PaperSpeckle: Microscopic Fingerprinting of Paper

Ashlesh Sharma[†]
ashlesh@cs.nyu.edu

Lakshminarayanan Subramanian[†]
lakshmi@cs.nyu.edu

Eric Brewer[*]
brewer@cs.berkeley.edu

[†]Courant Institute of Mathematical Sciences
New York University
[*]Department of Computer Science
University of California, Berkeley

ABSTRACT

Paper forgery is among the leading causes of corruption in many developing regions [2]. In this paper, we introduce PaperSpeckle, a robust system that leverages the natural randomness property present in paper to generate a fingerprint for any piece of paper. Our goal in developing PaperSpeckle is to build a low-cost paper based authentication mechanism for applications in rural regions such as microfinance, healthcare, land ownership records, supply chain services and education which heavily rely on paper based records. Unlike prior paper fingerprinting techniques that have extracted fingerprints based on the fiber structure of paper, PaperSpeckle uses the *texture speckle pattern*, a random bright/dark region formation at the microscopic level when light falls on to the paper, to extract a unique fingerprint to identify paper. In PaperSpeckle, we show how to extract a "repeatable" texture speckle pattern of a microscopic region of a paper using low-cost machinery involving paper, pen and a cheap microscope. Using extensive testing on different types of paper, we show that PaperSpeckle can produce a robust repeatable fingerprint even if paper is damaged due to crumpling, printing or scribbling, soaking in water or aging with time.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection
; I.4.7 [**Computing Methodologies**]: Image Processing And Computer Vision–Feature Measurement

General Terms

Algorithms, Design, Experimentation, Security

Keywords

paper fingerprinting, paper speckle

CCS'11, October 17–21, 2011, Chicago, Illinois, USA.
Copyright 2011 ACM 978-1-4503-0948-6/11/10 ...$10.00.

1. INTRODUCTION

Forgery of paper documents has been a major cause of corruption in several countries around the world, especially in developing regions. In these regions, most of the important essential services such as financial systems, healthcare, governance, land records completely rely on paper as the basic medium for storing critical information. In addition, these services use paper as the primary means for establishing identity and verifying the authenticity of information. Hence, any form of mass paper forgery can negatively impact the functioning of essential services and affect large populations. For example, the recent stamp paper scam in India unearthed an underground racket that produced fake stamp papers to forge important land and governance records resulting in a massive loss of over $10 billion dollars to the exchequer [2].

In this paper, we introduce PaperSpeckle, a robust system that leverages the natural randomness property present in paper to generate a fingerprint for any piece of paper.

The basic idea used in PaperSpeckle is the concept of *texture speckles* - when light falls onto an object, the scattered light when projected to a screen produces bright and dark regions formed due to paper's texture and underlying physical non-uniformities. Texture speckles are randomly distributed making them a worthy candidate for generating fingerprints. The concept of *speckles* has been used in the area of laser speckles to profile objects [15]. Typically, extracting good speckle patterns requires expensive equipment. In PaperSpeckle, we show a simple repeatable texture speckle pattern extraction using paper, pen and a low cost microscope. We capture the phenomena of multiple scattering of partially coherent light (natural light) from the complex microscopic structure (surface irregularities and particles) of the paper to obtain the texture speckle pattern and use this information to produce a fingerprint of a specific region of the document.

Based on our experiences working in several rural developing regions, PaperSpeckle also addresses many of the practical constraints that occur in developing regions. PaperSpeckle is low-cost, portable and compact. PaperSpeckle can work with sub-$100 USB microscope connected to a simple mobile device. Paper based documents are very poorly maintained in developing regions and can easily get damaged due to a variety of factors: bad storage environments, damage due to rain, crumpling and aging of paper. Based on extensive stress testing, we show that PaperSpeckle is robust and can withstand severe environmental conditions. To illustrate the magnitude of robustness and damage resistance, PaperSpeckle can match fingerprints even when the conditions are non-ideal: (a) crumpling; (b) soaking the paper in water; (c) generating the fingerprint under different lighting conditions (different microscopes) and (d) aging of paper over time.

The rest of the sections are organized as follows: Section 2 provides a brief description of speckles in general and introduces the concept of paper speckle and how texture speckle patterns can be extracted from a paper. Section 3 presents fingerprint generation algorithm, that is used to compare and distinguish speckle patterns, Section 4 presents our implementation of the PaperSpeckle system on a desktop/laptop setting and also on a cellphone. Section 5 presents a detailed evaluation of PaperSpeckle system and Section 6 discusses the application scenarios. In Section 7, we discuss some of the issues regarding cloning or fabrication of paper and some its limitations and we conclude with a look into the future in Section 8.

1.1 Related Work

There have been a variety of paper fingerprinting solutions [8, 18, 16, 9, 30, 25, 7, 31, 3, 29] that have been proposed to deal with this problem of paper forgery. One canonical solution is to manufacture watermarked paper that uses a special form of paper or ink material that is hard to reproduce [29, 8, 18, 16, 9, 30]. Another approach is to use different types of lithography techniques [10] to embed a unique watermark in paper that would be hard to remove or duplicate. The problem with both these approaches is that they require expensive machinery or access to specialized paper (which can be limited) which can constrain their applicability to specialized applications such as currency notes, checks, official government paper records *etc.* In addition, the paper/inks used in some of these techniques [29, 16, 9] are specially prepared (using physical or chemical means). Another problem with these watermarking techniques is that they embed the same watermark across a bulk collection of documents (eg. currency notes, checks, official paper); hence watermarked documents of the same type are *indistinguishable.* In many common applications (healthcare, finance) which use paper-based records, it is essential to distinguish individual paper documents from each other and prove its authenticity. While standard bar-coding techniques [27] can embed a unique code into each paper, such codes can easily be reproduced and duplicated. PaperSpeckle differs from these standard techniques in that it uses the natural randomness in paper and provides a low-cost distinguishable fingerprint for any piece of paper.

Smith *et. al.* introduced fiber fingerprinting [7, 25] which uses the fiber structure of the paper to provide unique signature of the paper. There have been patents on authenticating paper documents based on their grain or fiber structure [19, 12]. The Print Signatures work by Zhu *et. al.* [31] uses the randomness in character printing by a laser printer to provide a unique signature of that region of paper. They use a microscope to zoom into the minor ink splatters near a character and extract the random pattern associated in the ink splatter. Recently, Clarkson *et. al.* [4] used mid-range scanners to model the 3D fiber structure of a paper and provide unique fingerprints based on it. Cowburn *et. al.* [3, 1] use laser microscopy to look into the complex fiber structure of paper to produce a unique fingerprint of the paper.

Our approach differs from these related works in three significant ways. First, in PaperSpeckle we do not form a 3D representation of the underlying paper structure nor do we use laser microscopy. We capture the phenomena of multiple scattering of partially coherent light (natural light) from the complex microscopic structure (surface irregularities and particles) of the paper region using a microscope to obtain the texture speckle pattern and use this information to produce a unique fingerprint of a region of the document. The physical property that we use to fingerprint the paper is very different from existing solutions. The scale at which the speckle pattern is extracted is at a microscopic level: a pixel in a speckle pattern is about 1-2 microns. To give a sense of the scale, the size of a red blood cell in the human body is about 8 microns.

