Recognizing Human Faces under Disguise and Makeup

Tsung Ying Wang, Ajay Kumar

Department of Computing, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
cstywang@comp.polyu.edu.hk, csajaykr@comp.polyu.edu.hk

Abstract

The accuracy of automated human face recognition algorithms can significantly degrade while recognizing same subjects under make-up and disguised appearances. Increasing constraints on enhanced security and surveillance requires enhanced accuracy from face recognition algorithms for faces under disguise and/or makeup. This paper presents a new database for face images under disguised and make-up appearances the development of face recognition algorithms under such covariates. This database has 2460 images from 410 different subjects and is acquired under real environment, focuses on make-up and disguises covariates and also provides ground truth (eye glass, goggle, mustache, beard) for every image. This can enable developed algorithms to automatically quantify their capability for identifying such important disguise attribute during the face recognition. We also present comparative experimental results from two popular commercial matchers and from recent publications. Our experimental results suggest significant performance degradation in the capability of these matchers in automatically recognizing these faces. We also analyze face detection accuracy from these matchers. The experimental results underline the challenges in recognizing faces under these covariates. Availability of this new database in public domain will help to advance much needed research and development in recognizing make-up and disguised faces.

1. Introduction

With the rapid development in advanced algorithms and improvements in the hardware, the accuracy of recognizing clear human faces under well-controlled environment is widely believed to be very high. However, such accuracy is known to significantly degrade [16]-[19] for recognizing faces real life faces under a wide variety of covariates. One of the key obstacles in advancing the research and benchmarking efforts in this area is related to lack of representative and larger databases with ground truth annotations that can be used not only to evaluate disguised face recognition or detection performance but also to evaluate the disguised accessories like eye glass, mustache, beard or goggle.

Figure 1: Image pairs with different types of Makeup/Disguise

1.1. Related Works

Recognition of disguised and make-up faces is inviting increasing attention in the literature. The increasing need for security at public places and installations has necessitated development of advanced face recognition systems that can operate using surveillance cameras and

Table 1: Summary face image databases available for evaluating disguised face recognition capabilities.

Database	No. of subjects	Total number of images	Focus on facial disguise	Publicly Available	Availability of ground truth
This paper	410	2460	YES	YES	YES
Labeled Face in the Wild [1]	5749	58797	NO	YES	NO
Public Figures Face Database [2]	200	13233	NO	YES	NO
Disguised Face Database [5]	325	500000	YES	NO	NO
Makrup in Wild Database [14]	125	154	YES	YES	NO
IIIT-Delhi Disguise Face Database [10]	75	681	YES	YES	NO

match disguised faces under real environment. On the other hand, outlooks and makeups have always been a major consideration for humans since we invented clothes. No matter the purpose is to look better or to hide identity, we invented makeups, masks, hats and a wide variety of accessories to wear on our faces, as also visible from sample images shown in figure 1. Under such conditions, traditional face detection and recognition methods becomes very difficult and challenging.

There are some open access human face databases focus on. human face images acquired under natural environment, such as Labeled Face in the Wild (LFW) [1], Public figures face database (Pubfig) [2], and the Yale face database [4]. However these databases mostly focus on inherent imaging covariates with least or non-cooperative subjects. Therefore a wide majority of images can represent large variation in pose, lighting, expression, scene, camera, imaging conditions and parameters. However most of these subjects were not having any makeup(s) and this may not meet the research interest relating to evaluating face recognition capability under disguise and makeup. Furthermore, some databases are also smaller and much shallower regarding image acquisition and organization. The Disguised Face Database maintained by the University of Pennsylvania [5] is similar in spirit to our interest, but it is not freely available to the public anymore. Table 1 provides a brief comparison among some of the databases mentioned above. Makeup in wild database [14], [18]-[19], is another useful database available to ascertain covariates from facial makeup. The IIIT-D Disguise Face Database [10] is publicly available and well-organized, but some of its "disguise" covers too much of human face's area, which is not realistic in real world and nearly impossible to distinguish between different person, even by human eyes. This paper [19] utilized database acquired from several makeup tutorials and builds the database of 75 subjects. This is an exciting work but the numbers of subjects are quite limited while the images are mostly acquired under controlled or in indoor environment. The disguised and makeup face database introduced in this paper, which contains 2460 face images from 410 different subjects, has been gathered from the celebrities (mostly famous movie stars or other celebrities). All the subjects are also least-cooperative and acquired under uncontrolled or least controlled environment. Usually within the six images for a subject, 1–2 images are from clean/clear faces and the others have light or heavy makeup or are in disguise.

