Abstract—The need of sharing fingerprint image data in many emerging applications raises concerns about the protection of privacy. It has become possible to use automated algorithms for inferring soft biometrics from fingerprint images. Even if we cannot uniquely match the person to an existing fingerprint, revealing their age or gender may lead to undesirable consequences. Our research is focused on de-identifying fingerprint images in order to obfuscate soft biometrics. In this paper, we first discuss a general framework for soft biometrics fingerprint de-identification. We implemented the framework to reduce the risk of successful estimation of gender from fingerprint images using ad-hoc image filtering. We evaluate the proposed approach through experiments using a data set of rolled fingerprints collected at West Virginia University. Results show the proposed method is effective in preventing gender estimation from fingerprint images.

Index Terms—Image De-Identification, Gender Estimation, Fingerprint Recognition.

I. INTRODUCTION

Automatic fingerprint recognition has been widely adopted in biometric and forensic applications [1]. Unfortunately, it is not uncommon to encounter an application in which personal information is stored without subject’s approval or control [2]. The threat to privacy arises from the ability of a third party to access the information and use it without authorization. An identifiable fingerprint retains personal information about an individual. It also creates the potential for an invasion of privacy by linking together personal information about an individual from various sources. It is becoming important to control who has personal/private information about us. Currently, for example, an important challenge in biometrics is the ability to release data sets for research and testing purposes that guarantee confidentiality to the participating subjects.

De-identification is a required element of information integration in order to reduce the risks of unauthorized disclosure. The implicit goal of de-identification is to protect privacy and preserve data utility [3]. In de-identification, data content may be altered to remove or obscure personal information which would allow identification of individuals who are the source of the data. Image de-identification refers to the process of depriving the biometric data of information which would be useful to identify the source of the data. In many real-world applications, de-identification allows the release of data without sensitive information.

De-identification of fingerprint images can be performed in several ways. The first assumes that the image is transformed in such a way that it cannot be matched against the same person’s fingerprint in the gallery. The utility of this approach for biometrics is minimal, as the primary purpose of fingerprinting, identification, is eliminated. The second approach assumes that both the probe fingerprint image and the matching gallery image are transformed in the same non-obvious way. As the result, the fingerprint representations - the templates, still match, but their form does not reflect the appearance of the original fingerprint images thus protecting the identity of the person to those who may be able to penetrate the system during the operation. A good example of this approach are cancelable fingerprint templates [4]. The third approach, useful for releasing research data sets, does not significantly change the appearance of fingerprint images, but subtle changes disallow the successful extraction of soft biometrics. A critical step in this type of fingerprint image de-identification is the preservation of reliability of minutiae extraction [5]. If the ridge structure is not well-defined minutiae cannot be appropriately detected. Any corruption of the ridge structure while attempting de-identification could cause an undesired loss of the individuality in fingerprints as it may induce variations in the position and orientation of minutiae and / or introduce spurious minutiae. Therefore, a good de-identification algorithm is the one that will minimize performance degradation of biometric matching, while maximizing the difficulty of the extraction of soft biometric identifiers.

To the best of our knowledge, no previous work exists on de-identification of fingerprint images for the purpose of removing soft biometric identifiers. In this paper, we propose a method for reducing the ability to estimate gender from fingerprint images, while not degrading the matching performance. This technique is based on ad-hoc image filtering, attenuation or amplification of discriminative frequency components. Additionally, we propose a scheme to assess the impact of de-identification mechanisms on fingerprint recognition systems.

The paper is organized as follows. In Section II, we describe the state-of-the-art pertaining to biometric image de-identification. Section III presents the proposed approach for disabling gender estimation from fingerprint images. Section IV discusses the evaluation procedure while Section V presents experimental results. We draw conclusions in Section VI.

II. RELATED WORK

Soft biometrics, such as gender, ethnicity, age, height, weight and eye color, provide ancillary information about user’s identity, which is not discriminative or permanent enough to differentiate two individuals. Automatically extracted soft biometrics can be used during the decision making process in order to enhance the overall performance of the system [6]. Several recent studies examine differences in fingerprint images between genders. The most notable differences in the spatial domain appear to be related to ridges. It has been observed that females typically have a higher ridge density compared to males, due to finer epidermal ridge details. Ridge density is defined as the number of ridges within a certain space in fingerprint image [7]. In [8] and [9] fingerprints are classified based on gender using statistics such as white lines count and ridge count that are manually extracted. In 1999, Acree manually counted ridges in a well-defined area of fingerprints [10]. Recently, gender estimation has

been performed by a method based on discrete Wavelet Transform DWT and Singular Value Decomposition SVD [11].

