Attribute-based Continuous User Authentication on Mobile Devices

Pouya Samangouei, Vishal M. Patel, and Rama Chellappa
Center for Automation Research
University of Maryland, College Park, MD 20742
{pouya, pvishalm, rama}@umiacs.umd.edu

Abstract

We present a method using facial attributes for continuous authentication of smartphone users. The binary attribute classifiers are trained using PubFig dataset and provide compact visual descriptions of faces. The learned classifiers are applied to the image of the current user of a mobile device to extract the attributes and then authentication is done by simply comparing the difference between the acquired attributes and the enrolled attributes of the original user. Extensive experiments on two publicly available unconstrained mobile face video datasets show that our method is able to capture meaningful attributes of faces and performs better than the previously proposed LBP-based authentication method.

1. Introduction

Advances in communication and sensing technologies have led to an exponential growth in the use of mobile devices such as smartphones and tablets. Mobile devices are becoming increasingly popular due to their flexibility and convenience in managing personal information. Traditional methods for authenticating users on mobile devices are based on passwords, pin numbers, secret patterns or fingerprints. As long as the mobile phone remains active, typical devices incorporate no mechanisms to verify that the user originally authenticated is still the user in control of the mobile device. Thus, unauthorized individuals may improperly obtain access to personal information of the user if a password is compromised or if a user does not exercise adequate vigilance after initial authentication on a device.

To deal with this problem, various continuous authentication (also known as active authentication) systems have been developed in which users are continuously validated after the initial access to the mobile device. For instance, [8], [7], [17] proposed to continuously authenticate users based on their touch gestures or swipes. Gait as well as device movement patterns measured by the smartphone accelerometer were used in [5], [14] for continuous authentication. Stylometry, GPS location, web browsing behavior, and application usage patterns were used in [9] for active authentication. Face-based continuous user authentication has also been proposed in [10], [6], [13]. Fusion of speech and face was proposed in [13] while [3] proposed to fuse face images with the inertial measurement unit data to continuously authenticate the users. A low-rank representation-based method was proposed in [16] for fusing touch gestures with faces for continuous authentication. Finally, a domain adaptation method was proposed in [18] for dealing with data mismatch problem in continuous authentication.

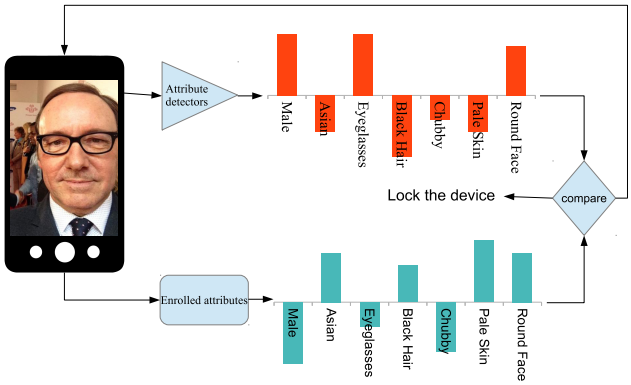


Figure 1. Overview of our attribute-based authentication method.

Most face-based authentication systems use representation-based features and hence perform poorly when the testing environment is different from where enrollment occurred. This is clearly explored in [6] where enrollment and testing sessions don't overlap leading to notable lower accuracy compared to when enrollment and testing sessions overlap. Facial attributes ideally should remain the same under different background or lighting conditions which makes them more robust to changes in acquisition conditions.

In this paper, we present an attribute-based continuous authentication system for smartphone users. Figure 1 gives an overview of the proposed attribute-based continuous authentication method. Given a face image sensed by the

front-facing camera, our pre-trained attribute classifiers provide a 44-dimensional attribute feature. The score is determined by comparing the extracted features with features corresponding to the enrolled user. These score values are used to continuously authenticate the mobile device user.

This paper is organized as follows. Section 2 gives the details of the proposed attribute-based authentication method. Experimental results on two publicly available mobile face video datasets are given in Section 3. Finally, Section 4 concludes the paper with a brief summary and discussion.

2. Attribute-based Authentication

In this section, we present the details of the proposed attribute-based authentication system. In particular, we describe the training data used to learn the attribute classifiers, how different classifiers are trained for each attribute and how verification is performed using the attributes.