The key challenge in this database is largely pertaining to matching face images acquired under real environment with several covariates including those in pose, distance, occlusion and illumination. Therefore it is extremely hard to utilize this database to evaluate performance of face recognition algorithms for face images acquired from close range with disguise and/or make-up as the key covariate. The development of algorithms to evaluate their invariance

to make-up and disguise is important consideration while selecting biometrics systems based on face imaging. Therefore there is a pressing need to develop a face image database that can present such frontal face images which has disguise and make-up as the key covariate and whose images have ground truth made available so that effectiveness of automatically identifying various attributes (eye glass, beard, mustache, etc.) can be evaluated by the algorithms that need to be developed by the biometrics community.

1.2. Our Work

The key contributions from this paper can be summarized as in the following. Firstly, this paper provides a new database of disguise and/or make-up faces in public domain. This database from 410 different subjects illustrates challenging face images acquired under real environments. Table 1 presents a comparative summary of earlier databases available and employed for developing face recognition algorithms. The database developed in this work is first of its kind and includes ground truth for every images that illustrate presence of glasses, goggles, mustache, beard, that will enable researchers to develop advanced algorithms in accurately recognizing such real faces under disguised appearances. Secondly, we also provide a comparative performance evaluation from the two popular commercial matchers and an effective matcher for recognizing disguised faces proposed in recent journal publication [7]. The experimental results also illustrate challenges in automatically detecting face images from two commercial matchers, along with publicly available trained detector [6], and need for further work in this area.

This paper is organized as follows. The acquisition of images and its availability for the researchers is described in section 2. This section also describes automated face detection and segmentation approach utilized for evaluating performance for the different matchers. Section 3 introduces the different matchers utilized in the performance evaluation. This section includes details on patched LBP which was also employed for the performance comparison. The experimental protocols and results from this database are presented in section 4 while the ley conclusions from this work are summarized in Section 5.

2. The Disguised and Makeup Faces Database

2.1. Acquisition of Images

The disguised and makeup faces database [12] is a human face dataset consisting of 2460 images of 410 different celebrities. The majority of images in this dataset are from celebrities which are movie/TV stars, while some of them are politicians or athletes. All the face images are collected

directly from the publicly available websites which are clearly cited in the database. We have ensured that the first image for each of the subject is frontal image without disguise or with least makeup. Within images of each subject, there must have at least one to two "clean" (pure face, without any makeups or disguises) face images. This serves as reference image that can be used to generate matches with other face images or in any other protocol to analyze the effectiveness of a face matcher. Rest of the images will have various types of disguises, including glasses, hairstyles, beard and other facial accessories.

Figure 2: Sample images for which the employed face detectors failed to correctly detect face region.

2.2. Face Segmentation and Preprocessing

The very first work after acquiring the images is to segment out the region of interest (ROI) for further processing. We evaluated the Viola-Jones face detection, implemented in the OpenCV library [6] as well as the face details extractor provided by VeriLook [8] to automatically segment the face region. The quality of some of images is not suitable to be detected by these face detectors. Figure 2 illustrates some image samples which were not able to be detected or generated very poor face detection results. Such face images show subjects which are usually heavily disguised (like large goggles, masks, hats, tattoos); or most of the important facial features are covered (but still recognizable by the human eyes). In order to accommodate such problem, we used two different approaches to compute and analyze the effect of these "detection failure" and is discussed in section IV. The face detection rate using our dataset was 77.4% (559 images out of 2460 images failed for the face detection either by using VeriLook [8] or Face++ matcher [9]).

