Time Series Matching for Biometric Visual Passwords

Kaustubha Mendhurwar
Concordia University
Montreal, Quebec

Sudhir Mudur
Concordia University
Montreal, Quebec

Tiberiu Popa
Concordia University
Montreal, Quebec

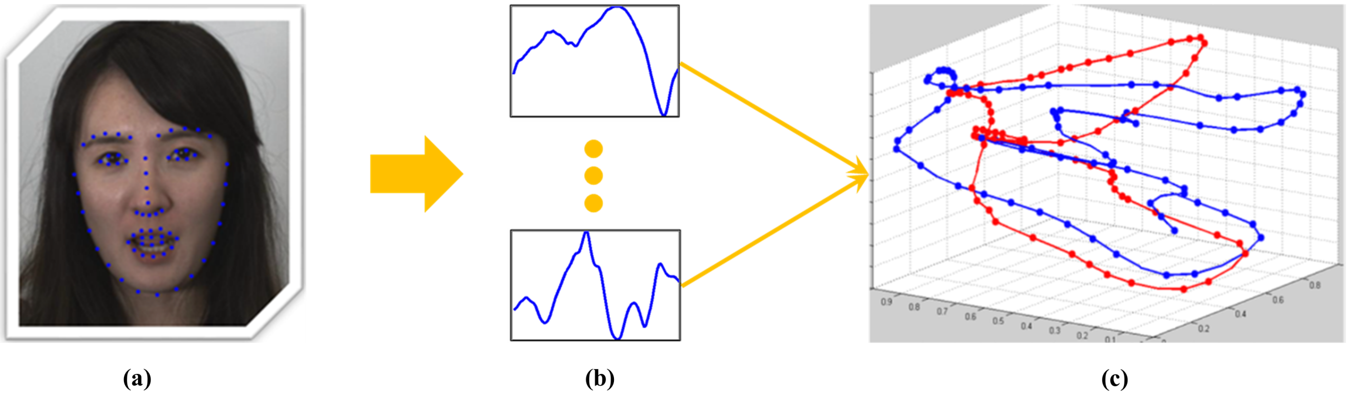

Figure 1: Biometric Visual Password Pipeline: (a) Face tracking points extracted from video streams using [Saragih et al. 2009]; only points around mouth are used for matching. (b) The salient degrees of freedom are extracted from this temporal data using [Mendhurwar et al. 2017]. (c) The final time dependent signal is matched using TSC [Mendhurwar et al. 2017].

ABSTRACT

User authentication through silent utterance of a secret phrase, a biometric visual password, has been previously attempted mainly using image based features extracted from video. Using state of the art face tracking, this problem can be framed as a high dimensional time series matching problem covering the motion of a select set of lip points. One major advantage is the small amount of training data needed. We deploy the time and shape correspondence (TSC) matching algorithm given its superior performance when dealing with multidimensional signals with shape and in the presence of noise. We report the results of a user study with 22 participants uttering the password ”siggraph rocks”. This data base along with other human action data bases we created for gait and gestures are made publicly available for comparison studies by other researchers.

CCS CONCEPTS

•**Computing methodologies** →**Computer graphics;**

KEYWORDS

lip motion tracking, TSC algorithm, biometric visual password

ACM Reference format:
Kaustubha Mendhurwar, Sudhir Mudur, and Tiberiu Popa. 2017. Time Series Matching for Biometric Visual Passwords. In *Proceedings of SIGGRAPH ’17 Posters, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
DOI: 10.1145/3102163.3102239

SIGGRAPH ’17 Posters, Los Angeles, CA, USA

DOI: 10.1145/3102163.3102239

1 INTRODUCTION AND RELATED WORK

Automatic lip reading from video footage is an important problem [Liu and Cheung 2014; Morade and Patnaik 2014; Wu and Ruan 2014] with many practical applications such as biometric visual passwords [Hassanat 2014]. A user is visually authenticated by uttering silently a secret word or sentence that she previously stored in front of a camera. Silent password authentication is becoming more popular over voice based authentication methods with several key advantages [Liu and Cheung 2014]. First, it is independent of the ambient noise, which is a big challenge for voice based authentication methods and second it does not reveal the password to a third party listener.

This problem of biometric visual password is traditionally tackled using image feature extraction and classification using variations of standard machine learning techniques such as Neural Networks (NN) [Lin et al. 1999], Gaussian Mixture Model (GMM) [Shafait et al. 2006; Wark et al. 1998] and Hidden Markov Model (HMM) [Luettin et al. 1996; Mok et al. 2004]. While these methods can provide reasonably accurate results, one of their major drawbacks is that they usually require a relatively large set of training data often making them impractical for real world application.

In this work we propose a re-framing of the problem of biometric visual passwords as a time series matching problem. We create a prototype based on Time Series Matching (TSC) algorithm [Mendhurwar et al. 2017]. We show using a user study that this method can achieve the same high accuracy as the above state of the art methods while requiring rather small training data. This initial investigation shows significant promise and we propose several novel research directions for future work.

2 PROPOSED METHOD

Using state of the art face tracking [Saragih et al. 2009] we extract in each video frame key points from the face, cf Figure 1a). This is encoded in the format of high dimensional time series data. Biometric password authentication is now re-framed as matching this data to a previously stored reference signal. In this work we are primarily interested in the matching aspect of the problem, specifically high accuracy matching, enabling high fidelity in user authentication.

