Facial Movement Based Person Authentication

Pengqing Xie
Department of Electrical and Computer Engineering
Iowa State University
Ames, IA 50010
carterp@iastate.edu

Yang Liu
Department of Electrical and Computer Engineering
Iowa State University
Ames, IA 50010
yangl@iastate.edu

Yong Guan
Department of Electrical and Computer Engineering
Iowa State University
Ames, IA 50010
guan@iastate.edu

ABSTRACT

In this position paper, we present an authentication method using spatial temporal facial movements. Facial movements are defined by a facial coding system and are obtained by making a facial expression in front of a 3D camera. Facial muscle movements are impossible to be replicated, thus our method is invulnerable to spoofing attacks. We briefly discuss our current work and present preliminary result.

1. INTRODUCTION

Security system often has to make authentication decisions to grant access to legitimate users. Currently, the most common method is to use a string of characters and numbers, which are known as the password. However, passwords can be transferred and stolen, thus are vulnerable to dictionary attacks and social engineering. Additionally, passwords require memorization which makes it difficult to manage. The trend of solving these problems is biometric authentication which utilizes physical appearances of certain body parts. Nevertheless, they suffer from spoofing attacks. For instance, a face authentication system can be fooled by manipulating a digital image of a legitimate user's face. Since most face recognition algorithms only detect and track static features such as points around the eyes, mouth, and nose, they cannot distinguish between real and counterfeited faces [1].

We propose a person authentication method based on facial movements. Given that each user makes a facial expression in a slightly different way, it would be unrealistic for imposters to break into the system by solely mimicking the user's inherent facial behaviors. Even if facial movements are captured, it would be difficult to replicate the moving process as in the spoofing attack. Apart from security concerns, our approach also benefits facial recognition: by utilizing depth data, it increases accuracy, and is invariant to changes in illumination, scale, translation and small rotations by a process of normalization.

Facial movements are defined by the Facial Action Coding System (FACS) which categorizes facial behaviors by 46 action units (AUs), each of which is anatomically related to a specific set of facial muscles. For instance, the lip stretcher (AU #20) is based on risorius and platysma [2]. Each AU typically lies within the range of -1 to +1, e.g. for the lip stretcher, -1 is interpreted as fully rounded, 0 represents neutral, whereas +1 means fully stretched (Fig.1). Thus the problem of face recognition is reduced to matching two time series – model and sample – for each AU acquired in a time interval.

2. Related Work

Numerous researches have been conducted on video-based face recognitions. In the work of Tistarelli, Bicego, and Grosso, physiological and behavioral cues for face recognition were derived from neural activation and infant behavioral studies. It discussed face recognition from a human's point of view - how does a human recognize a face in social situations. Functional magnetic resonance imaging (fMRI) studies revealed that visual tasks play a more influential role during face analysis and recognition, where not only facial features are extracted, but a coherent neural activity is stored in brain areas. This implies that a "dynamic signature" can be used for authentication [3]. For implementation, a dynamic face model is formed by multiple Hidden Markov Models (HHMs) where each state represents a potential facial expression. It is worth noting that the video sequence is unrolled before clustering since each face image is labeled independently based on a discrete number of facial expressions. Conversely to their approach, we want to model motion behavior in a continuous manner such that the "true" facial moving pattern is revealed. Similarly, most of the researches in the field took a HMMs approach [4] [5].

From the security point of view, the problem of playback attack can be prevented by detecting face liveness. Kollreider, Fronthaler, and Bigun used optical flow for analyzing the trajectories of certain parts of a live face: it considered movements of eyes, eyebrows, nose, mouth, and lip [6]. Bao, H. Li; N. Li, Jiang employed eye closity in their approach to describe the blinking of eyes for liveness detection [7]. Each technique has its own benefit and most of them could achieve an effective detection rate. However, since liveness detection comes naturally with our face recognition algorithm, we eliminate the need to make additional efforts.

The closest research to our approach is the one conducted by Biuk and Loncaric where a pattern trajectory in the eigenspace was built from a sequence of face images rotating from -90° to +90°. The prototype trajectory was later compared against incoming samples using a distance measurement of two time series [8]. Nevertheless, the measurement was accumulated in Euclidean, and thus would degrade rapidly with noise and sensitive to time variations [9]. Our approach utilizes the Longest Common Subsequence (LCSS) method to accommodate inputs that are noisy and time-shifted.

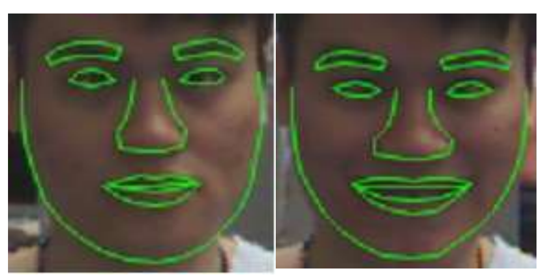

Figure 1. An illustration of lip stretcher. On the left is a neutral face (0), and on the right is a smiling face (+1).

