A further study of low resolution androgenic hair patterns as a soft biometric trait

Chan, Frodo Kin Sun; Kong, Adams Wai Kin

2018

Chan, F. K. S., & Kong, A. W. K. (2018). A further study of low resolution androgenic hair patterns as a soft biometric trait. *Image and Vision Computing*, 69,125-142.

<https://hdl.handle.net/10356/89500>

<https://doi.org/10.1016/j.imavis.2017.08.009>

© 2017 Elsevier B.V. This is the author created version of a work that has been peer reviewed and accepted for publication by *Image and Vision Computing*, Elsevier B.V. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1016/j.imavis.2017.08.009>].

Downloaded on 26 Apr 2023 21:19:49 SGT

A Further Study of Low Resolution Androgenic Hair Patterns as a Soft Biometric Trait

Frodo Kin Sun Chan and Adams Wai Kin Kong

School of Computer Science and Engineering, Nanyang Technological University,
Nanyang Avenue, Singapore 639798
(Emails: csfrodochan@ntu.edu.sg and adamskong@ntu.edu.sg)

Abstract — Soft biometric traits such as skin color, tattoos, shoe size, height, and weight have been regularly used for forensic investigation, especially when hard biometric traits, e.g., faces and fingerprints are not available. Recently, a new soft biometric trait, androgenic hair also called body hair, was evaluated. The previous study showed that low resolution androgenic hair patterns have potential for forensic investigation. However, it was believed that they are not a distinctive biometric trait because of the reported accuracy. To explore discriminative information in androgenic hair patterns, in this paper, a new algorithm, which makes use of leg geometry to align lower leg images, large feature sets (about 60,000 features) extracted through multi-directional grid systems to increase discriminative power and robustness, and class-specific partial least squares (PLS) models to utilize the features effectively, is employed. To further enhance the performance of the class-specific PLS models trained on very limited positive samples, one to three images per model in the experiments, and further enhance robustness against viewpoint and pose variations, a scheme is designed to generate more positive samples from a single image. Experimental results on 1,493 low resolution leg images with large viewpoint and pose variations from 412 legs demonstrate that low resolution androgenic hair patterns contain rich information and the impression of low discriminative power on androgenic hair is due to the method used in the previous study.

Keywords: Soft biometrics, biometrics, emerging biometrics, forensics

1. Introduction

Soft biometric traits, such as race, gender, height, tattoo, and skin mark, are regularly used by law enforcement agencies all around the world for narrowing down suspect lists and combining with hard

biometrics, e.g., faces and fingerprints, to achieve higher matching accuracy and speed [1, 38-39, 40]. Some soft biometric traits can be observed from long distance without user cooperation and can be remembered and described by witnesses easily, e.g., race of criminals. In some cases, even single soft biometric trait can provide accurate matching for particular persons, e.g., the angel tattoo on David Beckham's back and the skin mark on the forehead of Mikhail Gorbachev. Soft biometrics is especially critical when hard biometric traits are neither available nor with low quality. Because soft biometrics is not unique, universal, or permanent, development of new soft biometric traits is always demanded for combining other soft or hard biometric traits to improve matching performance and increase population coverage in different operational environments. Gait and comparative descriptions are two examples [2-3]. Recently, ¹androgenic hair also called body hair was evaluated mainly for identifying criminals and victims in riot, terrorist, and sexual offense images, where their faces and tattoos are not always visible because the criminals hide their faces and tattoos to avoid police identification and only a part of the victim body is visible.

With the advance of the digital technology, forensic investigation often deals with digital images, particularly in cases of masked gunmen, riots, sexual offenses, and child sexual abuse images (also known as child pornographic images). These crimes are widespread. In 2011, the England riot caused approximately 2,500 shops and businesses looted [21]. In 2011, the Stanley Cup riot in Vancouver [22] was about 100,000 people in the streets. On the other hand, the Statistics Canada reported that there were over 5 million child sexual abuse images on the Internet in 2008 and 15,662 websites hosting those images in 2009 [24].

To arrest criminals in these crimes, effective identification methods are essential. Latent prints, blood samples, DNA, dental records, tattoos, face images, and face sketches are used regularly. However, they cannot tackle cases with evidence images showing only non-facial skin of criminals without tattoos such as images of masked rioters and child sexual abuse images. Rioters are usually locals with casual

¹ Androgenic hair, which differs from head hair and the vellus hair, which is much finer and lighter in color, develops on body during and after puberty. The growth of androgenic hair depends on the level of androgens in different persons. [35]

dressing style, e.g., T-shirts, short sleeve shirts, and shorts. Their body parts are frequently exposed in front of cameras. Photos of masked rioters showing non-facial skin can be easily found. Fig. 1 shows riot images collected from the Internet. Paedophiles in child sexual abuse images are usually naked and their body parts are always captured in the images. Child sexual abuse images showing the faces or tattoos of paedophiles are rare because they know clearly that police can use these biometric traits to prosecute them. If androgenic hair is an effective biometric trait, it can be used with other soft biometric traits to narrow down suspect lists. Riot images can be captured by reporters, who always use high quality DSLR cameras. Child sexual abuse images are usually close-up images captured by smartphone cameras or consumer digital cameras. These images are not necessary captured by surveillance cameras, because by 2016 and by 2018, there will be 2.16 and 2.56 billion smartphone users worldwide, respectively [4]. Images captured by surveillance cameras are not within the scope of this paper. It should be highlighted that nowadays, almost everyone has a smartphone and its built-in camera is upgraded every year. For example, Samsung Galaxy S6 has a 16M pixels camera and Samsung Galaxy K Zoom and ASUS ZenFone Zoom have optical zoom functions. Thus, high quality evidence images can be available for criminal and victim identification.

Fig. 1. Riot images collected from the Internet [11, 50]. Lower legs are highlighted.

Though traditional soft biometric traits, including skin color, race, gender, and height, are applicable to these evidence images, their discriminative power is far from enough. Skin marks and blood vessels hidden in color images are recently considered for forensic investigation [31-32, 41]. However, they are not suitable for low resolution images because of their sizes. Even for high resolution images, they are not perfect. Body sites exposed in evidence images may have neither enough skin marks nor hidden blood vessel information. High concentrations of fat and melanin weaken visible light passing through the skin

and make blood vessel visualization difficult. Recently, skin texture [12] is also suggested for criminal and victim investigation on low resolution images. However, skin can be covered by dense androgenic hair.

Medical studies indicate that new follicles do not form naturally after birth in humans [30]. All androgenic hairs manifest a cycle. When one hair falls out, another new hair will grow at the same follicle [30]. Androgenic hairs have a long life cycle [37]. For example, a complete cycle of androgenic hairs on leg can be up to one year [29]. Fig. 2 shows two images from the same leg. Fig. 2(a) was collected in August 2009, while Fig. 2(b) was collected in October 2008. The color circles show partial corresponding androgenic hair follicles. In this small area, more than 40 corresponding androgenic hair follicles are found. More examples can be found in [23]. These properties found by medical researchers imply that androgenic hairs can be a useful biometric trait [23]. In this paper, androgenic hair is considered as a soft biometric trait because it is not universal same as tattoo, which is also considered as a soft biometric trait. According to a study done on 239 adult white males by Garn [36], 3% of them does not have androgenic hair on their lower legs or lower arms.

Fig. 2. Two skin images from the same leg [23]. (a) was collected in August 2009 and (b) was collected in October 2008. The color circles indicate the partial corresponding androgenic hair follicles.

Motivated by the medical studies, a method composed of a dynamic grid system for alignment, Gabor orientation histograms as features, and the Chi-square distance as a dissimilarity function was utilized to examine low resolution androgenic hair patterns in lower leg images [23]. In this paper, this method is called dynamic Gabor orientation histogram (DGOH) method. Their experimental results were

encouraging, but they gave an impression that androgenic hair is not very discriminative. It is suspected that this impression is due to DGOH, which is neither robust to viewpoint variation nor effective to extract discriminative features. In this paper, an algorithm, which uses leg geometry for alignment, local orientation, color, and spatial neighborhood information from multi-directional grid systems to form a large discriminative feature vector, a class-specific model trained for each leg image to effectively exploit the information, partial least squares (PLS) regression to handle unbalanced training data, and a positive sample generation scheme to increase robustness, is employed to examine low resolution androgenic hair patterns on lower legs. Lower legs are selected for this study because they most likely have androgenic hair, comparing with other body sites [36].