Second, the technique used in fingerprint generation is different from existing solutions and the fingerprint is much more compact. In addition to this, we also provide a detailed evaluation with respect to adverse environmental conditions and show that our system is robust in real world settings. Not only our system works without any modification to the paper document, it can be used with any specialized ink (such as the Uniball 207 Gel ink) to provide robust security in various scenarios. Texture speckle patterns can be extracted even when specialized ink or paper is used. We do not generate secure sketch of the fingerprint, since in developing regions authentication of a piece of paper is performed in an offline manner with an untrusted device. Also, the adversary might have access to the original document, which makes the secure sketch non-useful.

Third, unlike bulky equipment like scanners and laser surface authentication devices, we use a portable, handheld microscope to obtain the speckle pattern, that can be used in a widespread fashion in developing regions. Our system works both on a desktop/laptop and a cellphone. Cellphone is widely prevalent in the developing regions across the world and since our system also works on a cellphone, it can be readily used in a variety of settings.

Optical marks [17] are used to authenticate paper documents using latent images in different layers of a paper document. In our application, we do not manufacture paper documents, nor do we modify the document in any way. Due to the effect of multiple scattering of light through the structure of the paper, *texture speckle pattern* arises; and the fingerprint generated from this *texture speckle pattern* is used to authenticate that piece of paper.

2. SPECKLES

In this section we provide a brief background on laser speckles and then introduce paper speckles which are based on partially coherent light source (light of finite bandwidth). We discuss its advantages over laser speckles, describe the device setup and finally show how repeatable paper speckle patterns can be extracted using a microscope (with inbuilt LED) and a piece of paper.

2.1 Laser Speckles

When light falls onto an object and the scattered light is projected onto a screen, the screen is speckled with bright and dark regions which represents a speckle pattern [11]. A Speckle pattern is a random intensity pattern produced by the mutual interference of coherent or partially coherent wavefronts that are subject to phase differences or intensity fluctuations. At the screen these rays have a different op-

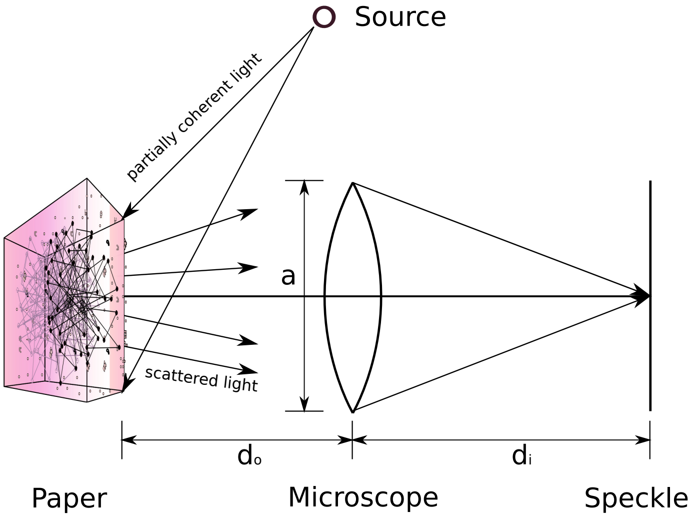


Figure 1: The light source is in the same plane as the observation point. The structure is the three dimensional cross section of the paper and the dark edges are light rays that undergo multiple scattering. The rays exiting the paper have a different optical path length that give rise to bright and dark regions representing a texture speckle pattern

tical path lengths; therefore the rays interfere and result in speckles.

The concept of speckles have been widely used in the field of laser speckles where a random speckle pattern is created when a coherent laser beam is reflected of a rough surface. Laser speckles have been used in research literature to fingerprint a wide range of objects including finger, paper documents, plastic cards, product packages [3]; the randomness in the laser speckle pattern can be used to uniquely identify the object. Using diffuse scattering of a focused laser, the fine structure of different surfaces can be extracted. However, the speckle pattern is very much dependent on the angle of measurement; hence laser speckle extraction requires expensive machinery to align laser as well as carefully extract the object surface pattern.

2.2 Paper Speckles

While laser speckles make a good candidate for extracting speckle patterns from paper, the underlying machinery is very expensive, delicate and fairly impractical in rural, developing country settings. Similarly, the extraction of speckles from partially coherent light source require expensive machinery, delicate experimental setup and careful modification of light source (to achieve either temporal or spatial coherence) [13, 23, 14]. Instead, we show that in practice we do not need to depend on lasers or specific partial coherent light sources and can capture light scattered from microscopic non-uniformities of an object illuminated by any partially coherent light source (such as LED) to identify and fingerprint the object. We call this type of patterns as *texture speckle patterns.* They are essentially, scattered light captured from the complex underlying microscopic structure of an object.

Consider the setup in Figure 1; using a simple microscope (with a 10-200X zoom) with an inbuilt LED source, we can extract a *texture* speckle pattern when we focus the microscope on a specific portion of the paper. Here, the texture speckle pattern represents how partially coherent light from the LED gets scattered through the rough surface and the underlying microscopic structures of the region of the paper focused on.

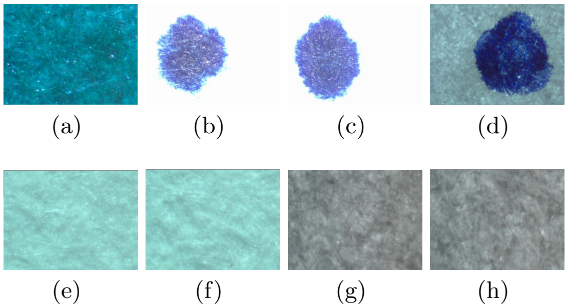


Figure 2: Speckles captured using USB microscope (a) a region of paper at 200X (b)-(d) speckles with ink stain (e)-(h) speckles without ink stain

These texture speckle patterns generated from a simple microscope with an inbuilt LED light source makes an excellent choice for a paper fingerprint due to a variety of factors. First, as shown in prior work on laser speckles [3], speckle patterns generated from lasers are tamperproof even if paper is soaked in water, crumpled, affected by aging etc. - the chances that the microscopic structure gets affected are small. As we show in this paper, the fingerprint generated from texture speckle patterns from partially coherent light source is sufficient to achieve the same objectives. Second, generating these texture speckle patterns is cheap (when compared to laser speckles), and is easy to use which make them a great choice for developing regions. Third, extracting just the fiber structure of the paper at a relative macroscopic level with a small zoom might be susceptible to cloning by the adversary. While the field of view of our microscopes are about 0.5mm (or much smaller based on the magnification) in diameter, fiber fingerprinting uses a much larger region of about 2.5 cm x 2.5 cm.

The light from the LED is focused on the paper and the scattered light from the paper is captured by the imaging system, which consists of an optical microscope arrangement. The microscope has two controls, one is the magnification and another is the focus. The images are captured using a fixed magnification of 200x, but the focus is varied to provide a more crisp image, with minimal blur.