.Every ROI image is normalized into 300×300 pixels, furthermore, to achieve better results, we applied two popular image quality enhancement methods [11], as shown in figure 3, *i.e.*, Single Scale Retinex (SSR) and quotient enhancement to eliminate or suppress the influence from shadow and/or illumination variations. In most of our tests, the ROI with SSR enhancement test set outperformed the other set. Therefore we only used this set of enhanced face images to perform further experiments and evaluations.
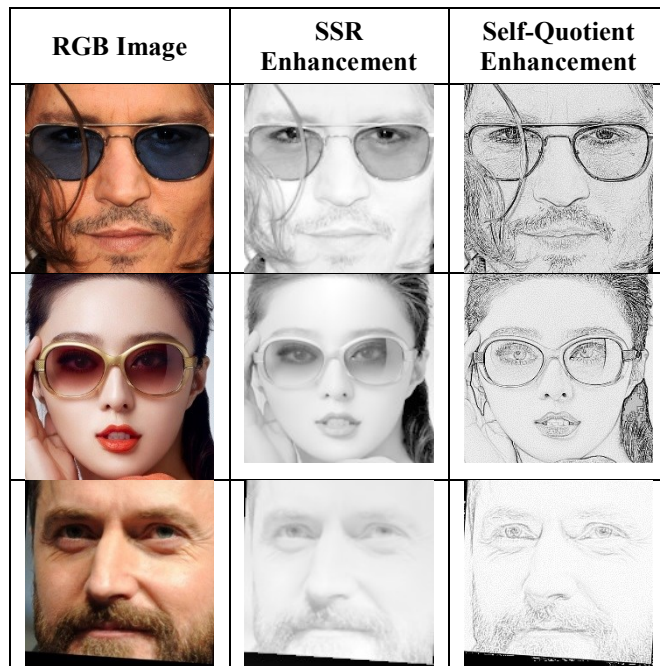

Figure 3: SSR and self-quotient enhancement on detected faces.

2.3. Ground Truth

A typical example for the usage of this database is for the surveillance or at border crossing checks, where the face images are usually acquired in less-constrained conditions with unclear facial feature, altered/occluded facial features due to makeup, or intentional disguises. Any automated alert on the level of disguise can help border crossing inspectors to initiate secondary or careful checks. Such makeup index will require automated detection of beard, mustache, goggles, eye glasses [13], or other makeup accessories influencing the appearance of natural biometric features. It is therefore necessary that automated detection of such accessories be accurately done and evaluating capability of available algorithms require database of images which have ground truth of such features (not available in other databases listed in table 1). It is also well-known that soft-biometrics features such as skin color [20]-[22], mustache/hair [15], [20] or gender [20]-[22] can be used to significantly increase the accuracy of face recognition systems/algorithms. However deployment of such capability requires automated extraction of such features and this again requires a database which has such ground truth on soft biometrics visible from facial images. A dynamic evaluation of such soft biometric attributes and key locations can also help to thwart spoof attacks using sophisticated facial masks. Therefore this paper develops such a databases that includes important attributes which can be considered as soft biometrics and also accessories to ascertain level of disguise while evaluating face recognition capability for images under make-up and

disguise. The database developed in this work provides following ground truth attributes corresponding to human inspection of each of the images in the database:

- File Name
- File Size
- Gender (Male; Female)
- Ethnicity (European; African; Asian; Others)
- Skin Color (Dark/Black; Yellow; Light/White)
- Hat (0 = No; 1 = Yes)
- Hair style (0 = Bang covering forehead; Others =1)
- Glasses (0 = No; 1 = Transparent; 2 = Dark)
- Beard (0 = No; 1 = Mustache; 2 = Goat patch; 3 = Chin curtain; 4 = Mustache + Goat patch; 5 = Mustache + Chin curtain; 6 = Goat patch + Chin curtain; 7 = All)

3. Experiments for Performance Evaluation

In order to ascertain the matching accuracy for the face images in the developed database, we performed a series of experiments using popular commercial and face matchers in the prior references/work. The LBP matcher has been used in earlier work to ascertain in matching accuracy of disguised faces and therefore this matcher was also selected for the performance evaluation. This matcher is briefly described in the following.

3.1. Local Binary Pattern (LBP)

The LBP features does not consider whole image as a high-dimensional vector, but describe only local features of the given image. The features extracted using this approach is expected to have a low-dimensionality implicitly. In our implementation, each image region of interest (ROI) is partitioned into 64 blocks (8*8); generate the histogram for each block; compare these histograms using Chi-squared distance to generate a similarity score. These histograms are commonly referred to as Local Binary Patterns Histograms. Our experiments employed basic LBP operator with 8 sampling points. More details on this matcher are available in [6]. Two arrays were constructed: one that stores all the image matrixes, the other one stores the ID labels (the images from the same person will be labeled as same ID). The genuine and impostor matches from the images in the database are used to generate

receiver operating characteristics for the performance evaluation.