2.1. Training Data

PubFig dataset [12] is one of the few publicly available datasets that provides facial attributes along with face images. We use this dataset to train our attribute classifiers. PubFig dataset consists of unconstrained faces collected from the Internet by using a person's name as the search query on a variety of image search engines, such as Google Images and flickr. However, there are several challenges that we have to overcome before this dataset can be effectively utilized for our application. Since the release of this dataset in 2009, many links to the images in this dataset are broken. Hence, not all the images listed in this dataset are available for downloading. As a result, we use a subset of this dataset where we could establish proper links to the images. Furthermore, the true attribute labels of the images are not provided, instead the output of their attribute classifiers are provided. As a result, we used a proper threshold to get the labels for each attribute of the available images to ensure that the classifier is certain enough about the label it is giving to the image. Finally, rather than using all 73 binary attributes in the PubFig dataset, we selected a more meaningful subset of 44 attributes in our implementation.

FaceTracer [11] is another publicly available dataset that has face images with 18 attributes. This dataset is smaller than PubFig dataset and again a greater portion of the hyperlinks to the images in this dataset are broken. Also, not all but a subset of attribute labels are provided.

2.2. Attributes Classifiers

Each attribute classifier $Cl_i \in \{Cl_1, ..., Cl_N\}$ is trained by an automatic procedure of model selection for each attribute $A_i \in \{A_1, ..., A_N\}$, where N is the total number of attributes. Automatic selection is necessary since each attribute needs a different model. Our models are indexed as follows

1 **Facial parts:** For each attribute, a set of different facial components can be more discriminative. The face components considered for training are: *eyes*, *nose*, *mouth*, *hair*, *eyes&nose*, *mouth&nose*, *eyes&nose&mouth*, *eyes&eyebrows*, and the *full* face. In total, nine different face components are considered.

2 **Features:** For different attributes, different types of features may be needed. For example, for the attribute "blond hair", features related to color can be more discriminative than features related to texture. The following features are considered in this paper: *LBP*[1], *ColorLBP*, *HoG*[4], and *ColorHoG*. ColorLBP and ColorHOG are obtained by concatenating the HoG/LBP feature of each RGB channel. In total, four types of features are extracted using the VLFeat toolbox [15].

3 **Locality of features:** In order to capture the local information, we consider different cell sizes of the HOG and the LBP features. In total, six different cell sizes, $6, 8, 12, 16, 24, 32$, are used.

We use a state-of-the-art publicly available fiducial point detection method [2] to extract the different facial components. Furthermore, the detected landmarks are also used to align the faces to a canonical coordinate system. After extracting each set of features, the Principal component analysis (PCA) is used with 99% of the energy to project each feature onto a low-dimensional subspace. An SVM with the RBF kernel is then learned on these features. This process is run exhaustively to train all possible models. For each attribute classifier, 80% of the available data is used for training the SVMs and 20% of the data is used for model selection. The face images in the test set do not overlap with those in the training set. Total number of negative and positive classes are the same for both training and testing. Finally, among all 216 SVMs, five with the best accuracies are selected.

For a given test face image F, a feature vector $[f_{a_1} ... f_{a_N}]$ is calculated by

$$f_{a_k} = \frac{\sum_{i=1}^{5} w_k^i Cl_k^i(F)}{\sum_{i=1}^{5} w_k^i}, \quad (1)$$

where $Cl_k^i(F) \rightarrow \{0, 1\}$ is the output of the ith accurate classifier for the kth attribute A_k on face image F, and w_i is the accuracy of Cl_k^i. The entire training pipeline of our method is shown in Figure 2.