The preliminary work [28], which employed leg geometry as an alignment feature and Coherent Point Drift (CPD) [27] as a registration method, focused only on alignment. It still used the DGOH method for matching. The other techniques, including multi-directional grid systems, class-specific models, PLS, and the positive sample generation scheme, were neither utilized by the preliminary work nor the DGOH method. Furthermore, the proposed algorithm is evaluated on 1,493 images, including low quality images, from 412 legs, while the preliminary work was evaluated on front pose images from 283 legs. In addition, 232 images of 67 legs are collected from the Internet for this evaluation. The experimental settings in this paper are closer to real scenarios because low quality images collected from diverse viewpoints and poses are included.

The rest of this paper is organized as follows. Section 2 presents our datasets and a preprocessing algorithm. Section 3 describes the details of the proposed multi-grid feature fusion algorithm. Section 4 reports the experimental results. Section 5 gives discussion and conclusive remarks.

2. Datasets and Preprocessing

Three datasets, two front pose datasets and one low quality dataset, were collected to evaluate the proposed algorithm. They contained lower leg images from different subjects (mainly Chinese, Indian and Vietnamese and some other Asians, Iranians, Europeans, Arabs, and South Americans) and different

genders (male: 71% and female: 29%). These images were taken by a Canon EOS 500D camera (the maximum resolution of 4,752×3,168 pixels) and a Nikon D70s camera (the maximum resolution of 3,008×2,000 pixels) in two occasions with an interval of at least one week. In each occasion, one to three front pose images were taken from each leg. The original images were collected from the left and right legs.²The left leg images were flipped and considered as right leg images from other subjects to enlarge the gallery sets in the experiments.

The front pose dataset collected in the first occasion is called Image Set 1, containing 534 images from 412 legs and the front pose dataset collected in the second occasion is called Image Set 2, containing 536 images from the same 412 legs. The front pose images are to simulate images collected from ex-prisoners and other suspects and good quality images collected in crime scenes. In the image collection, the poses of the subjects and the viewpoints of the cameras were not strictly controlled. Some samples of the front pose datasets are provided in Fig. 3. The relative scales of the images are retained to show the diversity of the images. Table 1 gives the distribution of the number of leg images in each occasion and the distributions of gender and ethnicity.

In addition, 423 low quality images were collected from 281 legs for evaluation. This dataset is called Low Quality Set. In this dataset, the subjects posed freely when the images were taken. Thus, they have problems of blur, out of focus, viewpoints, shadows, scaling, and illumination. These images are to simulate low quality evidence images collected in crime scenes. Also, the left and right legs of the same subject are considered as the legs of two subjects. Some samples in this dataset are shown in Figs. 4(b)-(e). In addition to the Image Set 1, the Image Set 2 and the Low Quality Set, 232 images of 67 legs were collected from the Internet for this study. These datasets will be shared to other researchers within three months after publishing this paper.

² The left and right legs of the same subject are genetically dependent and their androgenic hair patterns have some similarity. This experimental setting makes the testing sets more challenging. More detailed information about their similarity can be found in Section 4.

Four different resolutions (25 dpi, 18.75 dpi, 12.5 dpi, and 6.25 dpi) are studied in this paper. Their image sizes are respectively about 142×298 pixels, 106×233 pixels, 71×149 pixels, and 35×74 pixels. They simulate different distances between a subject (criminal or victim) and a camera. Their corresponding resolutions in terms of pixels per square centimeter are 97, 54, 24, and 6. Utilizing a Canon EOS 500D camera with a sensor pixel size of $2.2107 \times 10^{-5} \text{ mm}^2$ and a focal length of 55mm to capture images with these resolutions, the distances between subjects and the camera are respectively 11.9m, 15.8m, 23.8m, and 47.5m. These distances are calculated by the equation, $d = f \times (25.4 \times h_o) / h_{os}$, where f is the focal length, h_o is the object height in inch, $h_{os} = p_0 \times \sqrt{\rho}$ is the height of an object on sensor in mm, p_0 is the image height of the object in pixels, p_0/h_o is the resolution in dpi (dot per inch) and ρ is the sensor pixel size in mm^2 . Note that Canon EOS 500D is a six-year-old model. To demonstrate that the resolutions of the testing images can be found in real cases, Fig. 4 lists legs collected from rioters and terrorist images in the Internet and our testing images. Their relative sizes are retained for comparison.

Fig. 3. Some samples of left and right legs in the front pose datasets.

Fig. 4. A comparison between the testing images and criminal images in real cases. (a) rioter and terrorist images from the Internet and testing images with sizes of (b) 142×298 pixels, (c) 106×233 pixels, (d) 71×149 pixels, and (e) 35×74 pixels. The relative scales are retained for comparisons. The references of the Internet Images are listed in Appendix A.

Table 1 Details of the datasets (a) the distribution of the number of images per leg in the three datasets (b) the distribution of the gender and ethnicity in terms of leg.

(a)

#Number of images per leg	Number of legs (Image Set 1)	Number of legs (Image Set 2)	Number of legs (Low Quality Set)
1	290	290	139
2	122	120	142
3	0	2	0

*For example, 290 in the table means that 290 legs have one image in the Image Set 1.

(b)

	Number of legs (Image Set 1 and Image Set 2)		Number of legs (Low Quality Set)	
Ethnicity	Male	Female	Male	Female
Chinese	182	82	118	61
*South Asian	80	22	55	16
Caucasian & European	20	0	15	0
#Others	18	6	10	5
Unknown	2	0	1	0

* South Asian including Vietnamese, Bangladeshi, Filipino, Javanese, Indonesian, and Malaysian

Others including Iranian, Persian, Japanese, Korean, Mexican, and Eurasian

Preprocessing is applied to the front pose datasets for segmenting the legs from the background. The images in the Low Quality Set collected from diverse poses and viewpoints are to simulate low quality evidence images collected in crime scenes. They are segmented manually. Manual operations are not uncommon in forensic investigations. For example, before inputting a latent print to an automatic fingerprint identification system, fingerprint experts may manually detect minutiae. The front pose datasets are preprocessed by a scheme which is a modification of the preprocessing scheme used in the DGOH method [23] for improving its robustness to tilted legs [28]. The modified segmentation scheme is summarized as follows:

Step 1: Each channel of an input image (R, G, B) is filtered by a 3×3 median filter. Then, based on

Crowley et al.'s suggestion [18], the ratios $I_R = \frac{R_M}{R_M + G_M + B_M}$ and $I_G = \frac{G_M}{R_M + G_M + B_M}$, where R_M , G_M ,

and B_M are the three filtered channels, are calculated to separate the skin from the background.

Step 2: Pixels are regarded as skin pixels if $I_R > \bar{I}_R$ and $I_G < \bar{I}_G + \text{std}(I_G)$, where $\bar{x}$ and $\text{std}(x)$ are the mean and standard deviation of x , respectively.

Step 3: Afterwards, several morphological operators are applied to the image to remove noise.

Step 4: Then, for each leg, the region of interest which is illustrated in Fig. 5 is extracted. BB' is the shortest distance below MM' and AA' is the shortest distance above MM' . By using A , A' , M , M' , B and B' , the region of interest is defined.

Step 5: In order to increase the number of subjects for evaluation, the segmented left legs are flipped to the right and regarded as the right legs of new subjects. Note that segmented right legs are not flipped.

An illustration of these 5 steps is given in Fig. 6. Note that both left and right legs are segmented from each image. The segmented images are resized to 142×298 pixels, 106×233 pixels, 71×149 pixels, and 35×74 pixels. Figs. 4(b)-(e) show a set of testing images in different resolutions.

Fig. 5. The region of interest is defined by six points A , A' , M , M' , B , and B' (bounded by the red dotted lines).

Fig. 6. The proposed preprocessing scheme. (a) Input images, (b) filtered images with regions of interest, (c) segmented images, and (d) the images of the right legs and the flipped left legs. The images in the first and second rows are from the same person, but different sessions.

3. The Proposed Algorithm

Fig. 7 shows the schematic diagram of the proposed algorithm, which contains training and testing phases. The aim of the training phase is to generate a class-specific PLS model for each image in the training dataset. In the testing phase, input images are evaluated by the trained PLS models.

In the training phase, an image in the training dataset is first chosen to be a model image. This model image and other images from the same leg are considered as positive samples and the rest images in the training dataset are considered as negative samples. To increase robustness against pose variation and enhance the performance of the PLS model, a positive sample generation scheme which produces more positive samples by applying affine transformation to the model image is proposed (Section 3.1). Before extracting features, all other training images, except for the images produced by the positive sample generation scheme, are aligned with the model image. In the testing phase, the input images are also aligned with the model image before matching (Section 3.2). Then, local binary patterns (LBP) and Gabor features are extracted (Section 3.3). To enhance robustness against pose and viewpoint variations, grid systems, including the dynamic grid system reported in [23] and the proposed multi-directional grid systems are used to construct local LBP and Gabor orientation histograms (Section 3.4). These local histograms form large feature sets for extracting more useful information from androgenic hair. The training features are used to train a PLS model. PLS has been successfully applied to other image classification problems [25-26]. The total number of trained PLS models is equal to the number of training images because each of them is selected as a model image. The testing features are multiplied by the PLS regression coefficients β to measure their similarities to the legs of the model images (Section 3.5).