3.2. Local Binary Patterns with Blocks

Besides the traditional LBP method, we implemented the approach described in [7]. The basic idea in this approach is to distinguish the image blocks into two categories, biometric and non-biometric. The blocks that are expected to illustrate human skin, we classify as biometric. On the other hand, blocks illustrate non-human components, such as hair, scarves, glasses and mask, we can consider them as non-biometric block. We can discard the non-biometric block, only use and compute matching scores from the biometric blocks. This identification is achieved from the intensity of grey-levels in each block of the ROI image. A block is classified in (1), as the biometric if the intensity score is greater than or equal to a threshold, and is classified as non-biometric if the score is smaller than the threshold, as illustrated in figure 4.

$$B_{i,j} = \begin{cases} Biometric\ if & S \geq T \\ Non-biometric\ if & S < T \end{cases} \quad (1)$$

The threshold T is computed offline during the training phase and is kept fixed during the test or the evaluation phase.

3.3. Commercial Face Matchers

The performance from the two commercial matchers on the images in the developed database was also evaluated. These two commercial matchers were VeriFace from NeuroTech [8] and Face++ [9]. These two commercial matchers also provide free access to their evaluation software and was used in this work. These two commercial matchers also have their own face detection algorithm; therefore the raw image (instead of the ROI) was used to generate the matching scores from genuine and impostor matches.

4. Experiments and Results

In this section, we present experimental results using the database developed in this work. It may be noted that face detection accuracy of different commercial matchers and also the one available in OpenCV library [6] (used along with LBP matcher) are different. Therefore our experiments considered different protocols and sets for
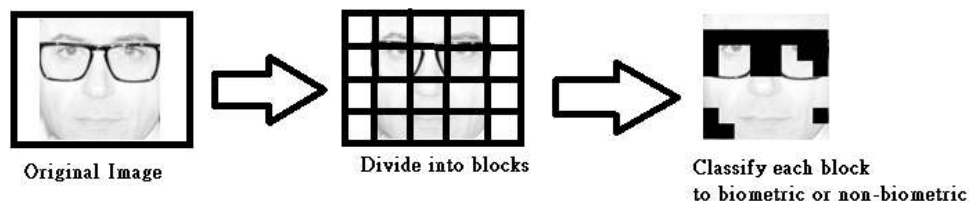

Figure 4: Matching disguised face images using blocks with biometric contents.

more fair assessment of matching accuracy. We conducted two sets of experiments using three different protocols: protocol A, protocol B and protocol C. Set A computes the matching scores on all-to-all basis, the most naïve method. While set B stores the largest (closest) matching scores as the score for one input image. For instance, if there are X subjects with each subjects has Y images, set A will generate XY(Y-1) genuine scores and XY(XY-Y) imposter scores, while set B will have XY genuine scores and XY(X-1) imposter scores. Set C uses the first or the frontal image with no or least makeup for the registration while rest of the images are utilized for test or probe. This evaluation can help to ascertain the capability of matcher to match disguised faces with single (training) image acquired during registration. The effectiveness of a matcher is evaluated using the ROC.

As discussed in section 2.2, two different approaches, that considers failure of face detection differently, were considered for the performance evaluation. The objective for choosing two different methods is to ascertain effectiveness of face matcher while considering *accurate* face detection and also to ascertain the performance in its totality, *i.e.*, considering face detection failure as a part of failure to match two face images.

- Method 1: This approach ignores all the face images where the face detection (from any of the three matchers) is unsuccessful.
- Method 2: This approach considers unsuccessful face detection (from any of the matchers) as a failed match and the match score is therefore assigned as zero (non-match).

The exact number of genuine and impostor match scores from three different protocols and two different methods is summarized in table 3. The ROCs corresponding to these different matchers is shown in Figure 5-10. The experimental results suggest superiority of commercial matcher (Face++) over the other two matchers considered in this work.

Table 2: Match scores under different protocols/methods.

	Genuine	Imposter
Method 1-A	10406	5090411
Method 2-A	14760	6036840
Method 1-B	2111	850287
Method 2-B	2460	167690
Method 1-C	1748	762148
Method 2-C	2050	809340

Table 3: Equal error rate from two commercial matchers.