2.3. Verification

We consider the continuous authentication problem as a verification problem in which given two pairs of videos

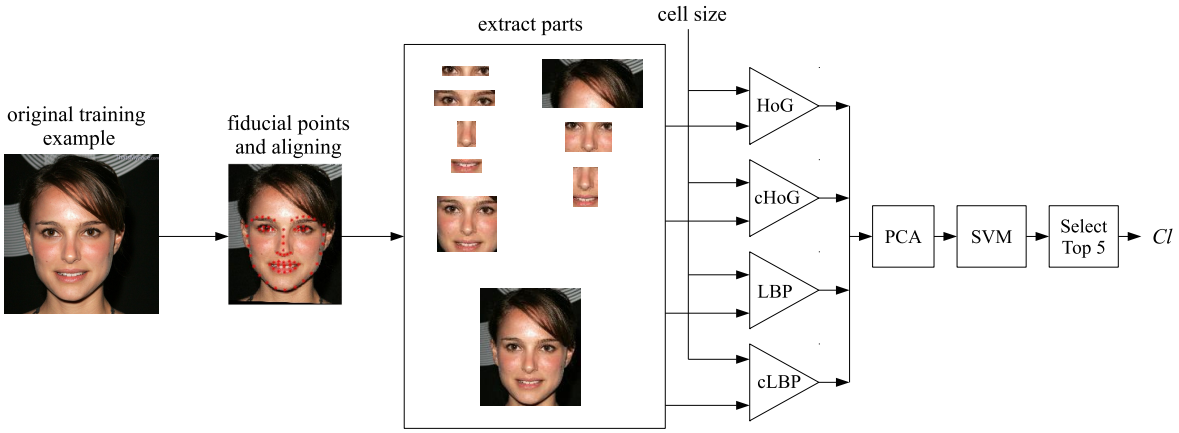


Figure 2. Training phase pipeline for each attribute classifier. Landmarks are first detected on a given face. Different facial components are then extracted from these landmarks. Then for each part, features are extracted with different cell sizes and the dimensionality of features is reduced using the PCA. Classifiers are then learned on these low-dimensional features. Finally, top five Cls are selected as our attribute classifier.

or images, we determine whether they correspond to the same person or not. The well-known receiver operating characteristic (ROC) curve, which describes the relations between false acceptance rates (FARs) and true acceptance rates (TARs), is used to evaluate the performance of verification algorithms. As the TAR increases, so does the FAR. Therefore, one would expect an ideal verification framework to have TARs all equal to 1 for any FARs. The ROC curves can be computed given a similarity matrix.

We use the proposed framework to extract the attribute vector from each image in a given video. We then simply average them to obtain a single attribute vector that represents the entire video. Then, the (i, j) entry of the similarity matrix S_{attrs} is calculated as

$$s_{i,j} = \frac{1}{\|\mathbf{e}_i - \mathbf{t}_j\|_2}, \tag{2}$$

where $\mathbf{e}_i$ is the ith attribute vector representing the gallery (or enrollment) video, and $\mathbf{t}_j$ is the jth attribute vector representing the probe video.

3. Experimental Results

We evaluate the performance of the proposed attribute-based authentication method on two publicly available mobile video datasets - MOBIO [13] and UMDAA-01 [6]. In addition to the ROC curves, Equal Error Rate (EER) is used to measure the performance of different methods. The EER is the error rate at which the probability of false acceptance rate is equal to the probability of false rejection rate. The lower the EER value, the higher the accuracy of the authentication system.

We use an LBP-based method as a baseline for comparison. In this method, each detected face is represented by the histogram of LBP features. The same aligned faces that are used for attribute feature extraction are also used to extract the LBP features. Similar to the attribute features, the LBP features from each image in a video are extracted and averaged to represent a single video. The LBP features are extracted using the VLfeat toolbox. The similarity matrix, S_{LBP}, is then built by comparing two feature vectors. This LBP-based method has been used for mobile face authentication in [13] and [10]. A third fusion score matrix, $S_{fusion} = \tilde{S}_{LBP} + \tilde{S}_{attrs}$, is calculated by z-score normalization

$$\tilde{s}_{i,j} = \frac{s_{i,j} - \bar{S}}{\sigma(S)}, \tag{3}$$

where $\bar{S}$ and $\sigma(S)$ are the mean and the standard deviation of the entries in similarity matrix S, respectively.

3.1. Results on Attribute Classifiers

In order to see how well our attribute classifiers work, Tables 1 and 2 contain the accuracies of the attribute classifiers trained using our system on the PubFig and FaceTracer datasets, respectively. As it can be seen from these tables, most of the accuracies are high.