Most of the literature on biometric image de-identification focuses on human faces. Existing approaches belong to two main categories, i.e., ad-hoc distortion/suppression techniques or formal methods such as K-Same [12]. In image distortion-based approaches for face de-identification the region occupied by the face is altered using image filtering or data suppression. Newton et al. [13] proposed the K-Same algorithm which scientifically guarantees protection of privacy when sharing video data. It preserves several facial details but assures low reliability of face recognition systems when operating on de-identified face images. Later, Gross et al. proposed an extension of the K-Same algorithm, which takes into account linear appearance variations of faces [14]. In their work, face de-identification is not performed in the image space but in that of the model parameters in which the face is represented through Active Appearance Model (AAM).

III. THE PROPOSED APPROACH

In the proposed framework, the soft biometrics de-identifier module operates on the original image and it outputs a de-identified image. Features for matching are extracted from de-identified images, as shown in Fig. 1.

A. Automatic Gender Estimation from Fingerprints

In this study, we used the automatic gender estimator recently proposed by Marasco et al. [15]. The features exploited by this algorithm are described below.

Local Binary Pattern (LBP). Texture is characterized by the spatial distribution of gray levels in a neighborhood. Its details cannot be captured in a single point. LBP characterizes the spatial configuration of local image texture. The texture is defined as the joint distribution of gray values of a circularly symmetric neighbor set of P image pixels on a circle of radius R (Eqn. (1)):

\[ T = t(g_c, g_0, \ldots, g_{P-1}) \]  

where \( g_c \) is the gray value of the center pixel of the local neighborhood and \( g_0, \ldots, g_{P-1} \) are the gray values of \( P \) equally spaced pixels on the considered circular symmetric neighbor set. First, the gray value of the center pixel of the circularly symmetric neighbor set is subtracted from the gray values of the circularly symmetric neighborhood. Then, features are extracted from histograms of LBP images computed as indicated in Eqns. (2) [16] [17]:

\[ LBP_{P,R} = \sum_{p=0,\ldots,P-1} s(g_p - g_c)2^p \]  

where \( s \) is defined as: \( s(x) = 1 \) if \( x \geq 0 \), else \( s(x) = 0 \).

Local Energy Concentration. The energy concentration is computed as the square magnitude of the Fourier spectrum. Local information can be obtained by computing the energy concentration as ring feature for different equally spaced frequency bands of the Fourier spectrum. Each ring was constructed by computing the difference of two equally spaced low-pass Butterworth filters [18]. The considered Region Of Interest to extract ring features ranges from 50 Hz to 250 Hz. Frequency components of the Fourier transform are related to spatial characteristics of the image. In particular, frequency expresses the rate of change in the image such that higher frequency components pertain to patterns in the image corresponding to occurrence of larger variations in grey-level intensity. Given an image \( I(x, y) \), the Discrete Fourier Transform (DFT) of size \( M \times N \) can be computed as follows:

\[ F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) e^{-i2\pi\left(\frac{xu}{M} + \frac{yu}{N}\right)} \]  

where \( i = \sqrt{-1} \), \( x \) and \( y \) are spatial variables, \( I(x, y) \) represents the gray-level intensity value at pixel \((x, y)\) of the image, \( u \) and \( v \) are frequency variables [19].

Local Phase Quantization (LPQ). Previous analysis focused on simply analyzing the magnitude spectrum of a fingerprint image. Given that the fingerprint can present different orientations, a rotation invariant Local Phase Quantization (LPQ) technique can point out the differences in gender in the phase spectrum. A novel set of features based on the LPQ of the fingerprint image has been defined in [20] [21]. The approach measures the difference between male and female fingerprints in terms of high frequency information loss.

Image Quality measures the degree of usefulness of a biometric sample for automated recognition. The quality of captured biometric data directly impacts the effectiveness of the matching process. Gother and Tabassi discussed the concept of predicting error rates based on quality values [22], as an indicator of matchability.

