California State University, Fullerton

Multimodal Biometric Authentication using Feature Fusion on MoBio mobile dataset

The team:

Dr. Mikhail Gofman, Yash Bhambhani, Oscar Olazabal, Nicole Traboulsi, Rushabh Shah

Yash Bhambhani

CPSC 499: Independent Study

Dr Mikhail Gofman

15 May 2020

1. Introduction

A biometric authentication is a promising approach to security in the Internet of Things (IoT). It relieves users from having to create and remember strong passwords, and largely eliminates the security threats resulting from using the same password on multiple accounts/devices. However, implementing such authentication on IoT devices is challenging. Ever since the creation of such authentication services, numerous cases of hackers bypassing the device have also been reported. For example, Apple iPhone 5s[1] from 2013 consisted of a fingerprint recognition biometric authentication system. Few days after the release, hackers were able to bypass the authentication by copying the authenticated fingerprint on the scanner by using gorilla glue. Samsung Galaxy S6[1] faced a similar fate, by hackers using wax this time around. Biometric authentication services as sophisticated as Apple iPhone X's Face ID has also been prone to unauthenticated bypass. Face ID was bypassed by using a 3-dimensional mask or 2-dimensional infrared images.[2] Furthermore, there have been reports of children authenticating their parents' devices since their faces look similar. Even putting these security risks aside, false rejections are common in uncontrolled conditions.[2] Conditions, where lighting isn't optimal for the biometric authentication to perform its matching measure or a dirty camera lens/fingerprint scanner[3], can throw off the authentication service. These problems can more or less be blamed on the system's reliance on one single type of biometric. Therefore, these limitations and safety measures can be successfully overcome by creating and implementing a multimodal biometric authentication check. In this paper, by following a similar pipeline of Gofman et. al.'s[4] previous work, we create a multimodal biometric authentication system using feature-level fusion to combine features from different biometric types. This project scope utilizes Idiap Research

Institute's MOBIO dataset[5], which consists of audio and video samples from one hundred and fifty-two subjects.

2. ***Preliminaries***

This paper provides a brief overview of the challenges and opportunities of utilizing the MOBIO dataset, implementing a multimodal biometric authentication system, Discriminant Correlation Analysis[6] for integrating multiple features together, and our methodology along with the classification algorithms used in the project scope to train and test the accuracy of the model.

2.1 Utilizing the MOBIO dataset

In order to train any classification models, an organized and vast dataset is essential. This requirement was fulfilled by McCool et. al.'s dataset, which was created in 2012. The dataset has a female-male ratio of nearly 1:2 and was collected from five different countries. The voice and face samples were captured on a Nokia N93i and 2008 Apple MacBook. The MOBIO dataset provides text lists with sample paths to already split training, development, and evaluation datasets. Each of these datasets consists of subject identification numbers and paths to their corresponding face and voice samples. This project scope reads in the file paths for training and evaluation samples and stores them in a data structure for further preprocessing. Then, four subjects and four of their bi-modal samples are randomly selected for the training phase equaling 16 face and 16 voice samples. Similarly, sixteen subjects with four of their corresponding face and voice samples are selected for the evaluation phase equaling 48 face and 48 voice samples. Finally, a single subject is selected as the 'genuine' sample, and the rest are deemed as the

‘imposter’ sample. This genuine and imposter split on the data doubles the samples in total, and total number of face and voice samples in training phase equals 32, and in the evaluation phase equals 96.

2.2 Implementing a multimodal biometric authentication system

Integrating identifying information from multiple biometric modalities is known to improve recognition accuracy. This is because different modalities, such as face and voice, provide independent sources of discriminating information that can be used for identification. Multiple modalities can also be more difficult for an attacker to bypass than a single one. Finally, high-quality identifying data from one modality can be used to compensate for low-quality data in other modalities to increase authentication accuracy.

To implement such a system, we needed to find the most identification retaining method of integrating features together from different modalities. We decided on utilizing the feature-level fusion, which combines features from two modalities into a single set. After fusing the features together, we classify the combined set as belonging to a legitimate user or an imposter one. Features, in our level of fusion, preserves more identifying information than other levels of fusion out there. Specifically, we combine the Histogram Oriented Gradient (HOG)[7] features and Local Binary Pattern (LBP)[8] features from the face with Mel Cepstral Coefficient (MFCC)[9] features and Linear Predictive Codes (LPC)[10] features from the voice. The biggest challenge during the fusion phase of these features was that these features are measure differently and hence cannot be combined directly. Second, combining these features can result in the curse-of-dimensionality problem where the combined set of features will be excessively

large. These large fused-features increases the computational power and the time required to train and evaluate our classifiers. To address these problems, we use Discriminant Correlation Analysis (DCA) for feature-level fusion.