The following subsections explain the details of the proposed algorithm. Section 3.1 mentions the positive sample generation scheme. Section 3.2 offers the alignment method. Section 3.3 describes the feature extraction process. Section 3.4 gives the details of the grid systems and Section 3.5 briefly discusses the PLS regression.

Fig. 7. A schematic diagram of the leg identification algorithm. Blue cross (+) represents the positive training samples and red minus (-) represents the negative training samples.

3.1. A Positive Sample Generation Scheme

The PLS model, like other classification methods, is trained by feature vectors obtained from positive and negative samples. However, the raw training set is imbalanced because the negative samples obviously outnumber the positive samples. Though PLS has been shown to be robust against imbalanced training data [5], to balance them and increase robustness against inaccurate alignment due to pose and viewpoint variations, a scheme which makes use of affine transformation to generate more positive samples is proposed. First, parameters are uniformly sampled from scale $s = \{s_{\min}, \dots, s_{\max}\}$, rotation $\alpha = \{\alpha_{\min}, \dots, \alpha_{\max}\}$, and translation $(x, y) = \{(x, y)_{\min}, \dots, (x, y)_{\max}\}$ to construct a set of affine transformation matrices, i.e.

$$A_i = \begin{bmatrix} s_i \cos \alpha_i & s_i \sin \alpha_i & x_i \\ -s_i \sin \alpha_i & s_i \cos \alpha_i & y_i \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Then, the affine transformation matrixes and the bicubic interpolation are applied on the model image to create new positive samples. Let the number of negative samples be N_n and the number of positive samples be N_p . The scheme constructs $N_n - N_p$ affine transformation matrixes and generates $N_n - N_p$ new positive samples for each model image. It should be emphasized that the proposed algorithm aligns the negative

samples with the model image, but no alignment is performed on the new positive samples. If the new positive samples are aligned with the model image, the effect of the affine transformation will be eliminated, i.e. the new positive samples and the model image will become the same. By using this scheme, more positive samples are generated and PLS training becomes balanced and more robust against inaccurate alignment. In addition, the positive sample generation scheme is only used in training and therefore it does not increase computational cost for matching.

3.2 A Low Resolution Alignment Method

To align two lower leg images, e.g., a model image with a negative sample and a model image with a testing image, a low resolution alignment method consisting of two parts, edge sampling and image registration, is proposed. Edge samples of two leg boundaries are first extracted. Then, angular samples of the leg boundaries defined through the centerline and the center point illustrated in Fig. 8 are used as registration points. In order to obtain the centerline and the center point, the distance transform [6] is defined as

$$D_i(p) = \min_{b \in B_i} \|p - b\|_2, \quad (2)$$

where $i \in \{L, R\}$, B_L and B_R are the left and right boundaries of a leg, D_L and D_R are the results from the left and right boundaries, and $p = (x, y)$ is a pixel location of the leg image. To combine the information from both boundaries,

$$D_o(p) = \max((D_L(p), D_R(p))) \cdot e^{-m}, \quad (3)$$

where $m = \lfloor |D_L(p), D_R(p)| \rfloor$ and $\lfloor \cdot \rfloor$ denotes a floor operator, is defined. The function e^{-m} is used to ensure that the points close to the midpoints of the left and right boundaries have high responses. The center point can be found as $p_c = \operatorname{argmax}_p D_o(p)$. If there is more than one point p_k such that $D_o(p_k) = D_o(p_c)$ and the number of p_k (including p_c) is odd, a point whose y-coordinate is the median is chosen as the center point. If the number of p_k is even, a point whose y-coordinate is the $(g/2)^{\text{th}}$ point is chosen as the center point, where g is the total number of p_k . The centerline, which passes through the center point, is obtained by using linear regression on the points $\{p_1, \dots, p_i, \dots, p_H\}$, where $p_i = \operatorname{argmax}_p D_o(p)$ and p is a pixel location in the i^{th} row of D_o .

Fig. 8. Illustration of the centerline and the center point defined by D_O . (a) The magnitude map of D_O , where brighter color represents a higher value of D_O . (b) The centerline (the red line), the center point (the blue cross) estimated using D_O and the angular samples (the green dots).

After obtaining the center point and the centerline, the points on the boundaries are uniformly sampled in terms of angular distance. Afterwards, pruning and normalization processes are applied to the two point sets obtained from an image pair (e.g., a model image and a negative training sample) to ensure rotational invariant and strengthen robustness of translation and scale variations. The affine CPD method [27] is used to align the two point sets and generates a transformation matrix. By using the transformation matrix and bicubic interpolation, one of the images is projected to the other.

For each class-specific PLS model, the positive and negative training images and the testing images are projected to the model image by using this alignment scheme, except for new images created by the positive sample generation scheme. The preliminary work [28] only projects one of the color channels for feature extraction. In this paper, all three color channels are projected and therefore, the resultant image is still a color image.

3.3 Feature Extraction

The DGOH method only employs grayscale intensity for identification, but neglects information in different color channels, which is shown to be useful for other recognition problems, e.g., image segmentation and face recognition [16-17, 33]. Thus, the Gabor and LBP features utilized in this paper are extracted from different channels of a color space, e.g., the RGB space. First, Gabor filters are convolved with a color channel I of an image and its Gabor orientation field is calculated by

$$O(x, y) = \arg_{\theta_k} \max_{m,k} |G_{\lambda_{mk}, \theta_k, \sigma_m, \gamma} * I|, \quad (4)$$

where $G_{\lambda_{mk}, \theta_k, \sigma_m, \gamma}$ is the real part of a zero DC Gabor filter with the orientation $\theta_k = k\pi/8$, λ_{mk} is the wavelength of the sinusoidal component, σ_m is the standard deviation of the elliptical Gaussian window, γ is the spatial aspect ratio, $m = \{1, \dots, M\}$ and $k = \{1, \dots, K\}$ are the scale and orientation indexes. In the implementation, 32 orientations and 4 scales are used [23].

LBP features are useful to represent image texture and invariant to monotonic gray-level changes [15]. The local binary patterns $LBP_{P,R}$ [14] are defined as

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad (5)$$

where P is the number of sampling points, R is the radius of the sampling circle, g_c is the center pixel value, g_p is the neighborhood pixel value and $s(x) = 1$ if $x \geq 0$, otherwise $s(x) = 0$. In the proposed algorithm, $LBP_{8,2}^{u2}$, which are the uniform LBP with at most two bitwise transitions from 0 to 1 or vice versa when they are considered as circular patterns, are extracted from the directional grid systems and $LBP_{8,1}^{riu2}$, which are the nine rotational invariant uniform binary patterns [14], are extracted from the dynamic grid systems. Two different forms of LBP are applied because the grid size of the dynamic grid system is relatively small and the histogram of $LBP_{8,2}^{u2}$ has more bins than that of $LBP_{8,1}^{riu2}$. If $LBP_{8,2}^{u2}$ are extracted from the dynamic grid system, its grid histogram may not be stable because of the small grid size and relatively large number of bins. Thus, $LBP_{8,1}^{riu2}$ are extracted from the dynamic grid system.

3.4 Two Grid Systems

For effectively extracting local information, gridding the Gabor and LBP features in images is essential. The dynamic grid system and the directional grid systems are used. The dynamic grid system, which is a part of the DGOH method [23], determines the block sizes based on the row width of the leg image (see Fig. 9(m)). The number of blocks in the i^{th} row is calculated by

$$N_i = \left\lfloor \frac{E}{2} \times \frac{\mu_{is}}{\mu_{ia}} + 0.5 \right\rfloor, \quad (6)$$

where $i = 1, \dots, E$, μ_{is} is the average number of skin pixels in the i^{th} row, μ_{ia} is the total number of pixels in the i^{th} row, including the background pixels, E is the number of rows to be divided and $\lfloor \cdot \rfloor$ denotes a floor operator. When the number of blocks is found, the block size can be calculated for each row easily. The corresponding histograms of Gabor feature $h_{OB_j}(k)$ and LBP feature $h_{LB_j}(k)$ respectively defined as

$$h_{OB_j}(k) = \sum_{(x,y) \in B_j} \delta(O(x,y) = k), \quad (7)$$

$$h_{LB_j}(l) = \sum_{(x,y) \in B_j} \delta(LBP_{P,R}^{riu2}(x,y) = l), \quad (8)$$

where B_j is a set containing coordinates of the pixels in the j^{th} block, δ is the Kronecker delta function, k is an orientation index and l is a binary pattern index, are extracted.