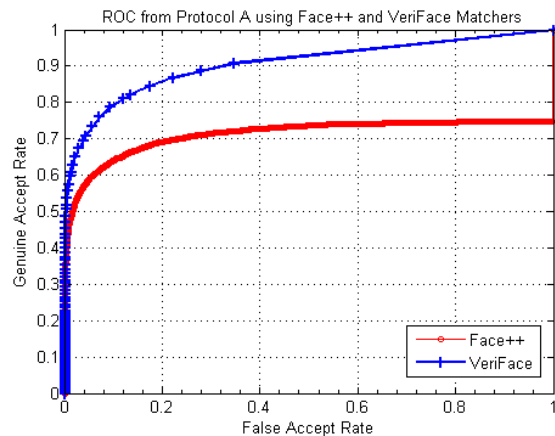| Protocol and Method	VeriLook	Face++
Protocol A Method 1	16.5%	12.2%
Protocol A Method 2	24%	16.2%
Protocol B Method 1	8.1%	8.3%
Protocol B Method 2	16.9%	23%
Protocol C Method 1	13.7%	10.5%
Protocol C Method 2	20.2%	14.06%

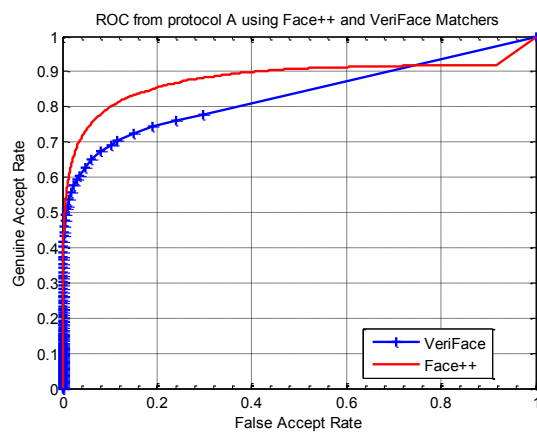**Figure 5**: The ROC from the commercial matchers using Protocol A with Method 1.

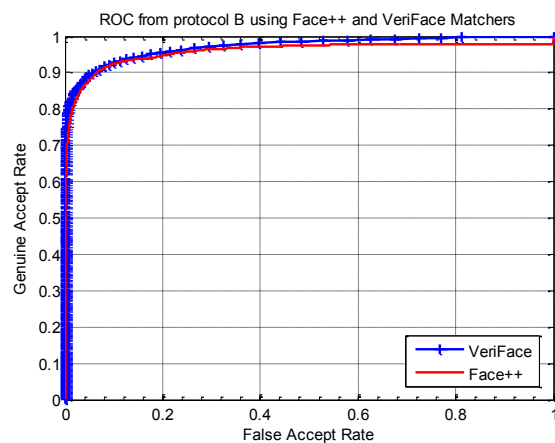**Figure 6**: The ROC from the commercial matchers using Protocol A with Method 2.

Figure 7: The ROC from the commercial matchers using Protocol B with Method 1.
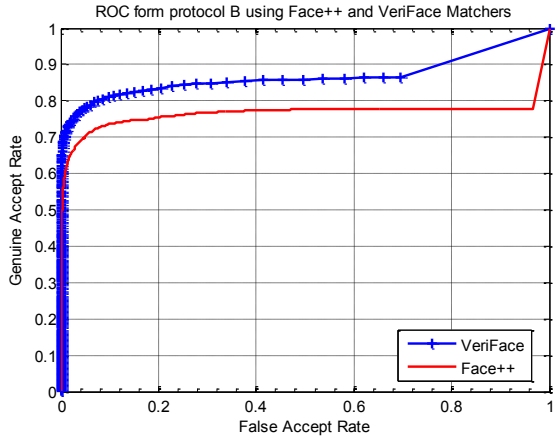
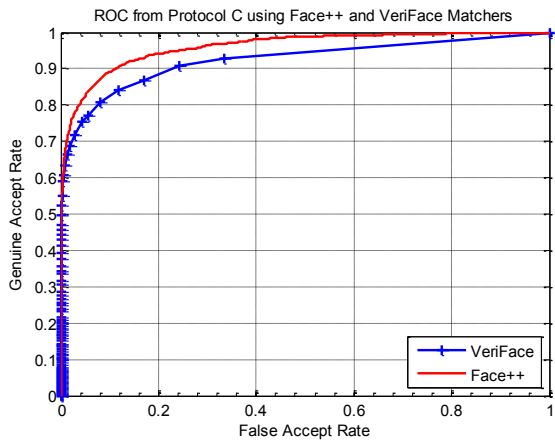**Figure 8**: The ROC from the commercial matchers using Protocol B with Method 2.