2.2.1 Biometric features

HOG features are among the most commonly used features in face recognition. Combining them together is also associated with improved recognition accuracy. HOG features are derived by partitioning an image into square cells, computing the histogram of gradient orientation in each cell, and then normalizing the result using a block-wise pattern. The process yields a dimensionless real number quantity that is the identifying feature.[7]

LBP features are another popular feature for face recognition and perform robustly in texture classification. The LBP algorithm splits the image into cells. For each pixel in each cell, neighborhoods are created. Every neighborhood consists of a center pixel and its surrounding pixels. The center pixel is then compared to its neighbors. If the neighbor is less than the center pixel, a one is written for that neighborhood, else a zero. This combination of 1s and 0s are then formed into a binary number which is converted to a decimal number. A histogram of the frequency of each seminal number with each cell is then calculated and saved.[8]

MFCC features are widely used in voice recognition. Each MFCC comprises of an audio signal. The feature vectors are computed by partitioning the signal and calculating filter bank coefficients with the use of a Fourier Transform. A discrete cosine transform is then applied in order to decorrelate the filter blanks.[9]

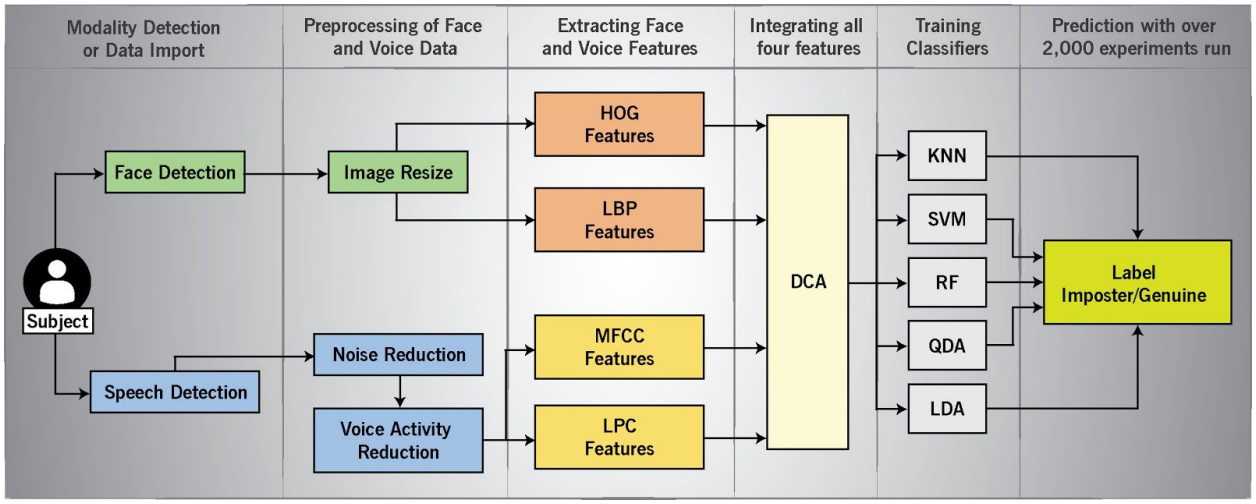
LPC features are also commonly used for medium to low bit rate voice samples. The LPC calculates a power spectrum of the signal by applying a formant analysis. An LPC feature is

quantized by the smaller number of bits compared to the original signal. A parametric model is computed based on the least mean squared error theory, this technique known as the linear prediction. By this method, the speech signal is approximated as a linear combination of its previous samples.[10]

2.3 Discriminant Correlation Analysis

DCA was originally proposed by Haghighat et. al.[6] as one of the first feature-level fusion algorithms that considers class structures. The algorithm combines two or more feature vectors into a single feature that incorporates the independent variation in each feature that is most predictive of the final class. This is accomplished by finding the optimal projection of features into a space that maximizes and standardizes between-class variance. Because DCA involves simple matrix operations on the individual features, it has a low computational complexity, which makes it suitable for almost every device.

2.4 Methodology



The figure above gives an overview of our feature-level fusion approach. First, a face image and voice recording are read in from the data structure for one subject. Then HOG and

LBP features are extracted after the face image is resized down to 150 x 150 pixels. We then perform noise reduction and voice activity detection on the voice sample and then use the detected, de-noised voice to compute MFCC and LPC features. The four feature sets are then fused into a combined feature set using DCA. DCA normalizes two sets of features at a time, combines them, and performs dimensional reduction on the fused set. The fused sets from the pairwise fusions are then concatenated. Finally, the fused features are fed into KNN, SVM, RF, QDA, and LDA classifiers that labels the current subject as either an imposter or genuine. Next, we discuss the details of the process.

