, and fourth lights, we obtain three more equations about $\mathbf{n}$, respectively. Therefore, a system of four equations $\boldsymbol{\zeta} = \mathbf{V}\mathbf{n}$ can be set up to solve for the normal vector at each location, where $\boldsymbol{\zeta} = [\zeta^{(1)},\ \zeta^{(2)},\ \zeta^{(3)},\ \zeta^{(4)}]^\top$ and $\mathbf{V} = [\mathbf{v}^{(1)},\ \mathbf{v}^{(2)},\ \mathbf{v}^{(3)},\ \mathbf{v}^{(4)}]^\top$.

VIII. PROOF: TURNING ONE LIGHT OFF IS BETTER THAN LEAVING ONE ON

In this section, we mathematically justify that turning off one light is better than leaving one light on. Without loss of generality, we compare two light configurations using the following two specific setups as illustrated in Fig. 8(a). In Setup 1 of "turning one light on", we turn on light #1 and leave lights #2, #3, and #4 off. In Setup 2 of "leaving one light on", we leave on lights #2, #3, and #4 and turn off light #1. It will be sufficient to prove that the vertical projection of the combined incident light vector due to three turned-on lights is greater than that of a single incident light vector due to one light.



(a) (b)

Fig. 8: (a) Two lighting configurations for the proof in Section VIII, and (b) relations among incident light vectors.

Before diving into the analysis of the two setups, we note the following facts that help to streamline the proof. First, the overall incident light resulting from all four possible lights is pointing straight down due to symmetry, which we denote as $\overrightarrow{OA}$. Second, the incident light vector solely due to light #1 is $\overrightarrow{OB}$. Third, the overall incident light vector due to lights #2, #3, and #4 is $\overrightarrow{BA}$, the difference between the aforementioned vectors. Fourth, it is easy to prove that $|\overrightarrow{OA}| > |\overrightarrow{BA}| > |\overrightarrow{OB}|$ and $\theta_1 > \theta_2$. We assume $|\overrightarrow{OB}| = 1$ and scale the other two vectors accordingly to simplify the following derivation.

In Setup 1, the expected reflected intensity at O is

$$\mathbb{E}\left[l_{\mathrm{r}}^{(\mathrm{Setup1})}\right] = \mathbb{E}\left[\lambda l \cdot \mathbf{n}^\top \mathbf{v}_1\right] \tag{3a}$$
$$= \lambda l \cdot \mathbb{E}[\mathbf{n}]^\top \mathbf{v}_1 \tag{3b}$$
$$= \lambda l \cdot [0,0,1]\mathbf{v}_1 \tag{3c}$$
$$= \lambda l \cdot |\overrightarrow{OB}| \cos\theta_1, \tag{3d}$$

where l is the arriving light intensity at O, which is the same for all four lights. In Setup 2, the expected reflected intensity at O is

$$\mathbb{E}\left[l_{\mathrm{r}}^{(\mathrm{Setup2})}\right] = \mathbb{E}\left[\sum_{i=2}^{4} \lambda l \cdot \mathbf{n}^\top \mathbf{v}_i\right] \tag{4a}$$
$$= \mathbb{E}\left[\lambda l \cdot \mathbf{n}^\top (\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4)\right] \tag{4b}$$
$$= \lambda l \cdot |\overrightarrow{BA}| \cos\theta_2. \tag{4c}$$

Finally, the ratio of the expected intensities between Setups 1 and 2 is

$$\frac{\mathbb{E}\left[l_{\mathrm{r}}^{(\mathrm{Setup1})}\right]}{\mathbb{E}\left[l_{\mathrm{r}}^{(\mathrm{Setup2})}\right]} = \frac{|\overrightarrow{OB}| \cos\theta_1}{|\overrightarrow{BA}| \cos\theta_2} \tag{5a}$$
$$= \frac{\sin\theta_2}{\sin\theta_1} \cdot \frac{\cos\theta_1}{\cos\theta_2} \tag{5b}$$
$$= \frac{\tan\theta_2}{\tan\theta_1} < 1. \tag{5c}$$

Here, (5a) implies that comparing the reflected intensities under two setups is reduced to comparing the vertical projection lengths of the incident light vectors. Fig. 8(b) gives a geometry interpretation. □