The texture speckle pattern is dependent on the properties of the imaging system: the type of incident light source, the distance between the light source and the object, the distance between the lens and the object d_o, the distance between the lens and the image capture plane d_i and the diameter of the lens a; the roughness characteristics and the configuration of microscopic structures in the object. Therefore, in order to obtain repeatable texture speckle patterns generated from partially coherent light, the properties of imaging system (d_i,d_o,a) have to be fixed, across trials. The two types of imaging systems of microscope assemblies that we use (Digital Blue QX5$^{\text{TM}}$and PC Gears AM2011$^{\text{TM}}$), can be configured such that its properties can be fixed across trials.

While we observe a texture speckle pattern, repeatedly extracting the pattern is not straightforward, as it is difficult to point to the exact region on the paper each time to obtain measurements. This is primarily due to the scale at which texture speckle is observed. To address this problem, we use a pen to stain a small region with an arbitrary contour; thereby producing an *arbitrary contour shaded region* which is about 0.5mm in diameter. Figure 2(a)-(d), illustrates different texture speckle pattern pictures extracted using this technique.

Using arbitrary contour shaded regions is important for several reasons. First, having a small shaded region aids in focusing within the same region for repetitive attempts since the goal would be to have the entire region appear within the microscope's field of view. Second, the arbitrary contour helps in orienting the images in a specific direction, as it helps in texture speckle image comparison.

Without the stain, texture speckles can be compared and matched, but the registration process (image registration) has to be standardized. The microscope and the paper need to be arranged in a specific way and this arrangement should provide the same texture speckle pattern from the same region of the paper on repeated measurements. Essentially, the extraction of the texture speckle pattern from the same region of the paper should be *repeatable*. If such a type of system is conceived, then the ink stain need not be used.

In this paper, we have used the ink stain, as we have performed experiments on a large scale to evaluate PaperSpeckle (on over 1500 pieces of paper), and not conformed to any sort of standardization. (In the rest of the paper, when we refer to speckles, we mean texture speckles).

3. FINGERPRINT GENERATION

To generate a fingerprint for a speckle pattern and to satisfy the properties that are required of the fingerprint we employ a technique that is used in texture analysis called Gabor transform and then use Singular Value Decomposition (SVD) to obtain eigenvalues (or singular values) of the Gabor transformed speckle. The details of the algorithm and the rationale are described below. First, we discuss Gabor transform and explain how Gabor bit sequences can be distinguished from each other. Second, we discuss the method of fingerprint generation from Gabor bit sequences.

3.1 Gabor transform

We convert the speckle image into a bit representation by applying Gabor transform to a speckle image. There are three reasons for using Gabor transforms:

1. It is used in laser speckle evaluation [22], iris recognition [6], as it is used to show statistical independence of speckles: empirically show that any two speckles are never the same.
2. Due to the statistical independence property obtained by using Gabor transform, it is a simple method to obtain a compact bit representation of a speckle image and distinguish any two bit patterns using the Hamming distance.
3. Gabor transform applied to speckles is robust to global changes in illumination and minor modifications of the speckle image [20].

Using the above technique, pairs of speckle patterns can be compared to find similarity (or dissimilarity). To compare two speckle patterns, we first convert each speckle image into bits using 2D Gabor transforms. Next, we compare the two speckle bits using the Fractional Hamming distance (FHD) metric to check if the speckles are similar or dissimilar. Comparing speckles using FHD shows the statistical independence property of Gabor transformed bit patterns.

i) Gabor transforms: We apply Gabor transforms [5] to speckle images and extract bits using the imaginary part of the complex phase of the Gabor wavelets,

$$g(x_0, y_0, f, \theta) = e^{-[\pi(a^2(x-x_0)^2+b^2(y-y_0)^2)]}e^{[i2\pi f(xcos\theta+ysin\theta)]}$$

The first term is a 2D Gaussian function located at (x_0, y_0) where a is width along the x-axis and b along the y-axis. The second term is a complex 2D sinusoid of frequency f and an orientation defined by θ. These parameters can be varied and Gabor transforms (or filters) can be applied to any location, scale or orientation of an image. The procedure to extract bits from the Gabor transforms can be stated as follows:

1. Compute the imaginary part of the complex phase of the Gabor transform for one orientation and one level.
2. Use the complex phase and apply zero as threshold to extract a binary sequence or a binary image.
3. Repeat this procedure for various orientations (θ) and levels (f).

The importance of using only the complex phase of the Gabor wavelet to extract bits is that, we eliminate any illumination effects, contrast or poor focus, present in the speckle image. Due to this, the extraction of speckle images need not be too precise. Also, mask bits are computed to remove the extraneous effects surrounding the speckle image. Mask bits are computed by thresholding the pixel values to 1 beyond the boundary of the speckle and 0 inside the speckle. If the speckle covers the entire image, then mask bits are not needed. Since, the region where the speckle pattern is extracted is known (marker, ink stain etc), any region outside of that is unnecessary. This unnecessary region is considered as the mask bits. These mask bits help in applying Gabor transforms to only the region within the image where speckles are present.

ii) Fractional Hamming Distance (FHD): Let A and B be the sequence of bits extracted from after applying Gabor transforms to two speckle images; $maskA$ and $maskB$ be the two mask bits of the respective speckle images. The similarity between two sequence of bits A and B can be computed using the Fractional Hamming Distance:

$$FHD = \frac{||(A \otimes B) \cap maskA \cap maskB||}{||maskA \cap maskB||}$$

$A \otimes B$ gives the difference between bits and the $\cap$ with mask bits prevents any extraneous (unnecessary) bits to be considered in computing FHD (|| || is the norm of the vector). The FHD provides a ratio, that defines whether the sequence of bits are either similar or dissimilar. Ideally a FHD of 0 would represent equal bits and therefore, a perfect match of the speckles images and a FHD of 0.5 would represent dissimilar bits (and therefore different speckle images), where the likelihood of 0 or 1 occurring in a bit sequence is equally probable.

Figure 3 shows the Gabor kernel convolved with a speckle image under various orientations (6 orientations, $\theta = 0$, 60^o, 120^o, 180^o, 240^o, 300^o) and levels (5 levels, where f is the scaling factor between successive levels) and the complex phase is thresholded to zero to extract the bit sequence. Level 4 and level 5 provide the required FHD, to statisti-

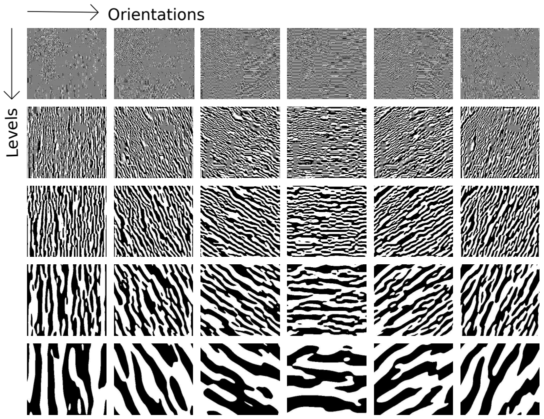


Figure 3: Gabor transform applied to a speckle image for various orientations and levels.

