ON THE INVERSION OF BIOMETRIC TEMPLATES BY AN EXAMPLE

Vanessa Testoni† and Darko Kirovski‡

† DECOM-FEEC, State University of Campinas, Campinas, SP 13083-852, Brazil
‡ Microsoft Research, One Microsoft Way, Redmond, WA 98033, USA

ABSTRACT

In this paper we analyze practical issues related to an adversarial inversion of biometric templates constructed without any cryptographically-secure protection features. The inversion in most cases is considered an ill-defined problem as the template captures only a small subset of the physical presence of a specific biometric trait. We apply our practical approach to an existing iris-based biometric system and demonstrate that “inverted” iris images pass matching tests even when encoded like barcodes.

Index Terms—Biometric template inversion

1. INTRODUCTION

Biometric features are physical traits of humans used for automated authentication. In recent history biometric systems, although abundant in research proposals [1], have seen slow adoption rates for many straightforward applications: from identity documents to automated access control. The reasons lie within several disadvantages: privacy concerns related to automated tracking of individuals enrolled in and checked by the system, dissatisfactory performance reflected in high false positives and false negatives rates that make enrollment of individuals in large populations an exceptionally difficult task, system costs, the inconvenience of scanning biometric traits, the fact that biometric features cannot be revoked, etc.

Here we are concerned about biometric skimming. Skimming is a term common in credit card fraud, where an adversary obtains data stored on the credit card’s magnetic strip, then manufactures its exact replica, and uses it fraudulently. In biometrics, the adversary obtains either:

A. a hi-fidelity scan of target’s biometric trait – this can be achieved using a “fake” or tapping into a “legitimate” scanning station or using sophisticated sensors, or

B. its template – in addition to the techniques used in case A, here the adversary could access a biometric template when it is stored in or communicated to/from a centralized database,

then proceeds to produce a physical representation which, once scanned, is marginally different from the original template and/or templates derived by scanning the target trait.

1.1. Preliminaries

Given an instance of a physical trait, a biometric system first obtains its digital scan, x. Then, the system computes a biometric template, t = f(x), using a function f(·) conceived to isolate and encode unique features of the human trait in a compact form, remove measurement variability via denoising, normalization, and alignment to a reference perspective, and minimize the probability of a false positive for a given probability of a false negative. For two different scans, x1 and x2, of the same trait, the templates are comparable in a sense that a specific well-designed distance metric d(f(x1), f(x2)) returns a relatively small scalar. Alternatively, d(f(x), f(y)), where scan y is an image obtained by scanning another trait instance, returns a relatively large scalar opening up a stage for a classic Neyman-Pearson hypothesis test – given two templates, establish a threshold d(t1, t2) ≤ θ at which they are considered to originate from the same biometric source.

In an adversarial scenario, the malicious party obtains either x or t = f(x) of a person already enrolled in the biometric system. In order to impersonate this individual, the adversary creates a “forged” biometric instance which, when scanned as xR, results in d(t, f(xR)) ≤ θ with an overwhelming probability; t denotes the registered template of the target. “Forging” a biometric template is a task which has already been achieved for numerous traits [2] [3] [4] with a notable example of “gummy fingers” [5]. The key problem in spoof detection is the fact that the adversary has the last word in the game – thus, it is realistic to assume that the adversary has an instance of the scanning hardware, can reverse engineer it, and gradually work out a technology that would fool the spoof detector. In such a setting it is unrealistic to assume that the detector will be resilient to biometric forgeries.

Cryptographically-strong techniques such as fuzzy vault [6] (reminiscent of a secret sharing protocol, is efficient on biometric data represented as sets, thus, common for fingerprint minutiae) and fuzzy extractors [7] aim to alleviate the problem with significant entropy loss and increase in false positives – both effects are difficult to survive considering the low entropy of biometric data acquired using modern sensing technologies. In addition, they only aim to de-randomize the biometric input as well as protect it from public display during both storage and matching, but do not prevent skimming.
In this paper, we analyze the general case when the adversary has access to a target \( t \) and \( f() \) represents a traditional feature extraction function. Even in the case when the adversary has \( x \), the attack route presented in this paper should have better performance for “fixed cost forging” because the forgery is optimized to the underlying \( f() \). Finally, since fuzzy extractors are used as post-processing to \( f() \), the process that we discuss in this paper, is essential in the inversion under fuzzy extractors as well – note that we are not interested in deriving \( x \) but any \( x_F \) that has \( d(f(x), f(x_F)) \leq \theta \).

An important feature of commonly deployed functions \( f() \) is that obtaining \( x \) from \( t \) is an ill-posed problem [8] as small or no noise in \( t \) could easily result in inverses that are entirely different from \( x \). We show that this is certainly not a problem for the adversary, as a matter of fact, we use this freedom to show that inverses represented with low resolution and very few quantization levels, thus requiring low-quality forgery equipment, could provide satisfactory attack performance. As a corollary, we argue that certain simple directions in computing biometric digests could make skimming more difficult.