3. Methodology

3.1 Enrollment

During the enrollment phase, the user is asked to follow instructions in a strict manner. The registration process is carried out in a relatively dark environment to ensure that the intensity of infrared light is not being affected. The user needs to remove any objects on their face, such as glasses and hats, to eliminate occlusions and reflection. The user is guided to move accordingly to fit the entire face in a rectangle, which is chosen to guarantee an optimal accuracy for the depth measurement. Finally, the user is instructed to perform one of the following tasks repeatedly: 1) pronouncing words; 2) imitating exemplar photos; 3) watching hilarious videos, all of which are meant to reveal facial movements either artificially (1 and 2) or naturally (3). In the subsequent steps, facial action units are computed from each video frame resulting in a number of temporal trajectories. Based on best average similarities against the other trajectories, one of these trajectories is determined to represent the identity model to guarantee that it is the nearest neighbor to all given trajectories. Additionally, the time and space matching thresholds used for LCSS similarity measurement are adaptively computed from enrollment trajectories and are stored along with the model.

3.2 Validation

When it comes to making an authentication decision, the user is asked to perform one of the data gathering tasks as described in section 3.1. An identity model is constructed and compared with existing legitimate models. LCSS is used as the comparison method with the similarity $S_{A,B}$ between the two models A and B defined as:

$$S_{A,B} = \frac{\mathrm{LCSS(A, B)}}{\min(length(A), length(B))}$$

4. Preliminary Experiment

Five subjects were invited to participate in our experiment. Each subject was asked to pronounce the word "WE" 30 times where each pronunciation was carried out within 4 seconds. A subset of the samples was used for building the model whereas the remaining ones were retained for validation.

The lip stretcher (AU #20) trajectories (Fig. 2) were extracted and analyzed using LCSS. The result showed a similarity higher than 70% for trajectories acquired from the same subject and as low as 10% if were taken from different subjects. Subsequently, another five AUs were extracted and fused together to form a combined similarity result. The similarities are being fused using the weighted product model to exploit priori knowledge such as what kind of facial expression was performed to generate the AUs. Since different AUs may respond differently to various expressions, we can assign a larger gain ratio to certain AUs. In this case, since the AUs are acquired by pronouncing the word "WE" (similar to smiling), it made sense to give more weights to the lip stretcher.

Since the facial expression in this experiment was carried out artificially, it was believed that up slope (neutral to maximal) facial movements might be similar for all subjects whereas down slope (maximal to neutral) ought to be distinctive. We would need to find it out in our future work.

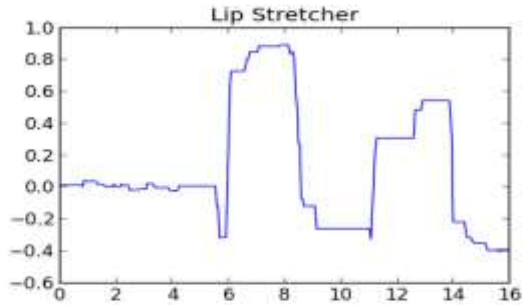

Figure 2. Lip stretcher produced by pronouncing the word "WE" for two times. Notice that the above two patterns were similar with the presence of noise and time variance.

5. REFERENCES

[1] Biggio, B.; Akhtar, Z.; Fumera, G.; Marcialis, G.L.; Roli, F., "Security evaluation of biometric authentication systems under real spoofing attacks," *Biometrics, IET* , vol.1, no.1, pp.11,24, March 2012

[2] P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

[3] Tistarelli, M.; Bicego, M.; Grosso, E., "Recognizing People's Faces: from Human to Machine Vision," Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on , vol., no., pp.1,7, 5-8 Dec. 2006

[4] Tsalakanidou, F.; Malassiotis, S.; Strintzis, M.G., "Face localization and authentication using color and depth images," Image Processing, IEEE Transactions on , vol.14, no.2, pp.152,168, Feb. 2005

[5] Xiaoming Liu; Tsuhan Chen, "Video-based face recognition using adaptive hidden Markov models," Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on , vol.1, no., pp.I-340,I-345 vol.1, 18-20 June 2003

[6] Kollreider, K.; Fronthaler, H.; Bigun, J., "Evaluating liveness by face images and the structure tensor," Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on , vol., no., pp.75,80, 17-18 Oct. 2005

[7] Wei Bao; Hong Li; Nan Li; Wei Jiang, "A liveness detection method for face recognition based on optical flow field," Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on , vol., no., pp.233,236, 11-12 April 2009

[8] Biuk, Z.; Loncaric, S., "Face recognition from multi-pose image sequence," Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on , vol., no., pp.319,324, 2001

[9] Michail Vlachos, Marios Hadjieleftheriou, Dimitrios Gunopulos, and Eamonn Keogh. 2003. Indexing multi-dimensional time-series with support for multiple distance measures. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '03). ACM, New York, NY, USA, 216-2