The dynamic grid system employed by DGOH is sensitive to viewpoint variation. When the viewpoint changes, the grids do not correspond correctly illustrated in Figs. 9(a) and (b). To overcome this problem, the directional grid systems are designed to increase robustness against this variation. In the directional grid systems, the blocks are rotated in different direction θ . The block height $H_{j\theta}$ and the block width $W_{j\theta}$ are respectively defined as $H_{j\theta} = \lfloor (H_\theta - 1)/N_H \rfloor$ and $W_{j\theta} = \lfloor (W_\theta - 1)/N_W \rfloor$, where H_θ is the distance between the top leg pixel and the bottom leg pixel along direction θ , W_θ is the distance between the leftmost and the rightmost leg pixels along the direction perpendicular to θ , N_W is the number of columns and N_H is the number of rows along direction θ . The corresponding histograms of Gabor feature $h_{OB_{j\theta}}(k)$ and LBP feature $h_{LB_{j\theta}}(l')$ respectively defined as

$$h_{OB_{j\theta}}(k) = \sum_{(x,y) \in B_{j\theta}} \delta(O(x,y) = k), \quad (9)$$

$$h_{LB_{j\theta}}(l') = \sum_{(x,y) \in B_{j\theta}} \delta(LBP_{P,R}^{u2}(x,y) = l'), \quad (10)$$

where $B_{j\theta}$ is a set containing coordinates of the pixels in the j^{th} block in the direction θ , δ is the Kronecker delta function, k is an orientation index and l' is a binary pattern index, are extracted. Figs. 10(a)-(l) illustrate the directional grid systems. In the implementation, $\theta = \{0^\circ, 30^\circ, \dots, 330^\circ\}$, $N_W = 8$, $N_H = 5$,

and the first and last rows are not used. The pairs of opposite directions (e.g., 90 and 270 degrees) are included because the blocks are placed from one side to the other side along each direction. For example, the blocks are placed from left to right along 90 degrees and from right to left along 270 degrees. Due to the floor operators used in the block height and width computation, blocks placed from opposite directions are not identical. Thus, blocks in opposite directions are included to generate more features for increasing distinctiveness of the feature set.

The blocks in the directional grid systems are larger comparing with those in the dynamic grid system and therefore, they are more robust against inaccurate alignment (Fig. 9). However, if only one grid, e.g., the one in Fig. 10(a), is used, the feature dimension will be very low and it may not contain enough discriminative information for achieving high accuracy. With this consideration, the directional grid systems are composed of many grids in different directions. Note that the leg boundaries are also covered by the blocks, which is different from the previous work [23]. The LBP histograms and the Gabor orientation histograms in each block of the dynamic grid system and the directional grid systems are extracted and normalized. The grid systems and the feature descriptors employed in this paper are not specific for androgenic hair and can also be applied to other texture [12].

For each image m , all the corresponding histograms $h_{O_mB_j}$, $h_{L_mB_j}$, $h_{O_mB_{j\theta}}$, and $h_{L_mB_{j\theta}}$ are concatenated together to form a large feature vector with a length of 59,766, i.e.

$$\mathbf{X}_m = \{h_{O_mB_1}, \dots, h_{O_mB_N}, h_{L_mB_1}, \dots, h_{L_mB_N}, h_{O_mB_{1,0}}, \dots, h_{O_mB_{N',330}}, h_{L_mB_{1,0}}, \dots, h_{L_mB_{N',330}}\}, \quad (11)$$

where N is the number of blocks from the dynamic grid system and N' is the number of blocks from the directional grid systems.

Fig. 9. A leg with four orange dots is captured at different viewpoints: (a) is the leg images, (b) and (c) are the results given by the dynamic grid system from the DGOH method [23] and one of the directional grid systems proposed in this paper. The first images in (a), (b) and (c) are the leg image captured from 0 degrees with respect to the leg and the second images are the leg image captured from 15 degrees with respect to the leg.

Fig. 10. (a)-(l) are the directional grid systems and (m) is the dynamic grid system.

3.5. PLS Regression

Partial Least Squares (PLS) is a widely used tool for processing chemical data for chemometric analyses [13]. Its underlying assumption is that there is a small number of latent variables to drive a system or process for generating the observed data. PLS regression is a model searching for the internal relationship between a feature matrix $\mathbf{X}_{(n \times p)}$ and a response variable $\mathbf{y}_{(n \times 1)}$, where $\mathbf{X}_{(n \times p)}$ and $\mathbf{y}_{(n \times 1)}$ are mean-centered and variance-scaled, n is the number of training data and p is the dimension of the feature vector. Latent variables are used to represent their internal relationship, which is defined as

$$\mathbf{X} = \mathbf{V}\mathbf{P}^T + \mathbf{E}, \quad (12)$$

$$\mathbf{y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \quad (13)$$

where $\mathbf{V}$ and $\mathbf{U}$ are the latent components, $\mathbf{P}$ and $\mathbf{Q}$ are the loading matrices, the superscript T represents a matrix transpose operator and $\mathbf{E}$ and $\mathbf{F}$ are the residue matrices. In this work, the NIPALS algorithm [13] is used and its computation details are given in Algorithm 1.

Algorithm 1. NIPALS algorithm

1. for $i = 1$ to d do
2. $\mathbf{W}_i = \mathbf{X}^T \mathbf{y} / \|\mathbf{X}^T \mathbf{y}\|$
3. $\mathbf{V}_i = \mathbf{X} \mathbf{W}_i / \|\mathbf{X} \mathbf{W}_i\|$
4. $\mathbf{X} = \mathbf{X} - \mathbf{V}_i \mathbf{V}_i^T \mathbf{X}$
5. $\mathbf{y} = \mathbf{y} - \mathbf{V}_i \mathbf{V}_i^T \mathbf{y}$
6. end
7. $\boldsymbol{\beta} = \mathbf{W} \mathbf{Q}^T = \mathbf{W} (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T \mathbf{y}$

where $\boldsymbol{\beta}$ is a $p \times 1$ regression coefficient vector of $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{f}'$, $\mathbf{f}'$ is a residual vector, d is the number of latent variables, $\mathbf{W}_i$ and $\mathbf{V}_i$ are the i^{th} columns of the matrices $\mathbf{W}$ and $\mathbf{V}$ in $\mathbf{V} = \mathbf{X} \mathbf{W}$. The second and third steps are matrix projection and the fourth and fifth steps are matrix deflation. In the training phase, the target responses of all the positive samples are set to one and the target responses of all the negative samples are set to negative one. In the testing phase, an estimated regression response $\hat{y}_N$ for an input feature vector $\mathbf{x}_N$ can be calculated by $\hat{y}_N = \bar{y} + \boldsymbol{\beta}^T \mathbf{x}_N$, where $\bar{y}$ is the sample mean of the training y . The model image corresponding to the highest PLS regression response is regarded as the best match. Note that $\bar{y}$ is ignored in the response calculation because it is a constant. PLS is robust against imbalanced training data [5]. The proposed algorithm trains one class-specific PLS model for each image in gallery sets. Since incremental PLS algorithm has been developed, the scalability of the proposed algorithm is not a problem [9].

4. Experimental Results

To evaluate the performance of the proposed algorithm, experiments were conducted on the datasets, including the Image Set 1, the Image Set 2, the Low Quality Set and the Internet images. Images with about 25 dpi were used as testing images for all the experiments except for the very low resolution image experiment. In that experiment, images with about 25 dpi, 18.75 dpi, 12.5 dpi, and 6.25 dpi were used and

their corresponding sizes were respectively 142×298 pixels, 106×233 pixels, 71×149 pixels, and 35×74 pixels. The Image Set 2 containing 536 images from 412 legs was regarded as a gallery set in all the experiments except the first experiment. The first experiment aimed to evaluate the performance of the proposed algorithm for the front pose leg images. The Image Set 1 and the Image Set 2 were used as a probe set and a gallery set, and vice versa. The second experiment aimed to evaluate the performance of the proposed algorithm for the low quality images. The Low Quality Set was used as a probe set and the Image Set 2 was used as a gallery set. In both of the experiments, the DGOH method [23] and the preliminary method [28] were implemented for performance comparison. The gallery sets were used for training and the probe sets were used for testing only. The Gabor and LBP features were extracted from the grayscale intensity I , the color channel C_r in the YC_bC_r color space and the color channel Q in the YIQ color space. C_r and Q were chosen based on Liu et al.'s and Choi et al.'s suggestions [16-17]. The Gabor features were obtained from the grayscale intensity I only because they are useful for locating strong edges, which are not always apparent in the C_r and Q components. In the positive sample generation scheme, the scale, rotation, and translation parameter sets were set to be $s = \{1.2, 1.1, 1.0, 0.9, 0.8\}$, $\alpha = \{-10, -6, -2, 2, 6, 10\}$, $x = \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$, and $y = \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$. Cumulative Match Characteristic (CMC) curves are used as a performance index. CMC curves are a standard approach to evaluate biometric algorithms developed for criminal and victim identification. They have been employed by numerous researchers.