Furthermore, in Figure 3 we show some sample outputs of our attribute classifiers. Results of the classifiers are scaled to be between -0.5 to 0.5. For the first face, eyeglasses, chubby, round jaw, Asian, male, no beard, sideburns, bangs classifiers give high scores. This clearly matches with the image shown on the left. For the second face, it is interesting to see that the Male classifier produces a negative score since the image corresponds to a fe-

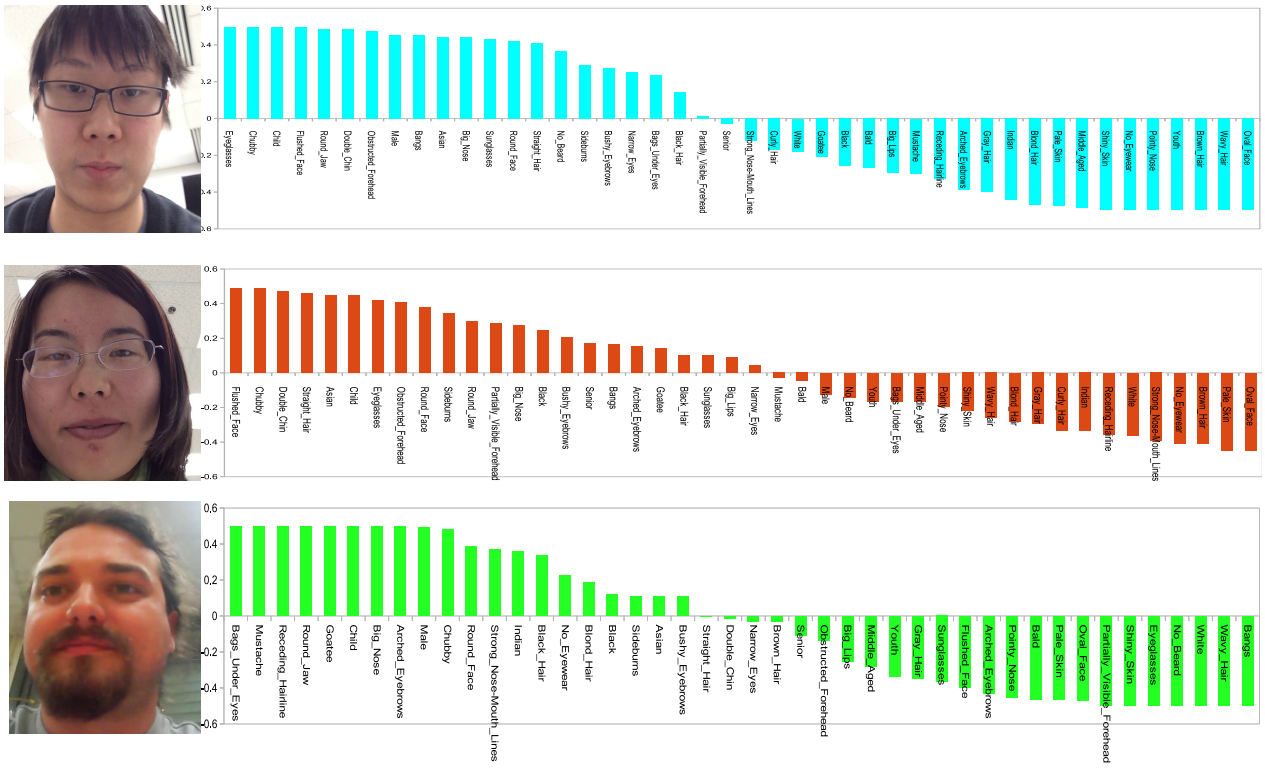


Figure 3. Illustration of our attribute classifiers on sample face images from the UMDAA-01 (first two images) and the MOBIO (last image) datasets.

Attribute	Accuracy	Attribute	Accuracy
Blond Hair	0.9089	Child	0.9538
Partially Visible Forehead	0.8645	Narrow Eyes	0.7777
Round Face	0.9156	Big Nose	0.8039
Indian	0.9714	Male	0.9451
Gray Hair	0.9091	Pointy Nose	0.816
Bags Under Eyes	0.8986	Asian	0.9225
Obstructed Forehead	0.8913	White	0.6992
Shiny Skin	0.9532	Youth	0.7299
No Eyewear	0.8875	Brown Hair	0.6725
Middle Aged	0.929	Bald	0.7909
Senior	0.8867	Wavy Hair	0.9357
Eyeglasses	0.9397	Straight Hair	0.7408
Sunglasses	0.9701	Bangs	0.9397
Mustache	0.8606	Arched Eyebrows	0.6462
Chubby	0.8815	Strong Lines	0.9308
Receding Hairline	0.8164	Pale Skin	0.793
Round Jaw	0.9357	Flushed Face	0.7819
Big Lips	0.7578	Double Chin	0.9727
No Beard	0.7766	Black Hair	0.8029
Goatee	0.9775	Curly Hair	0.8746
Black	0.7818	Bushy Eyebrows	0.836
Sideburns	0.8756	Oval Face	0.82