	Level 1	Level 2	Level 3	Level 4	Level 5
Instance 1	0.4898	0.4773	0.4636	0.3298	0.1654
Instance 2	0.4936	0.4798	0.3628	0.2146	0.1094
Instance 3	0.4507	0.4633	0.2863	0.1588	0.0662
Instance 4	0.4814	0.4897	0.4787	0.4245	0.2355

Table 1: Fractional Hamming Distance of a speckle with its own different instances across various levels. Each instance is a new measurement of the same speckle pattern.

cally distinguish two speckles. Table 1, shows the variation of FHD for various levels, for same speckle pattern extracted at different periods of time. (We have chosen orientation 5, which corresponds to $\theta = 240^0$ at each level, since the FHD values are similar for different orientations at a level). The FHD values at level 4 and level 5 are close to 0, which shows that these are the same speckle patterns. Table 2, shows the variation of FHD, when a candidate speckle is compared with different speckle patterns. The FHD values at level 4 and level 5 are close to 0.5, which show that these are different speckle patterns. We have found relatively similar FHD values (as described in Tables 1, 2) when trying to distinguish speckles in all of our experiments. In this paper, for statistical evaluation, we have chosen level 5 and orientation 5 as the Gabor bit sequence to compare speckle patterns using FHD. We have found similar statistical results for Gabor bit sequences extracted from different orientations of level 4 and level 5.

Figure 4 shows the FHD values of comparing 60,000 pairs of "different" Gabor bit sequences with mean 0.4875 and standard deviation 0.00577. When we say "different", we mean Gabor bit sequences in each pair correspond to speckle images taken from different regions of either the same paper or different paper. The least FHD value is 0.47. Each speckle image is of size 512 × 384 pixels. The Gabor transform of each speckle image provides a Gabor bit sequence of 196608 bits. Figure 5 shows the FHD values of comparing 200 pairs of "same" Gabor bit sequences with mean 0.1675 and standard deviation 0.0594. Here, "same" means Gabor bit sequences in each pair correspond to speckle images taken from the same region of the paper. The maximum FHD value is 0.28.

	Level 1	Level 2	Level 3	Level 4	Level 5
Speckle 1	0.4789	0.4737	0.4625	0.4591	0.4573
Speckle 2	0.4844	0.4803	0.4686	0.4410	0.4649
Speckle 3	0.4941	0.4662	0.4520	0.4706	0.4691
Speckle 4	0.4888	0.4876	0.4690	0.4608	0.4657

Table 2: Fractional Hamming Distance of candidate speckle with different speckles across various levels.

The two distributions are well separated and as seen from the figures any FHD value between 0.28 and 0.47 would distinguish two speckles and provide no false positive/negatives. How well are the two distributions separated? It would take the modification of at least 39321 bits for the two distributions to intersect each other, which around 20% of the total number of bits in the Gabor sequence. For the mean (peak) of two distributions to meet, around 32% (62914 bits) of the total number of bits in the Gabor sequence has to be modified.

We process 200 speckle images from four different kinds of paper and we examine each pixel value (intensity) of these speckle images. The mean of each pixel (across 200 speckle images) is computed and the mean value is plotted. The entropy of the Gabor bit sequence is maximized if each pixel value is 0.5 (the probability of each pixel being either 0 or 1). Figure 6 shows the probability of a bit being set in a Gabor bit sequence of a paper speckle is almost 0.5, which suggests that the entropy of Gabor bit sequence is high.

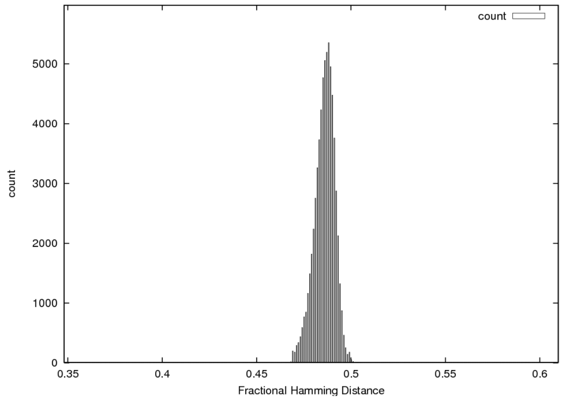


Figure 4: Fractional Hamming Distance of 60,000 pairs of different speckle image bits with mean = 0.48705 and standard deviation = 0.00577.

3.2 Fingerprint

Once we have the Gabor bit sequence for a speckle image, we convert the bit sequence into a binary matrix S_G such that its dimensions match the dimensions of the speckle image. Let S_G be the Gabor binary matrix. We apply SVD to S_G to obtain singular values, which we use as the fingerprint of the region of paper.

In SVD, we decompose S_G as, $S_G = U\Sigma V^T$ where orthogonal matrices U and V contain left and right singular vectors of S_G, respectively, and the diagonal of Σ contains the singular values of S_G. The singular values of any arbitrary matrix are uniquely determined [26]. These singular values are the square root of eigenvalues of $S_G S_G^T$.

Gabor binary matrix S_G is a binary (0,1) matrix and $S_G S_G^T$ is a real valued symmetric matrix. Therefore, the

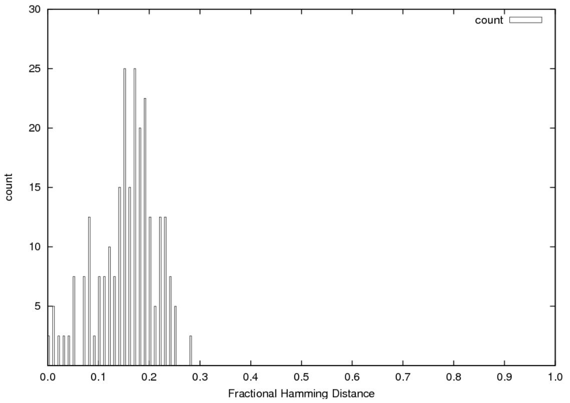


Figure 5: Fractional Hamming Distance of 200 "like" speckle bits with mean = 0.16751 and standard deviation = 0.05948.

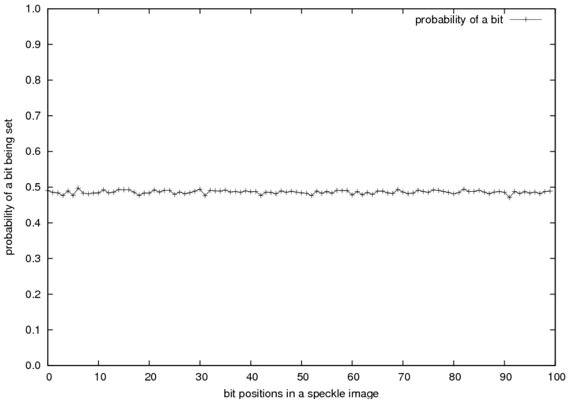


Figure 6: The probability of a bit being set in a Gabor bit sequence of the paper speckle image.

eigenvalues of $S_G S_G^T$ are real and well conditioned. Perturbations in $S_G S_G^T$ lead to perturbations of the same size in its eigenvalues (The proof is due to Stewart [26] and Papadimitriou *et. al.* [21] and states that if some large k singular values are sufficiently away from rest of the singular values, then the subspace spanned by the singular vectors are preserved if a small perturbation is added to S_G.)