B. Fingerprint De-Identification for Gender Estimation

The proposed approach is based on image filtering in the frequency domain. The general model used for filtering the image in the frequency domain is the following:

\[ G(u, v) = H(u, v)F(u, v) \]  

The considered transfer function \( H(u, v) \) is a zero-phase-shift filter and it affects real and imaginary parts of the Fourier Transform without altering its phase. It is constructed by using differences of two Butterworth functions as described in the previous subsection. This procedure allows for a band-specific scaling in the frequency domain. This linear filtering process applies blurring by attenuating high-frequency content or sharpening by increasing the magnitude of high-frequency components. Certain frequency components in the Fourier Transform are suppressed, while others are amplified. The Fourier coefficients of each band are multiplied by a scalar chosen ad-hoc for the specific band based on the energy distribution. The de-identified image is then obtained using the inverse of the Fourier Transform after suppression. The constant selected for the scaling varies according to the energy distribution pertaining to the given band. The main steps of the proposed approach are summarized by Table A.

IV. ASSESSMENT

The challenge of integrating a de-identified algorithm in a biometric system is to decrease the performance of soft biometrics estimators without compromising the correctness of the matching operation. The performance of the overall fingerprint system is obtained using metrics of a typical biometric systems, i.e. False Match Rate (FMR) and False Non-Match Rate (FNMR), when matching de-identified images. The de-identification algorithm can be evaluated by considering the performance of both the soft biometric estimator, which is desired to fail, and the biometric matcher which is desired to be accurate. At de-identification time, two types of errors can occur:

- Undermarking: failure to remove information as is required to inhibit inference of soft biometrics from the image. The True Soft Biometric Detection Rate (TSDR) measures this errors. It is defined as the proportion of de-identified images from which the estimated soft biometric characteristic is the correct one.
4. Compute scaling parameters $a_k$, $k = 1, \cdots, B$

\begin{align*}
    \mu_{k, w_0} &= \text{mean}(E_{k, w_0}) \\
    \mu_{k, w_1} &= \text{mean}(E_{k, w_1}) \\
    a_k &= (\mu_{k, w_0} - \mu_{k, w_1}) / 2
\end{align*}

5. Apply the scaling in the frequency domain:
   \begin{align*}
   \text{If } & w_1 \quad F^*(u, v) = (a_k + 1)F(u, v) \\
   \text{else } & F^*(u, v) = a_kF(u, v)
   \end{align*}

6. Compute the inverse of $F^*(u, v)$.

- **Overmarking**: removal of more information than is required. This error compromises the ability of appropriately matching de-identified images. When this error occurs, it does not matter if the de-identification technique succeeds or not, it causes an increase of the FNMR.

The de-identification system is expected to operate at an operating point which guarantees the best *trade-off* between them. There are three possible different scenarios for assessing the matching performance after de-identifying the images: gallery vs. probe image: a) original vs. de-identified; b) de-identified vs. original; c) de-identified vs. de-identified. In this study, we consider the case in which the algorithm for automatic de-identification is integrated in the sensor and all the acquired images are de-identified; this corresponds to the third scenario described above.

**Table A. Fingerprint De-Identification for Gender Estimation.**

**Input:**
Let $I(x, y)$ be the original fingerprint image.
Let $B$ be the number of frequency bands.
Let $w_0$ and $w_1$ represent the male and female classes, respectively.

**Output:**
$DI(x, y)$ de-identified image.

1. Compute the Discrete Fourier Transform $F(u, v)$ of the image $I(x, y)$.
2. Filter the image in the frequency domain:
   \begin{align*}
   G_k(u, v) &= H_k(u, v)F(u, v), \quad k = 1, \cdots, B
   \end{align*}
3. Estimate the energy distributions for $w_0$ and $w_1$:
   \begin{align*}
   E_{k, w_0} &= |G_k(u, v)|^2 \\
   E_{k, w_1} &= |G_k(u, v)|^2, \quad k = 1, \cdots, B
   \end{align*}
4. Compute scaling parameters $a_k$, $k = 1, \cdots, B$
   \begin{align*}
   \mu_{k, w_0} &= \text{mean}(E_{k, w_0}) \\
   \mu_{k, w_1} &= \text{mean}(E_{k, w_1}) \\
   a_k &= (\mu_{k, w_0} - \mu_{k, w_1}) / 2
   \end{align*}
5. Apply the scaling in the frequency domain:
   \begin{align*}
   \text{If } & w_1 \quad F^*(u, v) = (a_k + 1)F(u, v) \\
   \text{else } & F^*(u, v) = a_kF(u, v)
   \end{align*}
6. Compute the inverse of $F^*(u, v)$.