Table 1. Accuracies of the 44 attribute classifiers proposed in this paper on the PubFig dataset [12].

male subject. Finally, for the last face, mustache, Goatee, chubby and bags under eyes produce high positive scores

Attribute	Accuracy	Attribute	Accuracy
asian	0.8786	middle aged	0.7321
eyeglasses	0.7214	black	0.808
sunglasses	0.89	female	0.88
smiling false	0.8	senior	0.7933
no eyewear	0.7481	hair color blond	0.7875
child	0.8276	white	0.763
mustache	0.815	youth	0.692

Table 2. Accuracies of the attribute classifiers proposed in this paper on available attributes on the FaceTracer dataset [11].

which clearly match with the image shown on the left.

3.2. MOBIO Dataset Results

The MOBIO dataset [13] consists of video data taken from 152 subjects. The dataset was collected in six different sites from five different countries. In total twelve sessions were captured for each subject - six sessions for phase 1 and six sessions for phase 2. The database was recorded using two mobile devices: a NOKIA N93i mobile phone and a standard 2008 MacBook laptop computer. The laptop was only used to capture videos of part of the first session. So the first session consists of data captured with both the

laptop and the mobile phone. Figure 4 shows some samples images from the MOBIO dataset.

Figure 4. Sample images from the MOBIO dataset. One can clearly see different illumination conditions present in this dataset.

In the MOBIO protocol, for each person, the data from one session is used for enrollment and the data from the remaining sessions are used for testing. In the first set of experiments with the MOBIO dataset, we do not consider the data from the laptop session. The first mobile session is considered as the enrollment session and the data from the next 11 sessions are considered for testing. The ROC curves corresponding to this experiment are shown in Figure 5 for the entire dataset. As can be seen from this figure, our attribute-based method performs comparably to the LBP-based method. However, the best performance is achieved when the similarity matrices corresponding to the LBP and attribute features are fused. The EER values corresponding to this experiment are compared in Table 3.

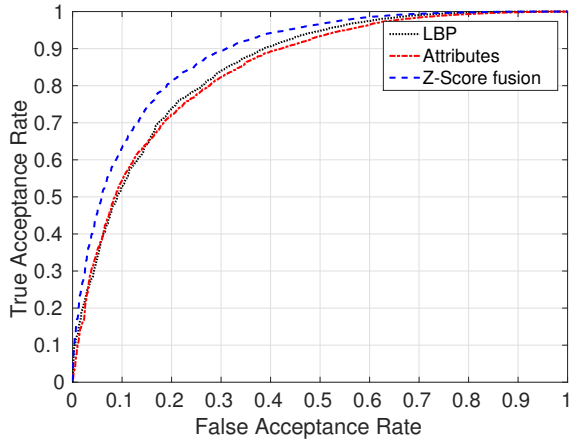


Figure 5. Performance evaluation on the MOBIO dataset.

Name	LBP	Attributes	Fusion
MOBIO_but	0.29	0.28	**0.25**
MOBIO_idiap	0.18	0.20	**0.14**
MOBIO_lia	0.31	**0.24**	0.25
MOBIO_uman	0.20	0.25	**0.18**
MOBIO_unis	0.24	0.28	**0.24**
MOBIO_uoulu	0.27	0.24	**0.23**
MOBIO_all	0.22	0.23	**0.19**

Table 3. The EER values for different methods on the MOBIO dataset.