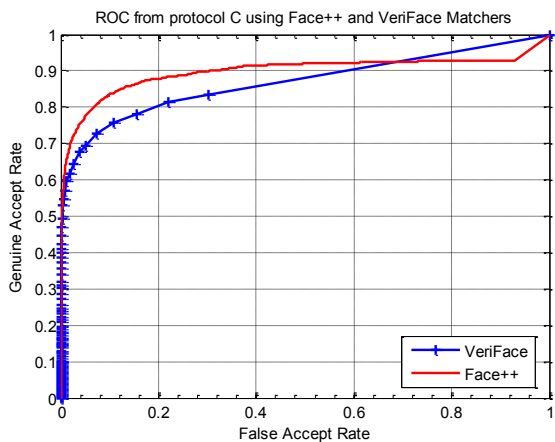**Figure 9**: The ROC from the commercial matchers using Protocol C with Method 1.

Figure 10: The ROC from the commercial matchers using Protocol C with Method 2.
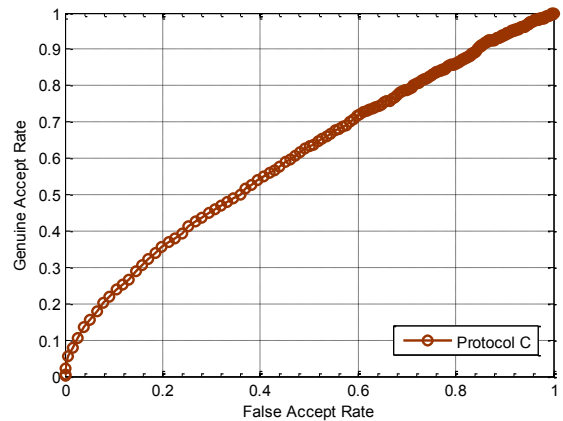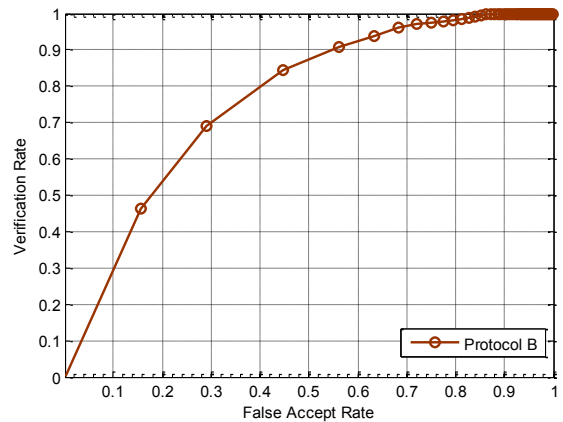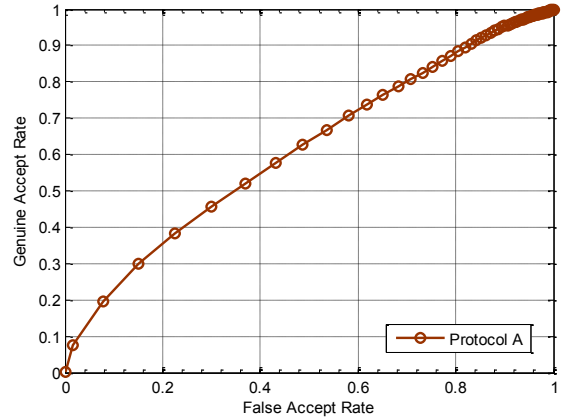

Figure 11: The ROC from the approach detailed in reference [7] using three different protocols.

Our experiments on evaluating the face detection capability suggest degradation in face detection performance for detecting a range of faces with accessories and makeup faces in this database. The best experiment results using this database is from the VeriFace matcher using protocol B which can be attributed to large number of training or gallery images employed with this protocol.

However it still has large misdetection rate (cannot correctly detect face, or the detected face region is incorrect). The accuracy of LBP based algorithm [7] largely depends on accuracy of labelling or predicting a biometric block under photometric variations in the ROI images. It is reasonable to expect that with the increase in number of training images the prediction accuracy for biometric clocks can increase. However, manual labelling of 300 images (for example) with their correct labels would need 300*64 = 19200 blocks to be correctly labeled. It is possible to significantly increase the performance using [7] if we can correctly label the biometric and non-biometric blocks. Such failure to correctly identify the biometric blocks is likely the reason for poor performance illustrated from ROCs shown in figure 11.