Let $S1_G$ and $S2_G$ be Gabor bit sequences of the same region (where $S1_G$ is the Gabor bit sequence of the speckle image extracted in the first trial and $S2_G$ is the Gabor bit sequence of the same region extracted in the next trial) and $S3_G$ be a Gabor bit sequence of a different region. When we compare $S1_G$ and $S2_G$, we know that the Fractional Hamming Distance (FHD) between these are small (less than 4% of the total number of bits) as evidenced from the analysis of the FHDs Gabor bit sequences of "same" speckles. Hence, the difference in singular values of $S1_G$ and $S2_G$ would be a small number, indicating that $S1_G$ and $S2_G$ are the same speckle pattern. When we compare $S1_G$ and $S3_G$ (or $S2_G$ and $S3_G$), their FHDs would differ by a large amount (more than 20%) as evidenced from the analysis of FHDs of Gabor bit sequences of "different" speckles. Hence, the difference in singular values of $S1_G$ and $S3_G$ would be large, indicating that these are not the same speckles.

Next, we discuss the variation of singular values of S_G as the number of bits are varied.

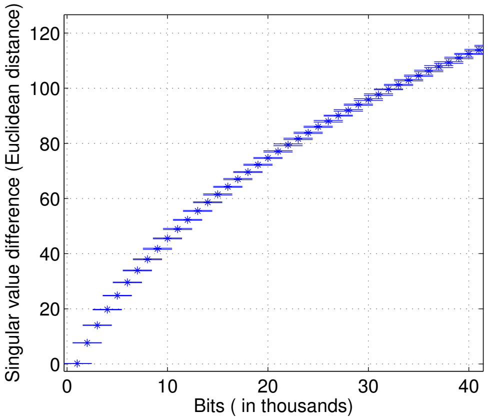


Figure 7: Singular value difference as the Gabor binary matrix is perturbed.

3.2.1 Singular value perturbation

We empirically examine the singular values of a Gabor binary matrix S_G as it is varied or perturbed. We consider one Gabor binary matrix S_G of size 512×384, and perturb the matrix randomly by modifying $10i$ bits in each step i of the experiment. In each step i, we perturb (flip) $10 \times i$ bits, where i is varied from 1000 to 196000. We conduct 20 trials in each step, compute mean and standard deviation and plot the results with error bars. Figure 7 shows the difference in the Euclidean distance between singular values of S_G and the singular values of the perturbed S_G, as the number of bits are modified. There is almost a linear increase in the difference of singular values as the perturbation of S_G increases. If the perturbation is small then the singular value difference is small, and if the perturbation is large, then the singular value difference is large. The difference in singular values of "same" Gabor bit sequences tend to be small as the change in bits is small (2000 bits). While the difference in singular values of "different" Gabor bit sequences tend to be large as the change in bits is large (20000 bits). Based on these results, we can distinguish fingerprints of "same" and "different" speckles and we show the detailed results in Section 5. Another important property that needs to be fulfilled is that the fingerprint has to be compact. To achieve this property, we analyze the scree plots of the singular values of large number of Gabor binary matrices. (Scree plot helps to analyze the relative importance of the singular values and a sharp drop in the plot signals that subsequent singular values can be ignored.) Figure 8, shows the singular values of one representative Gabor binary matrix, and it is clear from the scree plot that the "energy" or magnitude of singular values is clustered in the first few singular values. We choose the 64 and 128 largest singular values as the fingerprints of the speckle pattern. These 64 and 128 singular values hold 88% and 96% of the magnitude of the entire set of singular values respectively. The size of the fingerprint would be 768 digits if we use 128 singular values (using 6 digits of each singular value).

The evaluation of fingerprints across various types of paper are provided in Section 5.

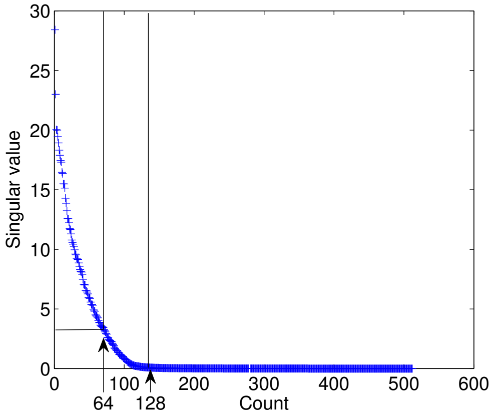


Figure 8: Scree plot of singular values. The first 64 and 128 singular values, which are inside the rectangles, contain 88% and 96% of the magnitude of the set of singular values respectively.

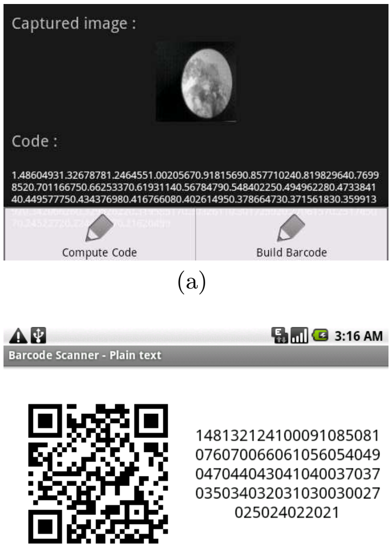
4. IMPLEMENTATION

We have implemented PaperSpeckle on two kinds of setup. One is a desktop/laptop attached with a USB microscope and another is a Google Nexus One mobile phone with a microscope attached to its camera.

In the laptop version we have tested the system with two types of microscopes: Digital Blue QX5 $^{\mathrm{TM}}$and PC Gears AM2011 $^{\mathrm{TM}}$. The detailed evaluation of speckles based on this system are presented in Section 5.

For the mobile phone version, we developed our application on the Android 2.1 platform on the Nexus One. We use standard image capture routines to capture the speckle patterns from the mobile device camera. The mobile phone camera is attached with a consumer grade microscope (Carson MM-200) that has a variable magnification from 10x to 100x, to extract speckles. (We fixed our resolution to 100x). These portable microscopes are low cost devices typically ranging from anywhere between $10 to $50. The fingerprint generation algorithm consists of two mathematically heavy operations: (a) computing Gabor transforms; (b) computing SVD of a large matrix. We implemented a lot of specific optimizations to reduce the compute time of these operations on the Android platform. Specifically, Android is relatively slow for floating point operations in comparison to integer operations; many of the previous Android phone hardware (1.5,1.6 phones) did not even support a hardware floating point unit. We implemented a lightweight floating point library using integer based calculations. While our initial unoptimized version took over 180 seconds of processing time to compute a fingerprint per image, our optimizations reduced the compute time to less than 5 seconds per image. In addition, our code can be easily ported to any Java enabled phone with an inbuilt camera. To support applications in emerging regions, making the system work on low-end mobile devices is essential.

Figure 9a shows the speckle pattern with a menu to generate the fingerprint (menu option: Compute Code) and the corresponding barcode (menu option: Build Barcode). Figure 9b shows the QR code with the corresponding fingerprint. This QR code can be scanned using any standard barcode scanner application on any cellphone to reveal the fingerprint. This fingerprint is compared for a match with the candidate speckle fingerprint that is computed after reading the speckle pattern. Our implementation can be used for: (a) generating fingerprints (2D barcodes) of new speckle images; (b) comparing a fingerprint of a new speckle pattern with a database fingerprints of speckle patterns; (c) reading a new speckle pattern and a barcode separately and comparing the fingerprint of the speckle pattern and the barcode to see if they both match.