**V. Experimental Results**

**A. Dataset**

The data set used in this study consists of fingerprints from 100 users, which is a subset of a collection consisting of 494 users carried out at West Virginia University. Fingerprints were acquired using a live-scan optical sensor with resolution factor of 500 dpi. Users provided two sets of fingerprints, in sequence, each consisting of rolled individual fingers on both hands, left slap, right slap, and thumbs slap. Fingerprints were collected without controlling the quality in acquisition, i.e., no fingerprint images were rejected and recaptured at that stage. Match scores between all image pairs were generated using the Identix BioEngine Software Development Kit. The NIST Biometric Image Software (NBIS) \(^2\) (the NFIQ function) was used to evaluate fingerprint image quality. NFIQ measures vary between 1 and 5, with 1 being the highest quality and 5 the lowest. Recent studies highlighted that NFIQ is effective in discriminating between good and bad quality images [23].

**B. Evaluation Procedure**

Experiments were carried out using a fingerprint gender estimator based on a K-Nearest Neighbor (KNN) model, where $K = 1$. The model was trained by performing a 10-Fold Cross Validation on the entire data set. This gender estimator provides an accuracy of 88.7% on original (unaltered) fingerprint images. The algorithm described in Table A was used to de-identify fingerprints. The gender estimator was then tested on the set of de-identified fingerprints. The Weka 3.6 software was used for the KNN classifier.

Match scores were computed in the following two scenarios: i) Unaltered gallery and probe images; ii) De-identified gallery and probe images.

**C. Results**

- **Visual Impact of Fingerprint De-Identification**. Fig. 2 shows that, visually, the impact of the de-identification process on the fingerprint images is not pronounced.

- **Gender Estimator Accuracy on De-Identified Fingerprint Images**. Fig. 3 illustrates the energy distributions for two different frequency bands before and after de-identification. Frequency components in original images separate females from males well. Conversely, corresponding frequency components pertaining to de-identified images from females and males overlap, therefore decreasing the discriminative power of ring features. Similar behavior was observed for the remaining bands. This

\(^2\)http://www.nist.gov/itl/iad/ig/nbis.cfm
Fig. 2. Fingerprint images before performing image de-identification ad-hoc for automatic gender estimation: (a) Fingerprint image pertaining to a female subject. (b) Fingerprint image pertaining to a male subject.

Fig. 3. Energy distribution across different frequency bands before and after performing de-identification. The selected frequency bands originally are discriminative with respect to gender but such a discriminative power is highly compromised when they refer to the de-identified version of those images.

- Matching Performance on De-Identified Fingerprint Images. Fig. 4 quantifies the effectiveness of the proposed de-identification algorithm with respect to matching performance. As desired, the variations induced in the images do not drastically affect verification error rates.

VI. CONCLUSIONS

Given the large number of applications which require sharing of data, de-identification is becoming an essential element for reducing the risk of unauthorized disclosure. In this paper, we used an image processing technique to hide information pertaining to gender in fingerprints. Images are analyzed in the Fourier domain where components in specific bands are attenuated or amplified based on their discriminative power. The proposed approach guarantees accurate matching between de-identified images. The discussed method does not fit a formal privacy model; subsequently, it does not guarantee that the privacy is actually protected. Reported results are specific for the gender estimator used to carry out experiments.

Future directions may include: i) Designing a suitable metric to measure the distance between the original image and the de-identified one to determine the degree of de-identification applied to the image. ii) Evaluating the robustness to potential re-identification attacks.

REFERENCES

[7] Y. Nigeria, “Towards Age & Gender Determination, Ridge Thickness to Valley Thickness Ratio (RTVTR) & Ridge Count on Gender Detection,”
Fig. 4. DET curve pertaining to the matching performance before and after performing image de-identification. The de-identifier module incorporated in the fingerprint recognition system does not significantly impact the verification performance.