2. AN EXAMPLE: IRIS TEMPLATES

In this paper, due to their relative efficiency, we focus on iris-based biometric systems and a common tool for computing biometric digests from iris scans, a publicly-available variant [9] of the Daugman Gabor-wavelet-based algorithm [10] [11].

2.1. Computing Iris-based Templates

Human irises are scanned in the infrared light spectrum for several reasons: reduction of glare, less noise in the reflectivity of human tissue, and invisible capture [10]. Once an iris image is captured, the system computes its biometric template as illustrated in Figure 1. We stress that the most widely used approach for computing iris templates is based upon Daugman’s algorithm [10]. However, since we are not aware of any publicly available implementation, we adopted Masek’s variant which has certain resemblances to Daugman’s work [9]. For details on the two algorithms, we refer the reader to review their technical reports. We continue with a brief description of the generic design of an iris template.

The first objective is to segment the iris region from the eye image. The method implemented in [9] is based on the Hough transform [12]. The purpose of this technique is to find imperfect instances of objects within a certain class of shapes by a voting procedure. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in a so-called accumulator space. The transform identifies oval areas that represent the iris and the pupil, and isolates noise such as occluding eyelids and eyelashes. The isolated iris region is then sampled using a polar coordinate grid and normalized into a fixed-size rectangular block of samples. The area of the rectangle which corresponds to the detected noise (e.g., eyelids, eyelashes) is masked using a binary mask \( m \). Exemplary values for the two matrices, \( x \) and \( m \), are illustrated in Figure 1.

The next step is to extract the iris texture. Here, Gabor filters are a traditional choice for obtaining localized frequency information. The 2-D Gabor filter can provide information about the positioning and frequency content of local image structures [10]. The 2-D Gabor function in space domain \( (x, y) \) has the functional form: \( g(x, y) = e(x, y) \cdot c(x, y) \), where \( e(x, y) \) is a 2-D Gaussian, denoted as an envelope, and \( c(x, y) \) is a complex sinusoid, denoted as a carrier. In [10], the Gaussian envelope is defined as:

\[
e(x, y) = \exp\left\{-\pi[(x - x_0)^2/\alpha^2 + (y - y_0)^2/\beta^2]\right\}, \quad (1)
\]
where \((x_0, y_0)\) specifies the peak of the envelope and parameters \(\alpha\) and \(\beta\) scale the two axes. The envelope assumes a circular shape for the following aspect ratio \(\beta/\alpha = 1\). The envelope defined in Eqn. 1, has a normalized magnitude and zero-rotation angle.

The complex sinusoid carrier is defined in [10] as:

\[
c(x, y) = \exp \left\{ -i2\pi (u_0(x - x_0) + v_0(y - y_0)) \right\},
\]

where \((u_0, v_0)\) are the spatial frequencies of the sinusoid carrier in Cartesian coordinates. The initial phase shift of the sinusoid carrier defined in Eqn. 2 is zero. The spatial frequency can also be defined in polar coordinates as \(\omega_0 = \sqrt{u_0^2 + v_0^2}\) with direction \(\theta = \arctan(v_0/u_0)\).

An alternative to the Gabor function is the log-Gabor function which is applied in [9]. While Gabor filters have Gaussian transfer functions when viewed on the linear frequency scale, log-Gabor filters have Gaussian transfer functions when viewed on the logarithmic frequency scale. The advantages are that Gabor filters do not have a DC component and that the transfer function of the log-Gabor filter has an extended tail at the high frequency end, which represents better visual content rich in textures and details, such as the iris images. The transfer function of the 1-D Log-Gabor filter on the linear frequency scale is:

\[
G(w) = \exp \left\{ -\frac{\log^2(w/w_0)}{2\log^2(k/w_0)} \right\},
\]

where \(w_0\) is filter’s center frequency; ratio \(k/w_0\) must be kept constant for varying \(w_0\). This filter is applied in the frequency domain for each row of the normalized iris in [9]. The inverse Fourier transform is then applied to the filtered iris and the sign of both the real and imaginary parts is used to compute the biometric template. Thus, for each row of samples \(x_r\) in the segmented rectangle \(x\) we compute:

\[
t_{rc} = F^{-1} \{ F(x_r) \cdot G \}
\]

where \(G\) denotes the 1-D log-Gabor filter as defined in Eqn. 3, and \(F()\) denotes a fast Fourier transform. The result, \(t_r\) (which is twice the length of \(x_r\) because it represents the signs of the real and imaginary parts of \(t_{rc}\), is then placed in the appropriate row of the resulting matrix \(t\). In [9], the authors mask out samples in \(x_{rc}\) whose magnitudes are smaller than a specific threshold. We will denote the overall mask as a binary matrix \(m\).

Finally, to compute a match between two templates, \(\{t_1, m_1\}\) and \(\{t_2, m_2\}\), one computes their relative Hamming distance in a hypothesis test:

\[
d(\{t_1, m_1\}, \{t_2, m_2\}) \equiv \frac{m_1 \cdot m_2 \cdot H(t_1, t_2)}{m_1 \cdot m_2} \leq \theta,
\]

where \(\theta\) is the detection threshold empirically set to a value that balances false positives and false negatives.