3.2.1 Cross-device Experiments

Images captured by different cameras have different characteristics. Since the MOBIO dataset has videos that were captured using different sensors, we conduct cross-session experiments in which the data from the laptop session are considered as the enrollment data and the data from the cell phone are used as the test videos. This experiment essentially allows us to study the robustness of different algorithms with respect to different image quality. Figure 6 and Table 4 show the the ROC curves and the EER values corresponding to this experiment. As can be seen from this results, attributes are more robust to camera sensor change than LBP features. In this experiment, fusion does not necessarily improve the performance over the attributes since LBP features perform poorly.

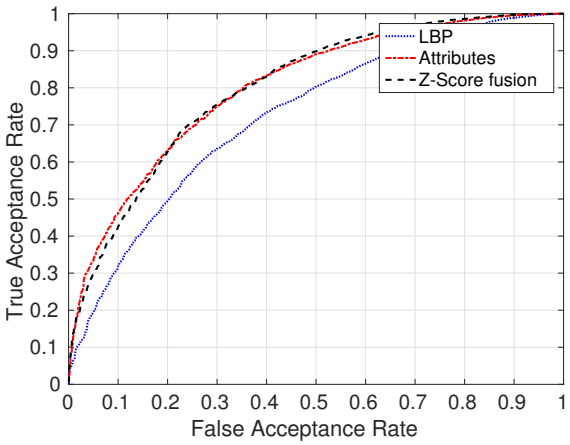


Figure 6. Cross device robustness. Laptop session videos are used for enrollment and the data from the remaining sessions are used for testing.

Enrollment	LBP	Attributes	Fusion
Laptop	0.33	**0.27**	0.27

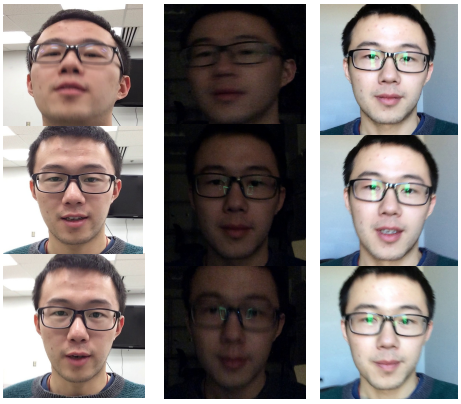
Table 4. The EER values corresponding to the cross-device experiment on the MOBIO dataset.

Name	LBP	Attributes	Fusion
UMDAA-01_1	0.2	**0.13**	0.13
UMDAA-01_2	0.32	**0.13**	0.16
UMDAA-01_3	0.23	**0.14**	0.14
UMDAA-01_all	0.34	**0.30**	0.30

Table 5. The EER values of different methods on the UMDAA-01 dataset.

3.3. UMDAA-01 Dataset Results

The UMDAA-01 dataset consists of 750 videos from 50 different individuals collected in three different sessions corresponding to three different illumination conditions. The UMDAA-01 dataset was collected using an app on an iPhone 5s. Each user performed five tasks in three sessions. The different tasks were enrollment task, document task, picture task, popup task and scrolling task. Figure 7 shows some sample images from the UMDAA-01 dataset where one can clearly see the different illumination conditions present in this dataset.



(a) (b) (c)

Figure 7. Sample images from the UMDAA-01 dataset. (a), (b) and (c) show some sample images from session 1, 2 and 3, respectively.

In the first set of experiments using this dataset, we use the data corresponding to the enrollment task as gallery and the data from the remaining tasks for testing. Figure 8 and Table 5 show the ROC curves and the EER values, respectively corresponding to this experiment. As can be seen from these results, our attribute-based method performs much better than the LBP-based authentication system. Fusion of the LBP and the attribute similarity matrices results in performance comparable to our method as the LBP features do not perform well on this dataset.

Furthermore, we conducted several session-wise experiments on this dataset. We used the enrollment data as gallery and the data from other tasks from the same session as probe. The ROC curves corresponding to these experiments are shown in Figure 9(a)-(c). It can be seen from

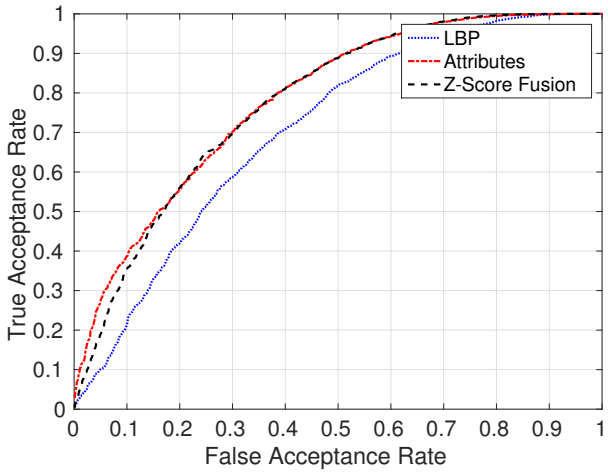


Figure 8. Performance evaluation on the UMDAA-01 dataset.