5. Conclusions and Further Work

This paper has developed a new database [12] of disguised and make up faces, along with ground truth labels to advance research on recognizing faces with disguise and makeup. The availability of ground truth labels like ethnicity, mustache, glasses or skin color will also help in the development of algorithms to automatically detect and recognize such soft biometrics accessories which can be used to enhance face recognizing accuracy or for indexing large scale face databases. We have also presented experimental results using two popular commercial matchers and another matcher investigated in recent reference [7]. The experimental results in section 4 using these matchers [8]-[9] are however indicative performances (rather than absolute) and can possibly be improved further with better choice of parameters or setting. Presence of accessories for makeup and disguise, like hairstyle, goggles, occlusions, hat and beard, can significantly impact or degrade the face detection capability of the face detection matchers available today. The experimental results presented in this paper suggest that face recognition capability of the considered matchers for recognizing disguised faces is quite poor. It is expected that availability of database from this work will help to advance research in the development of face matchers that can help to accurately recognize disguised and makeup faces for their commercial deployment.

Acknowledgment

References

[1]   Labeled Face in the Wild. http://vis-www.cs.umass.edu/lfw/

[2]   The Public figures face database.
http://www.cs.columbia.edu/CAVE/databases/pubfig/

[3]   The YouTube faces database.
http://www.cs.tau.ac.il/~wolf/ytfaces/

[4]   Yale Face Database.
http://vision.ucsd.edu/content/yale-face-database

[5]   The Disguised Face Database
http://www.cis.upenn.edu/~gibsonk/dfd/

[6]   Face Recognition with OpenCV .
http://docs.opencv.org/modules/contrib/doc/facerec/facerec_tutorial.html

[7]   T. Dhamecha, R. Singh, M. Vatsa, A. Kumar, "Recognizing disguised faces: human and machine evaluation," *PLoS ONE* 9(7): e99212. doi:10.1371/journal.pone.0099212, 2014.

[8]   Neurotechnology, VeriFace SDK.
http://www.neurotechnology.com/

[9]   Face++: Leading Face Recognition on Cloud.
www.faceplusplus.com/

[10] IIIT-D Disguise version 1 face database,
https://research.iiitd.edu.in/groups/iab/facedisguise.html

[11] T. H. N. Le, K. Luu, K. Seshadri, and M. Savvides, "Beard and mustache segmentation using sparse classifiers on self-quotient images," *Proc. ICIP 2012*, Orlando, USA, pp. 165-168, Sept. 2012.

[12] Weblink for downloading *The Hong Kong Polytechnic University Disguise and Makeup Faces Database* described in this paper, 2016, available
http://www.comp.polyu.edu.hk/~csajaykr/DMFaces.htm

[13] D. Yi and S. Z. Li. "Learning sparse feature for eyeglasses problem in face recognition," *Proc. Automatic Face & Gesture Recognition and Workshops*, pp. 430,435, Santa Barbara, USA, Mar. 2011.

[14] Makeup in Wild Database,
http://www.antitza.com/makeup-datasets.html

[15] E. S. Jaha, and M. Nixon, "Soft biometrics for subject identification using clothing attributes," *Proc. IJCB 2014*, Clearwater, Florida, Sep-Oct.. 2014.

[16] K. K. Kim, J. Y. Lee, H. S. Yoon, J. H. Kim, J. C. Sohn, "System for recognizing disguised face using Gabor feature and SVM classifier and method thereof," *US Patent No.* US8913798 B2, Dec. 2014.

[17] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," *Proc. ICB 2013*, Madrid, Jun. 2013. .

[18] C. Chen; A. Dantcheva, A.; Ross, "Automatic facial makeup detection with application in face recognition," *Proc. ICB 2013*, pp.1-8, June 2013.

[19] A. Dantcheva, C. Chen, A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?," *Proc. BTAS 2012*, pp.391,398, Sept. 2012.

[20] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE. Trans. Info. Forensics & Security*, vol.9, no.3, pp.464,475, Mar. 2014.

[21] D. A. Reid, M. S. Nixon, S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," *IEEE Trans. Pattern Analysis & Machine Intell.* vol.36, no.6, pp.1216,1228, Jun. 2014.

[22] D. A. Reid, M. S. Nixon, "Using comparative human descriptions for soft biometrics," *Proc. IJCB 2011*, Washington DC, Oct. 2011.