Section 4.1 compares the proposed algorithm with the DGOH method, the preliminary work [28], and two deep learning methods. Section 4.2 evaluates it under different settings. Section 4.3 studies the performance influenced by gender, race, hair density, and genetic dependence. Section 4.4 reports the results obtained from the Internet images.

4.1 Comparisons with the Existing Methods

In general, probe sets and gallery sets for forensic evaluation should not be swapped, because they have different characteristics. 2-fold cross validation was adopted in the first experiment, because the Image Set

1 and the Image Set 2 have similar quality. In the first fold, the Image Set 1 was used as a probe set and the Image Set 2 was used as a gallery set. In the second fold, their roles were exchanged. Table 2(a) summarizes the average accuracy of these two folds. The results show clearly that the proposed algorithm performs the best and there is no significant difference between Figs. 11(a) and (b) obtained from the two folds. Fig. 11(c) shows the CMC curves of the three methods obtained from the Low Quality Set. In this experiment, the Image Set 2 was used as a gallery set. The performance of the DGOH method [23] is the worst among the previous methods designed for androgenic hair identification. The preliminary work [28], which aligns the legs by using leg geometry and employs the DGOH method for feature matching, performs better than the DGOH method, but its performance is still far from the proposed algorithm. The proposed algorithm outperforms these two methods because it makes use of different color channels and grid systems to extract more discriminative information. Furthermore, the large blocks relax the requirement of accurate alignment and indirectly extract leg geometry information. It also learns a class-specific PLS model for each image in the gallery sets so that importance of different features for different legs can be taken into account. In addition, extra positive samples are generated to increase robustness against inaccurate alignment due to pose and viewpoint variations. To show clearly the difference among these three methods, Table 2 lists their accuracy at different ranks. The accuracy in Table 2(a) is the average accuracy of the two folds. Comparing with the DGOH method, the proposed algorithm offers improvement of over 15% in terms of rank-1 accuracy in all the cases. At the 20th rank, the proposed algorithm achieves accuracy of 100% for the Image Set 1 and the Image Set 2 and offers improvement of 24.35% for the Low Quality Set. Because the preliminary method and the proposed algorithm use the same alignment scheme, by comparing their accuracy with the accuracy of the DGOH method, improvement offered by the alignment scheme and the other computational components can be observed and they are listed in Table 2. Though the alignment scheme improves the performance, the other computational components play a more important role, especially for the Low Quality Set. For the rank-1 accuracy of the Low Quality Set, they offer over 14% of improvement, while the alignment scheme offers only 1.90% of improvement.

In addition to the previous androgenic hair matching methods, the proposed algorithm is also compared with two deep learning methods, NN2 and NN4 reported in [46]. Their architectures are restrained but the loss function proposed by Wen et al. [47] is employed. According to Wen et al., their loss function performs better. The comparison is given in Fig. 11. The two deep learning methods do not perform better than other methods because deep learning requires a very huge database to train the parameters in the deep nets. The current databases cannot fulfill their data requirement.

Fig. 11. CMC curves from (a) the first fold, (b) the second fold and (c) the Low Quality Set as the probe set and the Image Set 2 as the gallery set.

Table 2 A summary of the experimental results. (a) The Image Set 1 with the Image Set 2 and (b) the Low Quality Set with the Image Set 2

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	<u>94.39%</u>	<u>99.91%</u>	<u>100%</u>
The preliminary method [28]	84.49%	93.65%	96.08%
The DGOH method [23]	79.25%	91.50%	94.49%
Improvement offered by the alignment scheme	5.24%	2.15%	1.59%
Improvement offered by the other computational components	9.90%	6.26%	3.92%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.96%</u>	<u>59.10%</u>
The preliminary method [28]	10.17%	36.88%	43.97%
The DGOH method [23]	8.27%	30.26%	34.75%
Improvement offered by the alignment scheme	1.90%	6.62%	9.22%
Improvement offered by the other computational components	14.18%	16.08%	15.13%

4.2. Performance under Different Settings

Some may suspect that the proposed algorithm performs better than the DGOH method and the preliminary work solely because of the large feature sets, storing more discriminative information. To study it, the proposed algorithm and the preliminary method using the large feature sets were compared on the probe sets, the Image Set 1 and the Low Quality Set. In other words, both algorithms used the same feature set and the same alignment scheme. The preliminary method still employed the chi-square distance for matching. The corresponding CMC curves are given in Fig. 12. Fig. 12(a) shows that the preliminary method using the large feature sets performs even worse than the DGOH method for the Image Set 1 and Fig. 12(b) shows that the preliminary method using the large feature sets performs slightly worse than the preliminary method for the Low Quality Set. Note that the preliminary method and the DGOH method utilize short feature vectors. These results demonstrate clearly that without a proper classification scheme, only large feature sets cannot increase the identification rates and even degrade the performance of the original preliminary method. The proposed algorithm can take advantage from the large feature sets because the importance of different features for different legs can be taken into account by the class-specific PLS models. The experimental results point out that without a careful design, the large feature sets, which store more information, are useless or even harmful for the performance.

Fig. 12. Comparison between the proposed algorithm and the preliminary method using the large feature sets for the probe sets of (a) the Image Set 1 and (b) the Low Quality Set.

Someone may suspect that PLS performs better than chi-square distance regardless of the feature size, but it is not true for both probe sets. In Fig. 12(a) (12(b)), their rank-1, rank-10, and rank-20 accuracy is 85.21%, 94.19%, and 97.19% (10.17%, 29.31%, and 36.88%) from the CMC curves of the preliminary work using small feature sets and chi-square distance. In Fig. 13(a) (13(b)), the corresponding accuracy is 75.09%, 90.82%, and 94.01% (9.46%, 25.06%, and 33.81%) from the CMC curves of the Gabor features on the dynamic grid system using small feature sets and PLS. By comparing the accuracy of these CMC curves, it is shown that PLS performs worse than chi-square distance in small feature sets. This comparison pinpoints that the large feature sets are vital for the proposed algorithm.

To evaluate the effectiveness of the color information, the proposed algorithms with and without color information were compared on the probe sets of the Image Set 1 and the Low Quality Set. The Image Set 2 was used as a gallery set. When color information was excluded, features were extracted from the grayscale intensity I only. When color information was included, features were extracted from the grayscale intensity I , the color channel C_r in the YC_bC_r color space and the color channel Q in the YIQ color space. Their results are summarized in Table 3. The proposed algorithm offers higher rank-10 and rank-20 accuracy for the Image Set 1. It also outperforms the one without color information for the Low Quality Set at the 1st, 20th and 30th ranks. At the 30th rank, the color information provides improvement of 2.60%. These results show that color provides additional information to the proposed algorithm to achieve better performance.

Table 3 A summary of the experimental results of the proposed algorithm and the one without color information for the probe sets of (a) the Image Set 1 and (b) the Low Quality Set.

(a)			
	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	93.82%	<u>99.81%</u>	<u>100%</u>
The proposed algorithm without color information	<u>94.57%</u>	98.88%	99.25%

(b)			
	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.25%</u>	<u>59.10%</u>
The proposed algorithm without color information	24.11%	<u>52.25%</u>	56.50%

To further understand the performance of the proposed algorithm, experiments using different algorithmic settings were conducted. Figs. 13(a) and (b) show the CMC curves obtained from different features and different grid systems. The Image Set 1 and the Low Quality Set were respectively used as the probe sets of Figs. 13(a) and (b). There are four different settings: the Gabor features on the dynamic grid system, the LBP features on the dynamic grid system, the Gabor features on the directional grid systems and the LBP features on the directional grid systems. Figs. 13(a) and (b) pinpoint clearly that the LBP features on the directional grid systems achieve the best result among the four settings and the LBP features on the dynamic grid system give the worst result. The dimension of the LBP features on the directional grid systems is 45,312. The dimension of the Gabor features on the directional grid systems is 8,192. The feature dimensions of the other two settings are only 3,232 and 3,030. The CMC curves in Figs. 13(a) and (b) point out the importance of using large feature sets for androgenic hair pattern identification. The results in Figs. 13(a) and (b) also agree with Su et al.'s study that the Gabor features outperform the LBP features on the dynamic grid system [23]. When they are combined together, i.e., the proposed algorithm, it outperforms any individual setting. This result further confirms the necessity of using the features and the grid systems together in the proposed algorithm.