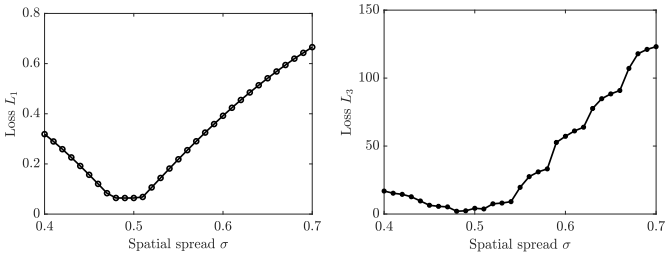
IX. CHOICE OF σ FOR SPATIAL BLURRING FILTER

In the proposed digital twin, we aim to synthesize photos of patches that resemble the high-level appearance of real patches by using the appropriate value of the spatial filter spread σ. One necessary condition is to ensure that the pixel intensity distribution of the final processed synthetic image in Fig. 5(c) matches that of the real patch image in Fig. 5(a). We construct three statistical grounded loss functions to characterize the difference in pixel intensity distributions as follows. All three losses are built on the estimated row-wise and column-wise AR(1) parameters for images.

The first two loss functions L_1 and L_2 are defined as follows. L_1 captures the difference of the mean AR parameters between a real image and a synthetic image, namely,

$$L_1(\sigma) = |\bar{\rho}_{\mathrm{row}}^{\mathrm{real}} - \bar{\rho}_{\mathrm{row}}^{\mathrm{syn}}| + |\bar{\rho}_{\mathrm{col}}^{\mathrm{real}} - \bar{\rho}_{\mathrm{col}}^{\mathrm{syn}}|, \tag{6}$$

where $\bar{\rho}_{\mathrm{row}}^{\mathrm{real}}$ and $\bar{\rho}_{\mathrm{row}}^{\mathrm{syn}}$ are the sample means of estimated row-wise AR parameters for the real and synthetic image, respectively. The quantities for columns are defined similarly.



(a) (b)

Fig. 9: Statistical grounded loss functions to enforce high-level similarities between synthetic and real patches: (a) correlation coefficient based, $L_1(\sigma)$, and (b) KL divergence based, $L_3(\sigma)$. Both are minimized around $\sigma = 0.5$.

L_2 captures the difference of the median AR parameters between a real image and a synthetic image, namely,

$$L_2(\sigma) = |\tilde{\rho}^{\text{real}}_{\text{row}} - \tilde{\rho}^{\text{syn}}_{\text{row}}| + |\tilde{\rho}^{\text{real}}_{\text{col}} - \tilde{\rho}^{\text{syn}}_{\text{col}}|, \tag{7}$$

where $\tilde{\rho}^{\text{real}}_{\text{row}}$ and $\tilde{\rho}^{\text{syn}}_{\text{row}}$ are the sample medians of the row-wise AR parameters for the real and synthetic image, respectively. The quantities for columns are defined similarly. The third loss function L_3 is based on the symmetric Kullback-Leibler (KL) divergence between the probability mass functions (PMFs) for the real and synthetic images defined below:

$$L_3(\sigma) = D_{\text{KL}}(P||Q) + D_{\text{KL}}(Q||P) \tag{8a}$$

$$= \sum_x P(x) \log\Big(\tfrac{P(x)}{Q(x)}\Big) + \sum_x Q(x) \log\Big(\tfrac{Q(x)}{P(x)}\Big), \tag{8b}$$

where $P(x)$ and $Q(x)$ are PMFs of estimated AR parameters of the real and synthetic images, respectively.

For each of the aforementioned statistical grounded loss functions, we iteratively searched for an optimal σ that ensures the maximum similarity between the real and the final processed synthetic image. We used an initial search range of $[0.3, 1.30]$ with a step size of 0.1, and then narrowed down the search range to $[0.4, 0.7]$ with a more precise step size of 0.01. The first two losses achieved their minima at $\sigma = 0.5$, as shown in Fig. 9(a), whereas the third loss achieved the minimum at $\sigma = 0.48$, as presented in Fig. 9(b). Since the difference between the optimal σ is small, we used $\sigma = 0.5$ to generate synthetic photos.