### 2.2. The Inversion

Considering that the adversary has access to \(\{t, m\}\), our first step in creating a “fake” iris would be to recover the filtered normalized iris values \(x\) from the template. Clearly, it is not possible to recover the exact values of the original data, \(x\) because the template only stores the signs of the real and imaginary parts of the filtered iris data, \(t_{rc}\). Thus, we adopt the following per-row procedure:

\[
y_r = A F^{-1} \{ G^{-1} \cdot F [\text{Re}(t_r) + \text{Im}(t_r)i] \},
\]

where a real scalar \(A\) scales result’s amplitude and \(\text{Re}(t_r) + \text{Im}(t_r)i\) recreates a complex number whose real and imaginary parts belong to \(\{0, \pm 1\}\) depending on the sign of the corresponding parts of \(t_{rc}\). When we apply \(F\) in Eqn. 7, we fill the DC component and the second half of \(t_r\)’s spectrum with low-magnitude complex noise to reproduce the effect of the log-Gabor filter.

As log-Gabor filters are not perfectly invertible due to losses in the high and near-zero frequencies, we apply the simple inverted transfer function \(G^{-1}\) and fill the two lowest frequency samples (including the DC) and the final two-thirds of the resulting spectrum with low-magnitude complex noise. Finally, we apply the inverse FFT to obtain the “forged” normalized iris \(y_r\) in the spatial domain.

By construction, if we reapply Eqs. 4 and 5 to \(y_r\), we are expected to nearly re-compute \(t_r\). The magnitude \(A\) is selected so to achieve an identical mean signal value between \(x\) and \(y\). One disadvantage of using \(y\) in this format is the precision in printing that needs to be achieved to accomplish its nuances. We relax this requirement by applying a non-uniform \(\mu\)-law quantizer with \(Q\) quantization values; typically \(Q \in \{2, 4, 8, 16\}\).

Finally, we produce the final “forged” iris image formatted as a radial barcode (see Figure 1). We set our quantization levels so to mimic the histogram of values represented in \(x\). Note that if \(x\) is not available to the adversary, it is reasonable to assume that she will have sufficient knowledge about the system to derive proper quantization levels. Examples of barcodes reconstructed at \(Q \in \{2, 4, 8, 16\}\) for a specific iris image are illustrated in Figure 1. We stress that from our experiments the lowest number of quantization levels that achieved successful results was 16; then, the relative Hamming distances between templates computed for \(x\) and \(y\) were on the average lower than 0.1.

### 3. EXPERIMENTAL RESULTS AND CONCLUSION

In order to evaluate the efficiency of our adversarial technique, we applied the attack to the CASIA iris database [13]—“Portions of the research in this paper use the CASIA-IrisV3 collected by the Chinese Academy of Sciences’ Institute of Automation (CASIA).” The database consists of 3-4 image
Fig. 2. Histograms for relative Hamming distance for a pair of templates coming from: A – two distinct, authentic scans of the same iris, B – one authentic scan and its inverse, C – one authentic scan and an inverse computed over a scan of the same iris but different from x, and D – two authentic scans of two different irises.

left-right eye scans for 100 subjects. We present the experiments using two plots. The first, shown in Figure 2, presents four histograms for \( d(\{t_1, m_1\}, \{t_2, m_2\}) \) for a pair of templates coming from: A – two distinct, authentic scans of the same iris, B – one authentic scan and its inverse, C – one authentic scan and an inverse computed over a scan of the same iris but different from x, and D – two authentic scans of two different irises.

In Figure 2 the C-histogram is only slightly shifted towards higher Hamming distance values compared to the A-histogram; also its overlap with the impostor D-histogram is minor. This indicates that the attack is likely to be successful in-field as the noise that impacts the detection statistic is marginal due to our attack. Figure 3 presents these results in the roc-format, where one can observe that the probability of a forgery failure is only slightly higher than the probability of a false negative in an authentic test.

We point the reader to the fact that once digitally “forged” an iris still needs to be printed using infrared ink on a contact lens. When \( x \) is of size 20 × 240, printing the barcode at 600dpi is likely sufficient to excite the correct response in the iris scanner. We did not execute this final step because all commercial iris scanners deploy Daugman’s algorithm [10], which is patented, thus, we would need to illegally reverse engineer a commercial scanner to launch the attack.

The attack could be prevented/made more difficult in several ways: a) by asking the subject under test to remove contact lenses – a task that substantially raises the inconvenience of the biometric test, b) by performing a liveness detection within the scanner by enforcing and detecting pupil dilation as proposed in [10] – a task that is non-trivial to execute and may not be fully successful under certain conditions such as the usage of mydriatics, etc., c) by substantially improving the level of detail captured from the human iris and represented by an iris template – tools such as super-resolution [14] may prove to be fundamental for this application however at a performance hit difficult to overcome especially for a “1-many” test typical for biometric enrollment.

4. REFERENCES