In order to compare the two grid systems, two experiments were performed on the Image Set 1 and the Low Quality Set. Figs. 13(c) and (d) show their CMC curves. In these experiments, the proposed algorithm only used one type of the grid systems in each of the test, but the other computational components remained the same. Figs. 13(c) and (d) demonstrate that the directional grid systems obviously outperform the dynamic grid system. These experimental results are expectable because the blocks in the directional grid systems are larger and more robust to inaccurate alignment and they are composed of a lot of grids to generate more informative feature vectors.

In this paper, a positive sample generation scheme is also used. To evaluate its effectiveness, two experiments were performed on the Image Set 1 and the Low Quality Set. Their CMC curves are given in Figs. 13(e) and (f). Two algorithmic settings, the proposed algorithm with and without the positive sample

generation scheme, were studied in these experiments. Fig. 13(e) shows that the performance of these two settings is close, but Fig. 13(f) shows that the proposed algorithm with the positive sample generation scheme performs much better. The results in Figs. 13(e) and (f) are different mainly because of alignment. Pose and viewpoint variations in the Image Set 1 (Fig. 13(e)) are not too large and the proposed alignment scheme based on leg geometry and the directional grid systems can handle these variations well. Therefore, the positive sample generation scheme cannot offer a significant improvement. When the positive sample generation scheme was turned off, the PLS models were trained on very imbalanced datasets, which contained only 1-3 positive samples. Fig. 13(e) shows that the performance without the positive sample generation scheme is close to that with the positive sample generation scheme. It means that the PLS models are robust to imbalanced datasets. However, Fig. 13(f) indicates that PLS alone is not sufficient for the Low Quality Set because pose and viewpoint variations in this dataset are very large. The PLS models trained with the samples produced by the positive sample generation scheme are more robust to inaccurate alignment because these samples are in fact distorted versions of the model images. Fig. 13(f) demonstrates that this scheme is essential for low quality images, e.g., images collected from crime scenes.

The proposed algorithm uses a large feature set to achieve high recognition rate. To demonstrate its necessity, Peng et al.'s feature selection method [49] is applied to the original feature set and 6000 features are selected. Figs. 13(g) and (h) show that the reduction of the features causes the accuracy drop. Extracting the large feature set requires more computational time and storage space comparing with smaller feature set. Matching accuracy is more important than speed in forensic applications, because they generally do not have real-time requirement and the development of parallel processing and hardware can speed up the method efficiently. With the recent development of storage devices, storage for the large feature set is not a problem.

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 13. CMC curves were obtained from different algorithmic settings to analyze different computational components in the proposed algorithm. (a) and (b) compare the Gabor and LBP features with different grid systems. (c) and (d) compare the directional grid systems with the dynamic grid system. (e) and (f) evaluate the effectiveness of the positive sample generation scheme. (g) and (h) evaluate the effectiveness of the large feature sets. The CMCs in the first column were obtained from the Image Set 1 and the CMCs in the second column were obtained from the Low Quality Set.

Table 4 summarizes the contribution of the alignment scheme, the directional grids, and the positive sample generation scheme. The directional grids make the greatest contribution for both datasets. It indicates that good feature sets are very important. The positive sample generation scheme and alignment scheme do not offer significant improvement for the Image Set 1 because the viewpoint variation of the Image Set 1 is not very large. However, they are vital for the Low Quality Set.

Table 4 A summary of the experimental results with different settings of the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	93.82%	<u>99.81%</u>	<u>100%</u>
The proposed algorithm without alignment scheme	92.51%	98.31%	99.06%
The proposed algorithm without directional grids	86.70%	97.94%	98.50%
The proposed algorithm without positive samples generation	<u>94.76%</u>	99.81%	100%
Improvement offered by the alignment scheme	1.31%	1.50%	0.94%
Improvement offered by the directional grids	7.13%	1.87%	0%
Improvement offered by the positive samples generation	-0.94%	0%	0%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.25%</u>	<u>59.10%</u>
The proposed algorithm without alignment scheme	13.24%	52.01%	58.39%
The proposed algorithm without directional grids	12.29%	33.33%	43.26%
The proposed algorithm without positive samples generation	17.97%	38.06%	45.15%
Improvement offered by the alignment scheme	11.11%	0.24%	0.71%
Improvement offered by the directional grids	12.06%	18.92%	15.84%
Improvement offered by the positive samples generation	6.38%	14.19%	13.95%

The proposed algorithm makes use of PLS instead of linear SVR, which is commonly employed by other researchers, for training regression models. An experiment was conducted to compare PLS and linear SVR on the proposed algorithm. In this experiment, all features and schemes in the proposed algorithm were the same, except that one algorithm used PLS and another applied linear SVR. Note that the positive sample generation scheme was also used with PLS and linear SVR. The linear SVR is implemented by using L2-regularized L2-loss support vector regression (primal) with the default parameters from LIBLINEAR [34, 48]. The results are given in Table 5. In the first ten ranks, the proposed algorithm using PLS performs better than that using linear SVR for both probe sets. The greatest difference is about 1.31% for the Image Set 1 and 3.55% for the Low Quality Set. It shows that PLS is a reasonable choice for the proposed algorithm.

Table 5 A comparison between PLS and linear SVR on the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.

(a)

	Rank-1 accuracy	Rank-5 accuracy	Rank-10 accuracy
The proposed algorithm with PLS	<u>93.82%</u>	<u>99.81%</u>	<u>99.81%</u>
The proposed algorithm with linear SVR	92.51%	98.88%	<u>99.81%</u>

(b)

	Rank-1 accuracy	Rank-5 accuracy	Rank-10 accuracy
The proposed algorithm with PLS	<u>24.35%</u>	<u>38.77%</u>	<u>43.74%</u>
The proposed algorithm with linear SVR	22.22%	35.22%	43.03%

To test the performance of the proposed algorithm on very low resolution images, another experiment was performed. To keep the parameters in the proposed algorithm unchanged, images in the probe sets and images in the gallery set were first resized to 142×298 pixels (25 dpi), 106×233 pixels (18.75 dpi), 71×149 pixels (12.5 dpi), and 35×74 pixels (6.25 dpi), and were then scaled back to 142×298 pixels. The Image Set 2 was used to generate the new gallery sets and the other two datasets were used to generate the new probe sets. Figs. 4(b)-(e) show images in these four resolutions. The proposed algorithm was applied to these new probe and gallery sets. The experimental results are given in Fig. 14. Fig. 14(a) shows that in most of the ranks, when the resolution decreases, the performance drops. However, the results in Fig. 14(b) are different. The CMC curves obtained from the images with 18.75 dpi and 12.5 dpi are better than those obtained from the images with 25 dpi and 6.25 dpi. The algorithm is more sensitive to inaccurate alignment for higher resolution images. Due to the low image quality of this dataset, accurate alignment is more difficult to be achieved. Note that the alignment is based on leg geometry, but not hair texture. High resolution images and low resolution images have roughly same alignment accuracy because leg geometry is low frequency information. The low resolution images can be considered as resultant images obtained by applying a low-pass filter to the high resolution images and a block in the grid systems, which determines a region to compute a particular feature, can be considered as a 2D box function under affine transform. The low resolution images are more robust to inaccurate alignment due to the low pass effect and the 2D box function, whose boundaries are discontinuous. Thus, the CMC curve obtained from the images with 25 dpi is not better than those from the images with 18.75 dpi and 12.5 dpi for the Low Quality Set. For the lowest resolution images (6.25 dpi), even though influence from inaccurate alignment diminishes, the CMC curve still drops because of the reduction of discriminative information in the images. For these very low resolution images, which can be found in real cases (Fig. 4), the proposed algorithm gives the rank-15 accuracy of 49.65% and the rank-30 accuracy of 57.45%. It demonstrates the effectiveness of the proposed algorithm for low resolution images. Nevertheless, the proposed algorithm performs similarly on images with different resolutions.

Fig. 14. CMC curves for different image resolutions. (a) The Image Set 1 and (b) the Low Quality Set.