2.4.1 Face feature extraction

First, we obtain *n* training images from the dataset. Each image is then resized to 150 x 150 pixels. 1 x 1296 HOG and 1 x 10000 LBP feature vectors are then extracted from each face sample to form two feature matrices. We use the OpenCV and Sci-Kit image processing libraries in python to perform face feature extraction. In particular, we utilize the cv2.HOGDescriptor module to compute the HOG feature and skimage.feature.linear_binary_pattern module to compute the LBP feature.

2.4.2 Voice feature extraction

Second, *n* voice samples are read in from the data structure. Then, we perform noise reduction and voice activity detection to de-noise and crop the voice sample where the speech was detected. Using this preprocessed sample, we compute a 1 x 13500 MFCC and 1 x 12960 LPC feature vectors. We utilize Librosa speech analyzing library in python to compute the voice features.

2.4.3 Feature-level fusion

DCA is applied to the HOG, LBP, MFCC, and LPC matrices. We fused pairs of features at a time and concatenated their outputs. We had a total of $\binom{4}{2}$ fusions, which resulted in four *n* x 2 matrices, where the first column is a projection of the first feature in the combination set and the second column is a projection of the second feature in the combination set. The four *n* x 2 matrices are then concatenated to create a fused matrix.

2.4.4 Classification

The fused matrix is used to train the KNN, SVM, RF, QDA, and LDA classifiers. Fused test data will be then fed into each of the classifiers to be labeled as genuine or imposter.

3. ***Experimental Results***

This experiment tested how well multimodal biometric authentication performed against unimodal authentication. We will be comparing our results to the results achieved in Gofman et. al.[11] Within the time frame of this paper, we were able to run 2,000 experiments. We obtained our result by averaging the equal error rates (EER) across the 2,000 experiments. The following table shows EERS for both unimodal and multimodal experiments from our best classifier in each modality. Additionally, it contains the result from Gofman et. al., which was the previous iteration of this same project.

Approach	**Fused (DCA)**	**Face (HOG)**	**Face (LBP)**	**Voice (MFCC)**	**Voice (LPC)**
Our approach	13.00%	20.52%	18.50%	39.52%	49.48%
Previous iteration	8.04%	16.91%	14.05%	43.76%	N/A

While EER of our fused DCA set, HOG set, and LBP set all went down a marginally small percent, our MFCC set did hike up 4.24%. The reasons for the negative progression include: inclusion of LPC features in this dataset and switching from a relatively small dataset to a large and different dataset.

4. References

[1] M. Schwartz, "Apple iPhone 6 touch ID hacked," Bank Info Security, September 23, 2014. [Online]. Available: https://www.bankinfosecurity.com/apple-iphone-6-touchid-hacked-a- 7348.

[2] A. Greenberg, "Hackers just broke the iPhone X's face ID using a 3D-printed mask," Wired, November 13, 2017. [Online]. Available: https://www.wired.co.uk/article/hackers-trick-apple-iphone-x-face-id-3d-mask-security.

[3] K. Spirina, "Biometric authentication: The future of IoT security solutions," IoT Evolution World, July 5, 2018. [Online]. Available: https://www.iotevolutionworld.com/iot/articles/438690-biometric-authentication-future-iot-security-solutions.htm.

[4] M. Gofman, N. Sandico, S. Mitra, E. Suo, S. Muhi, and T. Vu, "Multimodal biometrics via discriminant correlation analysis on mobile devices," Int'l Conf. Security and Management (SAM), pp. 174-181, 2018.

[5] C. McCool, S. Marcel, A. Hadid, M. Pietikäinen, P. Matějka, J. Černocký, N. Poh, J. Kittler, A. Larcher, C. Lévy, D. Matrouf, J. Bonastre, P. Tresadern, and T. Cootes, "Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data", IEEE ICME Workshop on Hot Topics in Mobile Multimedia, 2012.

[6] Haghighat, M., Abdel-Mottaleb, M. and Alhalabi, W. (2016). Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition. IEEE Transactions on Information Forensics and Security, 11(9), pp.1984-1996.

[7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human Detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, San Diego, CA, USA, 2005, pp. 886-893

[8] T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no.7, pp. 971-987, 2002.

[9] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[10] N Dave, “Feature extraction methods LPC, PLP, and MFCC in speech recognition,” *International Journal for Advanced Research in Engineering and Technology*, vol 1, pp 145-167, 2013.

[11] M. Gofman, S. Mitra, O. Olazabal, Y. Bai, Y. Choi, N. Sandico, and K. Pham, “Multimodal Biometrics for Enhanced IoT Security.” [Online] Available: https://www.dropbox.com/s/u2no41fvdzekj3u/OlazabalIoTMultimodal.pdf