Captured image :

Code :

```
1.48604931.32678781.2464551.00205670.91815690.857710240.819829640.7699
8520.701166750.66253370.61931140.56784790.548402250.494962280.4733841
40.449577750.434376980.416766080.402614950.378664730.371561830.359913
```

(a)

```
148132124100091085081
076070066061056054049
047044043041040037037
035034032031030030027
       025024022021
```

(b)

Figure 9: (a) A speckle image taken on the cellphone and the fingerprint of the speckle image; menu options: i) Compute Code, or ii) Build Barcode. (b) QR code of the fingerprint which is shown on the right side.

The mobile phone is attached with a microscope with a different magnification (100x) and field of view than the microscope attached to the laptop. So, the fingerprints generated by this setup would be different than the laptop setup. The microscopes used with the laptop can be varied to extract speckles at 100x magnification, and we use a reduced field of view that is equal in dimensions to the field of view of the mobile microscope, to obtain the same fingerprint as the mobile version.

5. EVALUATION

We evaluate the fingerprints across different types of paper in both ideal and non-ideal conditions.

We considered four completely different categories of paper in our analysis: (a) Letter size print paper; (b) thin notebook style paper (ruled notebooks); (c) Thick print poster-style paper; (d) brown-colored binding/package paper. Also, we tested our results on three different microscopes; two by Digital Blue$^{\mathrm{TM}}$and one by Dino-Lite$^{\mathrm{TM}}$AM2011. The reason for considering different microscopes is to ensure better validity of our results across different microscopes. In this section, we provide important evaluations that stress test the fingerprint under varied conditions. Due to space limitations, we have only presented the results for the fingerprints that represent the 64 singular values. The results for fingerprints that represent 128 singular values are similar to the fingerprints that represent 64 singular values. The evaluations shown here are done on the laptop version of the system (with 200x magnification). We have also evaluated mobile phone version on the four different types of paper (which are stated above) and have found that the results are similar to the laptop version.

5.1 Ideal conditions

In ideal conditions, where the paper was not tampered, we extracted 300 speckles from each type of paper, totalling up to 1200 different speckles. To analyze the variation in fingerprints of speckles of different paper (or region), we made $\binom{300}{2} = 44850$ comparisons. To analyze the variation in fingerprints of "same" speckles , we made 300 pairs of comparisons of each type of speckle. The results are presented in Figures 10a, 10c, 10e. In each figure, the distribution to the left represents the difference in the Euclidean distance of fingerprints of "same" speckles. "Same" speckle means pairs of speckles are compared, where each pair represents two speckle patterns that is extracted from the same region of the paper (multiple measurement of the same speckle pattern). In each figure, the distribution to the left represents the difference in the Euclidean distance of fingerprints of "different" speckles. "Different" speckle means pairwise comparison of speckle patterns extracted from regions of different papers. The two distributions are well separated and any value that is in between the two distributions (greater than the maximum of the left distribution and lesser than the minimum of the right distribution) can be used as a threshold to provide no false positives in identifying or matching two fingerprints. The mean for "same" fingerprints across various types of paper is around 1.5 and the mean for "different" fingerprints across various types of paper is around 85. Based on our analysis in Section 3.2, we know that if the difference (of Euclidean distance of singular values) between a Gabor binary matrix and a randomly perturbed Gabor binary matrix is around 1.5, then the number of bits they disagree is less than 1000. If the difference is above 85, then the number of bits that they disagree is more than 20000. By applying the same analysis to this context, we could say that for the mean of two distributions to meet at least 19000 bits of a Gabor binary matrix have to be modified. Similarly, for a single false positive to occur, at least 9200 bits of a Gabor binary matrix have to be modified.

5.2 Non-ideal conditions

To simulate non-ideal conditions, we tampered the paper in four ways: crumpling, printing or scribbling, soaking in water and aging.

5.2.1 Crumpling

We extracted speckles from 50 pieces of paper and then crumpled each paper thoroughly. Since, the scale at which we operate is around 1 micron, most of the region of interest was not tampered. On an average, 5% of each of the speckle images were modified. This statistic was found out by extracting the 50 speckles again from the same regions and comparing it with the original 50 speckles. Figure 10b, shows the result of comparing "same" speckle on the left and "different" speckle on the right. As we can observe, the two distributions are well separated and the fingerprints can be distinguished with no false positives. For a single false positive to occur, at least 12000 bits of a Gabor binary matrix need to be modified. This shows that PaperSpeckle is able to distinguish fingerprints even when the paper is crumpled.

5.2.2 Printing

Once a speckle is extracted from a region, if the entire region is scribbled or printed, the speckle from the same region would be different when speckle is extracted in the next trial. Ink on the surface of the paper scatters the light in unexpected ways and changes the original speckle pattern. (On the other hand speckle pattern can be extracted after the text is printed or written). In spite of this limitation, our approach sustains printing or scribbling up to 6% to the speckle region. We extracted speckles from 50 papers and then printed or scribbled near the region of extraction. On an average, 6% of each of the speckles were modified. Figure 10d, shows the results of comparing "same" fingerprints (on the left) and "different" fingerprints (on the right). The two distributions are well separated and the fingerprints can be distinguished with no false positives. For a single false positive to occur, the at least 7000 bits of a Gabor binary matrix need to be modified. This shows the even if there is printing or scribbling on the paper, PaperSpeckle is able to distinguish fingerprints.

5.2.3 Soaking in water

We extracted 25 speckles from different pieces of paper and submerged the pieces of paper under water for a few minutes. To make sure we extract speckle from the same region, we used UniBall 207 Gel pen to pigment the papers and mark the region (special ink from the Uniball pen doesn't fade when a paper is soaked in water). After pigmenting the papers, it was submerged in water for a few minutes. Figure 10f, shows the results of comparing "same" fingerprints (on the left) and "different" fingerprints (on the right). The two distributions are well separated and the fingerprints can be distinguished with no false positives. For a single false positive to occur, at least 5000 bits of a Gabor binary matrix need to be modified. This shows that PaperSpeckle is robust and is able to withstand extreme conditions such as water soaking.