4.3. Performance Influenced by Gender, Ethnicity, Hair Density and Genetic Dependence

The previous experiments evaluate the performance of the algorithms in different settings. The following experiments are to study the matching performance influenced by gender, ethnicity, hair density, and genetic dependence. The corresponding CMC curves are given in Figs. 15-18, respectively. Fig. 15(a) shows that male legs, which normally have higher hair density, perform for better than female legs for the Image Set 1 but not for the Low Quality Set. It was expected that male legs with denser hair should provide strong signals such that the extracted features are more stable and give better performance. Fig. 15(a) fits this expectation. However, Fig. 15(b) computed from the Low Quality Set is not. The images in the Low Quality Set are suffered from various distortions, including viewpoint and pose variations. If these distortions on the male legs and the female legs are different, it is hard to conclude that female hair patterns perform better than male hair patterns for low quality images. Fig. 16(a) shows the CMC curves of different ethnicities from the Image Set 1, where South Asians perform better than Chinese and Chinese perform better than Caucasians and Europeans. From our observation, South Asians have denser hair comparing with Chinese. Though Caucasians and Europeans also have dense hair too, their hair color and skin color are similar. In other words, their signals may not be as strong as those from South Asians. However, it should note that the number of Caucasians and Europeans in sets are not a lot (Table 1). Thus, the remark is not conclusive. Fig. 16(b) shows that the Caucasians, Europeans, Chinese and South Asians have different

performance for the Low Quality Set because of different levels of distortion discussed before. Furthermore, the hair patterns are classified into three groups, high, medium, and low hair density. Table 6 shows the distribution of hair density on Image Set 1, Image Set 2, and Low Quality Set. Fig. 17 shows the corresponding CMC curves. Fig. 17(a) shows that the hair patterns with medium density, many of which are from South Asians, perform the best. It was expected that hair patterns with higher density should perform better. The figure shows that hair density is not the only factor to determine the performance. The difference between skin color and hair color, which directly affects the stability of the extracted features, may also be a factor. The algorithm does not explicitly separate hair with different densities. Because of the density difference, the responses to the feature descriptors are also different. PLS trained on the databases automatically assigns weight to each feature and to each hair pattern. These weights indicate the discriminative power of the features. In other words, the algorithm makes use of their density difference implicitly. Figs. 15(a)-17(a) indicate that gender, ethnicity, and hair density have minor impacts on performance and Figs. 15(b)-17(b) indicate that these impacts are far lower than the impacts from other distortions, e.g., viewpoint and pose variations.

In addition to gender, ethnicity, and hair density, an experiment is performed to study the genetic dependence of androgenic hair patterns. Genetically identical biometric samples, including those from the same subjects (e.g., the left iris and the right iris of the same subject) and those from identical twins, are considered as the most similar biometric samples. Genetically identical biometric traits, e.g., fingerprint[43], palmprint[32], face[44] and vein[45], have been studied. The hair patterns of the left leg and the right leg of the same subject are also genetically identical. In this experiment, the hair patterns of the right leg images (flipped left leg images) are used as probe images and their targets in the gallery sets are the corresponding hair patterns of the flipped left leg images (right leg images). Fig. 18 shows the original CMC curves, which are labelled as genuine matching, and the CMC curves from genetically identical hair patterns, which are labelled as virtual twin matching. It is clear that the performance of the genetically identical hair patterns are much lower than the original one. However, it pinpoints that androgenic hair patterns are neither completely genetically independent nor completely genetically

dependent. It is similar to other biometric traits, such as fingerprint, palmprint, and vein, which partially genetically dependent. In the previous experiments, the hair patterns of the left legs are flipped and considered as hair patterns from other subjects. It in fact makes the experiment settings hard for the algorithm.

Fig. 15. CMC curves of male and female, (a) theImage Set 1 and (b) the Low Quality Set.

Fig. 16. CMC curves of different ethnicities, (a) the Image Set 1 and (b) the Low Quality Set.

Fig. 17. CMC curves of hair patterns with different densities, (a) the Image Set 1 and (b) the Low Quality Set1.

Fig. 18. Comparison of genetically identical hair patterns, (a) the Image Set 1 and (b) the Low Quality Set.

Table 6 The distribution of hair density on the Image Set 1, the Image Set 2, and the Low Quality Set

	Number of legs (Image Set 1 and Image Set 2)	Number of legs (Low Quality Set)
Low hair density	122	93
Middle hair density	201	132
High hair density	89	56

4.4. Matching Results from the Internet Images

Even though the images in the Low Quality Set have very large viewpoint and pose variations, to demonstrate the applicability of androgenic hair in real cases, 232 images of 67 lower legs were collected from the Internet. Fig. 19 shows some sample images collected from the Internet. 132 of them from the 67 lower legs were combined with the Image Set 2 to form a gallery set with 668 images from 479 lower legs. The other images collected from the Internet were used a probe set. Since the leg boundaries are not a stable

feature to align these images, which may have viewpoint, scaling, illumination, pose, blurry, out of focus, and shadow problems, the alignment scheme was skipped and the positive sample generation scheme was still used with the same set of parameters for handling different poses of legs. Fig. 20 gives the CMC curves. The experimental results show that the proposed algorithm can achieve the rank-1 accuracy of 29% and the rank-10 accuracy of 56%. As with the previous experiments (Fig. 11), the proposed method outperforms Su et al.'s method and the deep learning methods with significant margins. One difference is that the two deep learning methods perform better than Su et al.'s method. They automatically discover the similarity between the Internet images in the probe and gallery sets and therefore, they perform better than Su et al.'s method. However, the size of dataset is not enough for them to learn effective features to identify individuals. The experimental results demonstrate that androgenic hair has good potential for addressing the forensic needs.

Fig. 19. Sample images in the Internet dataset. (a) and (b) are images in the probe set and the gallery set, respectively. The numbers in (a) and (b) indicate the corresponding images from the same legs. Note that the relative scales are kept for comparison and they are resized to 142×298 pixels for the experiments.

Fig. 20. CMC curves computed from the images collected from the Internet.

5. Discussion and Conclusions

The previous study exposed the potential of androgenic hair patterns in low resolution images as a soft biometric trait for forensic applications [23]. However, the previous results gave an impression that androgenic hair is not a discriminative biometric trait, like other soft biometric traits such as gender, age, and race. To further study androgenic hair as a soft biometric trait, in this paper, a new algorithm, which is composed of an alignment scheme based on leg geometry, multi-directional grid systems, class-specific PLS models, and a positive sample generation scheme, is utilized. The proposed algorithm achieves accuracy of over 99% at the 10th rank for the frontal leg image database and over 59% at the 30th rank for the low quality image database. It also achieves accuracy of 56% at the 10th rank for the Internet images. These results demonstrate that the impression is due to the previous method, DGOH, but not androgenic hair.

In forensic applications, none of single biometric trait is suitable for all cases. For example, face is not applicable for identifying masked rioters. Even though hard biometric traits are available, the quality of signals may not be enough for accurate matching. Traditional soft biometrics have been used to combine hard biometrics and to narrow down suspect lists. Comparing with other soft biometric traits such as gender, race, and age, androgenic hair even in low resolution images has valuable discriminative information and is worth for further study.

The testing database in this study contains 1,725 images from 479 legs, which is the largest one comparing with all the previous studies. We admit that its size is not large comparing with face recognition databases, which have been studied nearly for a half-century and by numerous researchers. However, comparing with other public biometric databases such as UBIRIS.v1 iris database collected from 241 irises [10], CMU keystroke dynamic database collected from 51 subjects [19], and CASIA multi-spectral palmprint database collected from 100 subjects [20], the size of our database is comparable. In some forensic cases, the gallery set is not necessary very large because other information can significantly narrow down suspect lists. John Jihadi's case is an example. His British accent and terrorist activities gave important hints to law enforcement agencies to limit the suspects to U.K. terrorists in Middle East. In some extreme cases, it is in fact a verification problem, one-to-one matching. For example, gait recognition was applied to a legal case for verifying a suspect [2].

The rank-1 accuracy of the proposed algorithm for the Low Quality Set is not very high. However, its rank-20 and rank-30 accuracy is respectively more than 50% and 59%, which are 20% more than the previous work. Some may comment that this accuracy is still low. It should be noted that accuracy of forensic identification is generally low because of quality of signals. Paulino et al. reported that VeriFinger, a commercial fingerprint matcher, achieves accuracy of 30.6% for NIST SD27 latent fingerprints with ugly quality [8]. Comparing with other soft biometric traits such as gender and age, the accuracy reported in this paper has demonstrated the discriminative power of androgenic hair. It should be re-emphasized that images captured by surveillance cameras are not within the scope of this study.

Acknowledgement

This work is partially supported by the Ministry of Education, Singapore through Academic Research Fund Tier 2, MOE2012-T2-1-024.