Gallery	LBP	Attributes	Fusion
Session 1	0.36	0.33	**0.32**
Session 2	0.35	0.31	**0.30**
Session 3	0.38	0.33	**0.31**

Table 6. The EER values corresponding to the cross-session experiments on the UMDAA-01 dataset.

these figures that our attribute-based method works significantly better than the LBP-based method on the same-session experiments.

Finally, similar to the cross-device experiments on the MOBIO dataset, we conducted cross-session experiments on the UMDAA-01 dataset. We used the data from the enrollment task from one session as gallery and the data from the other sessions as probe. This experiment shows the robustness of our attribute-based method to different illumination conditions. From Figure 9(d)-(f), we see that even when the illumination conditions are different, our attribute-based method is more robust than the LBP feature-based method. From Figure 9(d)-(f) and Table 6 we see that in all cases, attributes performed better than LBP and the fusion of both gives the best results.

3.4. Runtime

The prediction and matching algorithm were tested on one core of Intel Xeon(R) CPU E5620 clocked at 2.4GHz with 12GB of RAM. Per video frame, the algorithm took 3.2s on average to extract face components and 0.09s to extract attribute features. The attributes prediction part took 5MB of memory on average which is reasonable for mobile device. The fast runtime and low memory usage is due to the order of algorithm being linear with image size. Without any improvement on face component extraction, on a Nexus 5 device which has four 2.3GHz CPU cores of Qualcomm MSM8974 Snapdragon 800, one can reliably extract attributes every 4 to 5 seconds with a very low memory us-