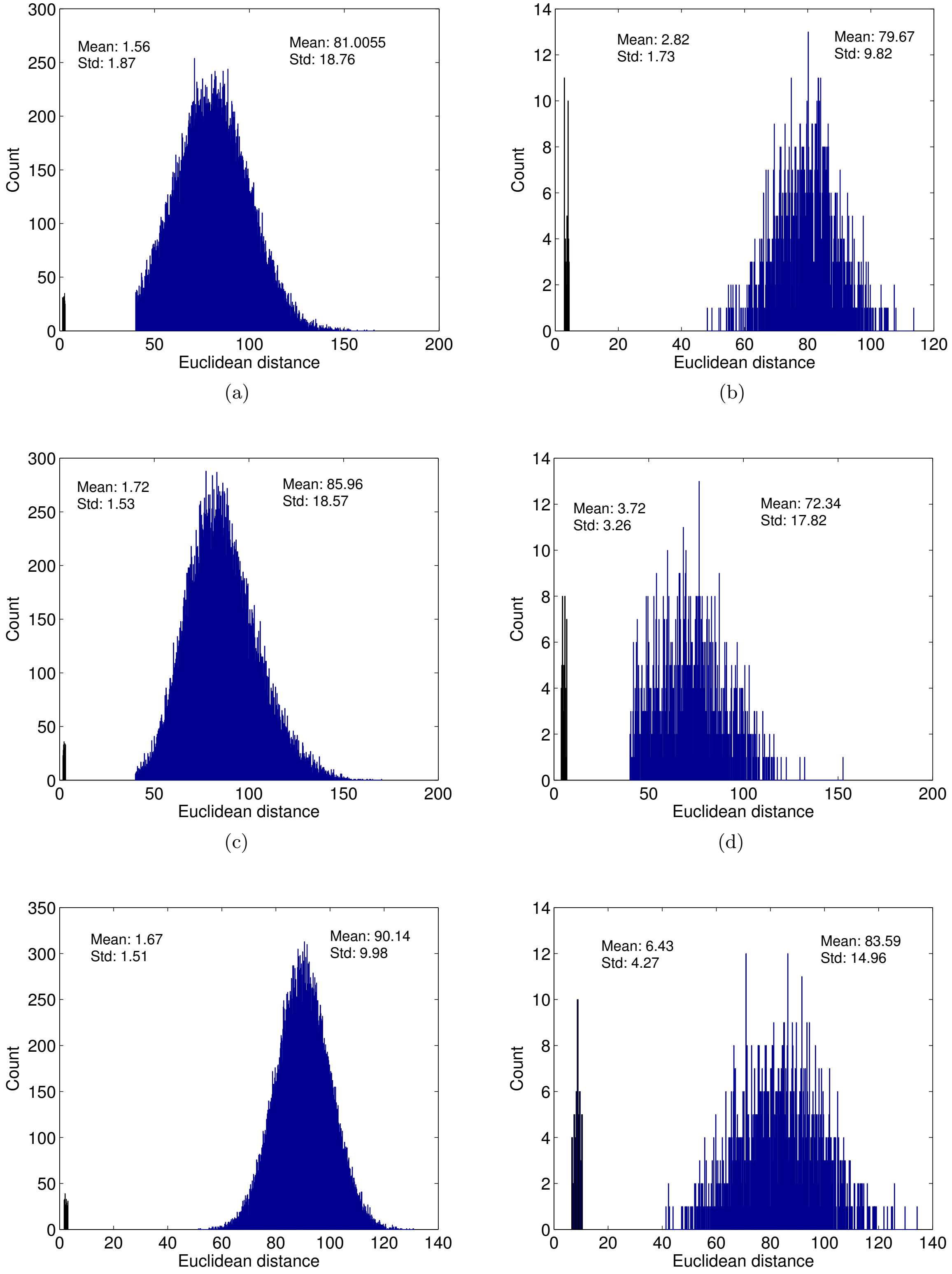
5.2.4 Aging

We extracted 50 speckles from different pieces of paper and stored it for around two years in a storage closet. Then, we extracted the speckles from the same set of papers, generated the fingerprints and compared them. The mean and standard deviation for "same" fingerprints is 1.52 and 1.61. The mean and standard deviation for "different" fingerprints is 87.56 and 12.98. For a single false positive to occur, at least 10000 bits of a Gabor binary matrix need to be modified. This shows that PaperSpeckle works on aged paper without any false positives.

6. APPLICATIONS

PaperSpeckle can be used in a variety of ways where verifying the authenticity of a paper document is of utmost importance.

The main benefit of fingerprints corresponding to a speckle pattern is that the fingerprint can be printed on the same piece of paper containing the paper speckle. Any piece of paper can be made *self-verifiable* by extracting the speckle pattern from a small region on the paper and imprinting the fingerprint of the speckle pattern of the region in the



(a)

(b)

(c)

(d)

(e)

(f)

Figure 10: Histogram of pairwise Euclidean distances between fingerprints. (a) Letter size print paper; (c) thin notebook style paper (ruled notebooks); (e) Thick print poster-style paper; (b) Crumpling. (d) Printing and scribbling. (f) Soaking in water.

same paper. This self-verifiable paper can be used for offline verification of paper documents.

Paper check, receipt or voucher is issued by a trusted authority such as a Bank or Government Authority. The first step of the offline authentication mechanism is the paper check generation. The issuing authority is associated with a cryptographic key pair $K = (P, Q)$, where P is the public-key and Q is the corresponding private-key. While Q is only known to the trusted authority creating the official paper documents, P is known to any entity who wishes to authenticate the paper document in an offline manner. Given any piece of paper, the issuing authority can convert the paper into an authentic offline verifiable paper document using the following paper check generation steps. i) Make the paper self-verifiable by reading a region of the paper, extracting the speckle mark M and imprinting the corresponding fingerprint $D(M)$ computed using any computing device with an attached microscope. ii) The Bank signs $D(M)$ using its private key Q. We denote this as $S_Q(D)$ and the bank prints $S_Q(D)$ on the paper check. In practice, $S_Q(D)$ can be represented as a compact 2D barcode. iii) The paper check contains three artifacts: speckle pattern M, fingerprint $D(M)$, signed number $S_Q(D)$. In practice, this can be extended to authenticate any additional information T embedded such as serial number (for currency notes, checks), transaction details or personal information. To achieve this, we can simply replace $S_Q(D)$ with $S_Q(D, T)$. To authenticate this paper in an offline manner, a mobile device (cellphone) equipped with a microscope, extracts the speckle image M from the region, generates the fingerprint $D(M)'$ of the speckle M and checks if generated fingerprint $D(M)'$ matches $D(M)$ printed on the paper. (If $D(M)$ is represented as a 2-D barcode, then $D(M)$ is read using inbuilt camera in the mobile device coupled with a standard barcode library). If the fingerprints do not match, the paper is a counterfeit. Else, if the fingerprints match, then it is guaranteed that the paper is genuine. Once the paper is determined to be genuine, the next step is to verify whether the fingerprint was signed by the trusted authority (bank). $S_Q(D)$ which is represented as a 2D barcode is read using the barcode scanner in the mobile device. Using the authentic public-key P stored in the mobile device, it is checked if $S_Q(D)$ is a valid signature of $D(M)$.

In a similar fashion, PaperSpeckle can be used to authenticate currency notes, lottery tickets, land records, degree certificates, receipts in microfinance and in other areas where the possession of paper is used as a primary record of ownership.

7. DISCUSSION

In this section, we briefly discuss the robustness of PaperSpeckle in the face of photocopying or fabrication of paper.

Photocopying paper: Photocopying paper will clearly not preserve the speckle pattern in the original paper since the microscopic region of the copied paper is inherently very different. In addition, the speckle region in a paper is in the micrometer range. The field of view of a microscope with 200x magnification is 2mm and each pixel in the image is around 1-2 microns. Even the shape of the contour at a microscopic level may be very different in the photocopied paper. This was observed by Zhu *et. al.* [31] where they state that due to the halftoning effects of the photocopying process, the shape of the contour would differ from the original one. To test this hypothesis, we photocopied 100 speckles and compared their fingerprints. The results we obtained were similar to the results we obtained for "different" fingerprints in Section 5.