References

- [1] A. Dantcheva, P. Elia, and A. Ross, What else does your biometric data reveal? A survey on soft biometrics, *IEEE Trans. Inf. Forensics Security* 11 (2016) 441-476.
- [2] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, On using gait in forensic biometrics, *J. of Forensic Science*, 56 (2011) 882-889.
- [3] D.A. Reid, M.S. Nixon, and S.V. Stevenage, "Soft biometrics: human identification using comparative descriptions", *TPAMI* 36 (2014) 1216-1228.
- [4] eMarketer, 2 billion consumers worldwide to get smart (phones) by 2016. <http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694>, 2014 (accessed 11.12.14).
- [5] H. Qu, G. Li, and Q. Xu, An asymmetric classifier based on partial least squares, *Pattern Recognition* 43 (2010) 3448-3457.
- [6] C.R. Maurer, R. Qi, and V. Raghavan, A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions, *TPAMI* 25 (2013) 265-270.
- [7] M.R. Islam, F.K.S. Chan, and A.W.K. Kong, A preliminary study of lower leg geometry as a soft biometric trait for forensic investigation, *ICPR* (2014) 427-431.
- [8] A.A. Paulino, J. Feng, and A.K. Jain, Latent fingerprint matching using descriptor-based Hough transform, *Int. Joint Conf. on Biometrics* (2011) 1-7.
- [9] X.Q. Zeng and G.Z. Li, Incremental partial least squares analysis of big streaming data, *Pattern Recognition*, 47 (2014) 3726-3735.
- [10] H. Proenca and L.A. Alexandre, UBIRIS: a noisy iris image database, *Int. Conf. on Image Anal. and Process.* 1 (1996) 790-777.
- [11] Business Insider, Huge protests continue in Brazil — small bands of radicals break into stores and loot. <http://www.businessinsider.com/huge-protests-continue-in-brazil--small-bands-of-radicals-break-into-stores-and-loot-2013-6>, 2013 (accessed 19.06.13).
- [12] F.K.S. Chan, X. Li, and A.W.K. Kong, A study of distinctiveness of skin texture for forensic applications through comparison with blood vessels, accepted in *IEEE Trans. Inf. Forensics Security*.
- [13] R. Rosipal and N. Kramer, Overview and recent advances in partial least squares, *Proc. Int. Conf. Subspace, Latent Structure and Feature Selection* (2005) 34-51.
- [14] T. Ojala, M. Pietikainen, and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *TPAMI* 24 (2002) 971-987.
- [15] T. Ahonen, A. Hadid, and M. Pietikainen, Face recognition with local binary patterns, *ECCV* (2004) 469-481.
- [16] Z. Liu and C. Liu, Fusion of color, local spatial and global frequency information and face recognition, *Pattern Recognition* 43 (2010) 2882-2890.
- [17] J. Y. Choi, Y.M. Ro, and K.N. Plataniotis, Color local texture features for color face recognition, *IEEE Trans. Image Process.* 21 (2012) 1366-1380.
- [18] J.L. Crowley and F. Berard, Multi-modal tracking of faces for video communications, *CVPR* (1997) 640-645.
- [19] K. S. Killourhy and R. A. Maxion., Comparing anomaly detectors for keystroke dynamics, *Annu. Int. Conf. on Dependable Syst. and Networks* (2009) 125-134.
- [20] Y. Hao, Z. Sun, T. Tan, and C. Ren, Multi-spectral palm image fusion for accurate contact-free palmprint recognition, *ICIP* (2008) 281-284.
- [21] The Guardian, A new kind of riot? From Brixton 1981 to Tottenham 2011. <http://www.guardian.co.uk/uk/2011/dec/09/riots-1981-2011-differences>, 2011 (accessed 09.12.11).
- [22] CBC News, A tale of two riots. <http://www.cbc.ca/news/canada/story/2011/06/16/f-vancouver-riot-1994-2011.html>, 2011 (accessed 16.06.11).
- [23] H. Su and A.W.K. Kong, A study on low resolution androgenic hair patterns for criminal and victim identification, *IEEE Trans. Inf. Forensics Security* 9 (2014) 666-680.

- [24] Beyond Borders Au- delà des frontières ECPAT CANADA, Child abuse images – statistics. <http://www.beyondborders.org/wp/wp-content/uploads/Child-Abuse-Images-Statistics.pdf>, 2012 (accessed 30.06.12).
- [25] A. Kembhavi, D. Harwood, and L.S Davis, Vehicle detection using partial least squares, TPAMI 33 (2011) 1250-1265.
- [26] W.R. Schwartz, H. Guo, J. Choi, and L.S. Davis, Face identification using large feature sets, IEEE Trans. Image Process. 21 (2012) 2245-2255.
- [27] A. Myronenko and S. Xubo, Point set registration: coherent point drift, TPAMI 32 (2010) 2262-2275.
- [28] F.K.S. Chan and A.W.K. Kong, Using leg geometry to align androgenic hair patterns in low resolution images for criminal and victim identification, ICPR (2014) 495-500.
- [29] W. Montagna, R.A. Ellis (eds), The biology of hair growth, Academic Press, New York, 1958.
- [30] K.S. Stenn and R. Paus, Controls of hair follicle cycling, Physiological Reviews 81 (2001) 449-491.
- [31] A. Nurhudatiana, A.W.K. Kong, K. Matinpour, S.Y. Cho, and N. Craft, Fundamental statistics of relatively permanent pigmented or vascular skin marks for criminal and victim identification, Proc. Int. Joint Conf. on Biometrics (2011) 1-6.
- [32] C. Tang, A.W.K. Kong, and N. Craft, Uncovering vein patterns from color skin images for forensic analysis, CVPR (2011) 665-672.
- [33] M. Mignotte, Segmentation by fusion of histogram-based k-means clusters in different color spaces, IEEE Trans. Image Process. 17 (2008) 780-787.
- [34] LIBLINEAR – A library for large linear classification. <http://www.csie.ntu.edu.tw/~cjlin/liblinear/> (accessed 27.09.15).
- [35] U. Blume-Peytavi, A. Tosti, D.A. Whiting and R.M. Trüeb (Eds) Hair Growth and Disorders, Springer, 2008.
- [36] S. M. Garm, Types and distribution of the hair in man, Annals of the New York Academy of Sciences 53 (1951) 498-507.
- [37] A. Vogt, K.J. McElwee, and U. Blume-Peytavi, Biology of the hair follicle, in U. Blume-Peytavi, A. Tosti, D.A. Whiting and R.M. Trüeb (Eds) Hair Growth and Disorders, Springer, 2008.
- [38] A. K. Jain, S. C. Dass, and K. Nandakumar, Can soft biometric traits assist user recognition?, Proc. SPIE 5404 (2004) 561-572.
- [39] A. K. Jain, S. C. Dass, and K. Nandakumar, Soft biometric traits for personal recognition systems, the 1st Int. Conf. Biometric Authentication (2004) 731-738.
- [40] A. Dantcheva, A. Singh, P. Elia, and J.-L. Dugelay, Search pruning in video surveillance systems: Efficiency-reliability tradeoff, ICCV workshop (2011) 1356-1363.
- [41] A. Nurhudatiana, A. Kong, K. Matinpour, D. Chon, L. Altieri, S.-Y. Cho, and N. Craft, The individuality of relatively permanent pigmented or vascular skin marks (RPPVSM) in independently and uniformly distributed patterns, IEEE Trans. Inf. Forensics Security 8 (2013) 998-1012.
- [42] A.W.K. Kong, D. Zhang, and G. Lu, A study of identical twins' palmprints for personal verification, Pattern Recognition 39 (2006) 2149-2156.
- [43] A.K. Jain, S. Prabhakar and S. Pankanti, On the similarity of identical twin fingerprints, Pattern Recognition 35 (2002) 2653-2663.
- [44] N. Srinivas, G. Aggarwal, P.J. Flynn and R.W. Vorder Bruegge, Analysis of facial marks to distinguish between identical twins, IEEE Trans. Inf. Forensics Security, 7 (2012) 1536-1550.
- [45] H. Zhang, C. Tang, X. Li, and A.W.K. Kong, A study of similarity between genetically identical body vein patterns, IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, 2014.
- [46] F. Schroff, D. Kalenichenko, and J. Philbin, FaceNet: A unified embedding for face recognition and clustering, CVPR (2015).
- [47] Y. Wen, K. Zhang, Z. Li and Y. Qiao, A discriminative feature learning approach for deep face recognition, ECCV (2016) 499-515.
- [48] C.H. Ho and C.L. Lin, Large-scale linear support vector recognition, Journal of Machine Learning Research, 13 (2012) 3323-3348.
- [49] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," TPAMI 27 (2005) 1226-1238.
- [50] Daily Mail, The £25bn bombshell: rioters stone MP after Greece pushes through austerity package. <http://www.dailymail.co.uk/news/article-2009437/Greek-vote-Riotersstone-MP-Greece-pushes-austerity-package.html>, 2011 (accessed 30.06.11).