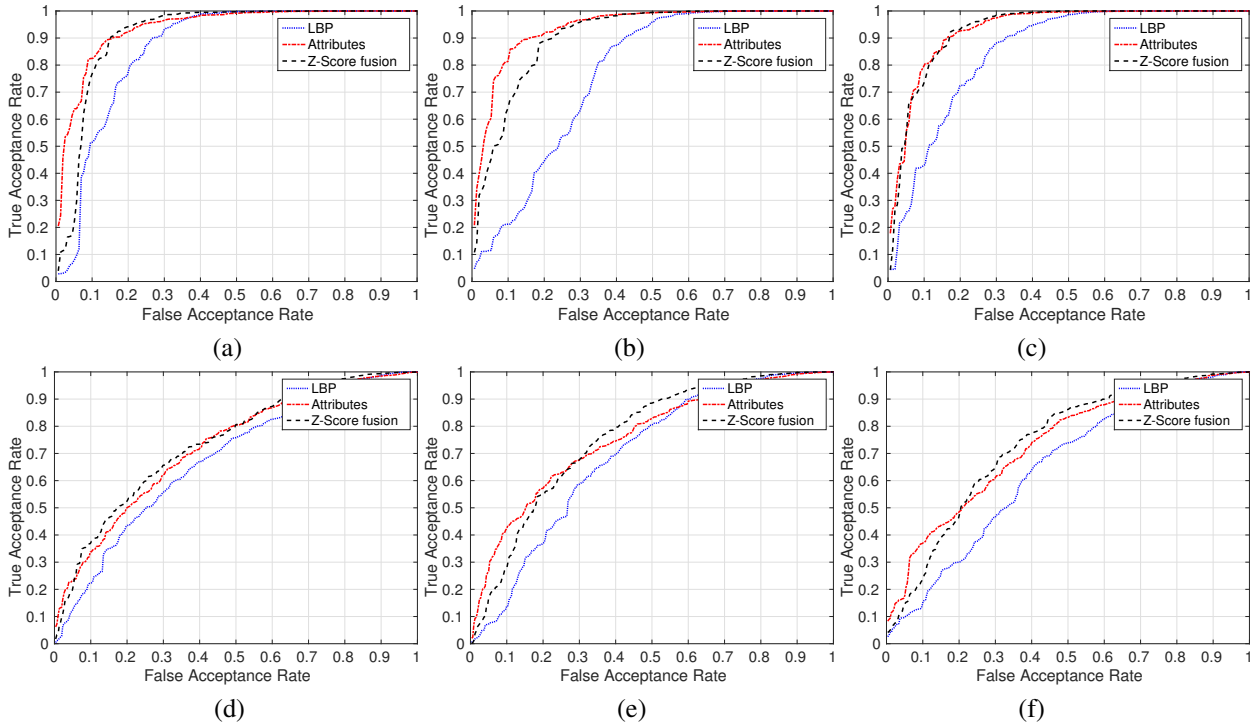


Figure 9. Session-wise performance evaluations on the UMDAA-01 dataset. (a) Gallery and probe data from session 1. (b) Gallery and probe data from session 2. (c) Gallery and probe data from session 3. (a) Gallery data from session 1 and probe data from sessions 2 and 3. (e) Gallery data from session 2 and probe data from sessions 1 and 3. (f) Gallery data from session 3 and probe data from sessions 2 and 1.

age. However, the face component extraction part can be done much faster if the detectors are trained on the face parts using methods such as [10]. As a result, one can detect facial parts in less than 0.5s [10] and hence the algorithm can run in real time.

4. Conclusion and Future Directions

We presented a novel continuous face-based authentication method using facial attributes for mobile devices. We trained 44 binary attribute classifiers and showed their effectiveness as feature vectors for active authentication with extensive experiments. We showed that attribute-based scores alone can improve the verification results. Furthermore, in situations where the representation-based features are also reliable, verification results can be further improved by fusing attribute-based scores.

In the future, we are planning on exploring how attributes can be detected more reliably from mobile images with sparse-representation-based methods and also how we can effectively adapt the attribute classifiers to changing attributes of the user, like aging or facial hair change by exploiting classifiers with feedback.

Acknowledgement

This work was supported by cooperative agreement FA8750-13-2-0279 from DARPA.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3444–3451. IEEE, 2013.

[3] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain. Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data. In *International Conference on Biometrics*, 2015.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[5] M. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 306–311, Oct 2010.

[6] M. E. Fathy, V. M. Patel, and R. Chellappa. Face-based active authentication on mobile devices. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015.

[7] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen. Continuous mobile authentication using touchscreen gestures. In *IEEE Conference on Technologies for Homeland Security*, pages 451–456, Nov 2012.

[8] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, Jan 2013.

[9] L. Fridman, S. Weber, R. Greenstadt, and M. Kam. Active authentication on mobile devices via stylometry, gps location, web browsing behavior, and application usage patterns. *IEEE Systems Journal*, 2015.

[10] A. Hadid, J. Heikkila, O. Silven, and M. Pietikainen. Face and eye detection for person authentication in mobile phones. In *ACM/IEEE International Conference on Distributed Smart Cameras*, pages 101–108, Sept 2007.

[11] N. Kumar, P. N. Belhumeur, and S. K. Nayar. FaceTracer: A Search Engine for Large Collections of Images with Faces. In *European Conference on Computer Vision (ECCV)*, pages 340–353, Oct 2008.

[12] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009.

[13] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J.-F. Bonastre, P. Tresadern, and T. Cootes. Bi-modal person recognition on a mobile phone: using mobile phone data. In *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, July 2012.

[14] A. Primo, V. Phoha, R. Kumar, and A. Serwadda. Context-aware active authentication using smartphone accelerometer measurements. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 98–105, June 2014.

[15] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[16] H. Zhang, V. M. Patel, and R. Chellappa. Robust multimodal recognition via multitask multivariate low-rank representations. In *IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2015.

[17] H. Zhang, V. M. Patel, M. E. Fathy, and R. Chellappa. Touch gesture-based active user authentication using dictionaries. In *IEEE Winter conference on Applications of Computer Vision*. IEEE, 2015.

[18] H. Zhang, V. M. Patel, S. Shekhar, and R. Chellappa. Domain adaptive sparse representation-based classification. In *IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2015.