Fabrication of paper speckle: Pappu [24] discusses some of the current 3D fabrication techniques. Fabricating paper using photolithography techniques is an expensive process usually in the order of hundreds of millions of dollars [24]. Also, these processes manufacture large number of identical structures and they are not economically suited for producing just one cloned copy of a single microscopic paper structure.

Optical scattering based systems are hard to copy/clone for two reasons [28]. i) The light diffusion obscures the location of the scatterers (scatterers or scattering elements, are particles that scatter light in an object). The state-of-the-art techniques can probe strong diffusive materials only up to a depth of 10 scattering lengths (scattering of light between particles). ii) Even if we know the position of the all scatterers, the precise positioning of large number of scattering elements is very expensive.

Even the forward problem is hard. Given the details of all the scattering elements, computing or simulating the speckle pattern is computationally expensive where the complexity increases exponentially with the number of scattering events [28].

Destroying a speckle pattern: The simplest manner to destroy a speckle pattern is to scratch the speckled region with a sharp object that destroys the surface characteristics of the region of interest. Alternatively, the speckled region can be excessively scribbled upon (using a pen or pencil) thereby changing the boundary specification of a speckled region or one can tear the piece of paper at exactly the speckled spot. Any of these techniques have to be precise since the speckled region is typically very small in diameter (0.5 mm). To enhance the robustness of PaperSpeckle in the face of inadvertent incidents, a simple approach is to speckle a piece of paper at multiple points coupled with their corresponding self-verifiable compact codes of each of the speckle regions. A determined adversary who wishes to destroy a speckled piece of paper has to destroy each of the speckle points in the piece of paper which would essentially render the paper useless.

8. CONCLUSIONS

In this paper, we have presented PaperSpeckle, a low-cost, robust, portable paper fingerprinting system that can identify and authenticate paper. The key contributions of the paper can be summarized as follows. We show how to extract repeatable texture speckle patterns from a region of paper and present an algorithm to generate fingerprint from a region of paper. We provide detailed evaluation of our fingerprinting algorithm across different types of paper and also show how our system is robust against tampering by evaluating texture speckles in adverse environmental conditions. We implement the fingerprinting mechanism on a cellphone and discuss how our system can be used in an offline manner, which has a high potential in mitigating forgery and enhancing physical security of paper in developing regions.

Acknowledgements

We thank Varun Tamminedi, Alpesh Shah and Brinsley D. Cunha for their involvement in the early part of the project (NYU class project report: “Speckles: Low-cost Paper Watermarking”).

9. REFERENCES

[1] Ingenia technology ltd. http://www.ingeniatechnology.com/.

[2] Telgi Scam. http://www.financialexpress.com/news/telgi-scam/83736/.

[3] J. D. R. Buchanan, R. P. Cowburn, A.-V. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D. A. Allwood, and M. T. Bryan. Forgery: ’Fingerprinting’ documents and packaging. *Nature*, 436:475, July 2005.

[4] William Clarkson, Tim Weyrich, Adam Finkelstein, Nadia Heninger, Alex Halderman, and Ed Felten. Fingerprinting blank paper using commodity scanners. In *IEEE Security and Privacy*, 2009.

[5] J.G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7):1169–1179, Jul 1988.

[6] John Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, February 2003.

[7] N. Salzman E. Metois, P. Yarin and J. R. Smith. Fiberfingerprint identification. In *Third Workshop on Automatic Identification*, 2002.

[8] Borowski Jr. et. al. Surface treated security paper and method and device for producing surface treated security paper, US Patent Number 5,193,854, 1993.

[9] E. B. Greene et al. Coatings and ink designs for negotiable instruments, us patent number 6,155,604, 2000.

[10] Kailath T et al. Multivariable control, simulation, optimization and signal processing for the microlithographic process, 1997.

[11] K. L. van der Molen F. van Beijnum, E.G. van Putten and A. P. Mosk. Recognition of paper samples by correlation of their speckle patterns. *Arxiv.org preprint physics/0610089*, 2006.

[12] Lin Feng. US Patent Application Number: 20100067691, Document certification and authentication system, 2010.

[13] H. Fujii and T. Asakura. A contrast variation of image speckle intensity under illumination of partially coherent light. *Optics Communications*, 12(1):32 – 38, 1974.

[14] Nicholas George and Atul Jain. Speckle reduction using multiple tones of illumination. *Appl. Opt.*, 12(6):1202–1212, Jun 1973.

[15] J. W. Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America (1917-1983)*, 66:1145–1150, 1976.

[16] E. B. Greene. Negotiable instrument, us patent number 4,634,148, 1987.

[17] Sheng Huang and Jian Kang Wu. Optical Watermark (WO/2002/023481), 2002.

[18] Kimura and Yoshihiro. Woven security label, us patent number 6,068,895, 2000.

[19] Roger D Melen. US Patent Number: 5325167, Record document authentication by microscopic grain structure and method, 1994.

[20] Margarita Osadchy, David W. Jacobs, and Michael Lindenbaum. Surface dependent representations for illumination insensitive image comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:98–111, January 2007.

[21] Christos H. Papadimitriou, Hisao Tamaki, Prabhakar Raghavan, and Santosh Vempala. Latent semantic indexing: a probabilistic analysis. In *PODS ’98: Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 159–168, New York, NY, USA, 1998. ACM.

[22] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. Physical One-Way Functions. *Science*, 297:2026–2030, September 2002.

[23] G. Parry. Speckle patterns in partially coherent light. In *Laser Speckle and Related Phenomena*, volume 9 of *Topics in Applied Physics*, pages 77–121. Springer Berlin / Heidelberg, 1975. 10.1007/BFb0111437.

[24] Pappu Srinivasa Ravikanth. *Physical one-way functions.* PhD thesis, 2001. Chair-Stephen A. Benton, Massachusetts Institute of Technology.

[25] Joshua R. Smith and Andrew V. Sutherland. Microstructure based indicia. In *Proceedings of the Second Workshop on Automatic Identification Advanced Technologies*, pages 79–83, New York, NY, USA, 1999. ACM.

[26] G. W. Stewart. *Matrix Algorithms: Volume 2 Eigensystems.* SIAM, 2001.

[27] J. Swartz T. Pavlidis and Y. P. Wang. Fundamentals of bar code information theory. *Computer*, 23(4):74âĂŞ85, 1990.

[28] P. Tuyls, B. Skoric, T. Akkermans, W. Ophey, and S. Stallinga. Security analysis of physical uncloneable functions, 2004.

[29] R. L. van Renesse. *Optical Document Security, Second Edition.* Artech House, Inc, Norwood, MA, 1998.

[30] Daniel; Zeira, Eitan; Ellett. Verification methods employing thermally–imageable substrates, us patent number 6107244, August 2000.

[31] Baoshi Zhu, Jiankang Wu, and Mohan S. Kankanhalli. Print signatures for document authentication. In *ACM CCS ’03*, pages 145–154, New York, NY, USA, 2003.