The authentication process to be employed is similar to other password authentication techniques. The person chooses a secret phrase - a few words. In the password setting stage, the person faces the video camera and utters this word a number of times, say five or so, and the camera captures this. The face tracker processes each utterance separately and then the tracks for a select set of points around the lips and mouth region are gathered. These points are collectively used to form a multi-dimensional time series signal. Around five to seven signals form our reference for this person. For authentication, the person utters this secret phrase one or more times. The system extracts the tracks for the same collection of points and matches the resulting signal(s) with the reference signals. As already mentioned earlier, high accuracy matching is important, and to avoid user frustration it should have very high recall, ideally with zero false positives. The choice of the matching algorithm is therefore vital. Since facial movements are governed by the underlying face structure of the person, there is a certain characterizing shape to the movement of points on the face. It has been shown that the time and shape correspondence (TSC) time series matching algorithm [Mendhurwar et al. 2017] works extremely well for multi-dimensional signals in which the individual degrees of freedom (DOFs) have shape and even when the signal has some noise. In our experiment we compare the results of matching obtained using TSC and DTW, the latter being the most widely researched time series matching technique.

3 USER STUDY EXPERIMENT

With no public database available suitable for this study, we created a database of 22 people uttering 10 times the phrase *siggraph rocks*. The participants ranged from 22 to 68 years and from varied regions, North America, South America, Europe, India and China. Some of them we randomly choose for reference and the rest for testing. Audio was not recorded. Since the passwords are typically short taking less than a second to utter, we use high temporal resolution at around 300 frames per second at 720*p*. We used the face tracker from [Saragih et al. 2009] that tracks 66 points around the face. We only considered the subset of 16 points around the mouth. We considered as DOFs vectors between pair of markers in order to maintain translation invariance.

4 COMPARISONS AND RESULTS.

We used the entire training database consisting of 5 instances of the utterance. We selected 1000 random testing signals to match. We repeated this experiment 10 times and averaged the results. Using the TSC algorithm [Mendhurwar et al. 2017] we obtained an average accuracy of 95.5%. We compared our result with Dynamic Time Warping (DTW), a popular time series matching algorithm. DTW yielded an average accuracy of 91%. Our accuracy is similar to other state of the art methods for this application, but the main advantage of using this approach is that it has very small requirement in the size of the training data.

5 CONCLUSION AND FUTURE WORK

We have recast the problem of lip motion based biometric visual password authentication into that of time series matching using a state of the art face tracker. Compared to previous works, which rely on their own feature definition and extraction techniques, our method can benefit from all advances in face tracking, which are bound to take place given its importance and wide applicability. We show how the lip motion obtained through face tracking can be transformed into a multi-dimensional format and how a time series matching algorithm can be effectively used for visual password authentication. We had to create our own data base for a user study. This data base along with various other databases we created including an extensive human gait and gesture data base are made publicly available for comparison studies. For the immediate future, we would like to conduct a bigger user study with much larger number and wider demographic of participants, and varying face capture conditions.

For the long term, our initial experiments suggest that this approach can be generalized to more complex problems such as lip-reading. Similarly to speech recognition where time series matching algorithms such as Dynamic Time Warping (DTW) have been extensively used, lip-reading can also be cast as a time series matching problem. Furthermore, new generation time series algorithm such as TSC are proven to consistently outperform DTW on signals obtained from human motion because they inherently embed significant shape information that can be leveraged to improve the accuracy of the matching. [Mendhurwar et al. 2017].

REFERENCES

Ahmad Basheer Hassanat. 2014. Visual Passwords Using Automatic Lip Reading. *arXiv preprint arXiv:1409.0924* (2014).

Chin-Teng Lin, Hsi-Wen Nein, and Wen-Chieh Lin. 1999. A space-time delay neural network for motion recognition and its application to lipreading. *International Journal of Neural Systems* 9, 04 (1999), 311–334.

Xin Liu and Yiu-ming Cheung. 2014. Learning multi-boosted HMMs for lip-password based speaker verification. *IEEE Transactions on Information Forensics and Security* 9, 2 (2014), 233–246.

Juergen Luettin, Neil A Thacker, and Steve W Beet. 1996. Speaker identification by lipreading. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, Vol. 1. IEEE, 62–65.

K. Mendhurwar, Q. Gu, S. Mudur, and T. Popa. 2017. The Discriminative Power of Shape An Empirical Study in Time Series Matching. *IEEE Transactions on Visualization and Computer Graphics* PP, 99 (2017), 1–1.

Lin Leung Mok, Wing Hong Lau, Shu Hung Leung, Shi-Lin Wang, and Hong Yan. 2004. Lip features selection with application to person authentication. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, Vol. 3. IEEE, iii–397.

Sunil S Morade and Suprava Patnaik. 2014. Lip Reading by Using 3-D Discrete Wavelet Transform with Dmey Wavelet. *International Journal of Image Processing (IJIP)* 8, 5 (2014), 384.

Jason M Saragih, Simon Lucey, and Jeffrey F Cohn. 2009. Face alignment through subspace constrained mean-shifts. In *Computer Vision, 2009 IEEE 12th International Conference on.* IEEE, 1034–1041.

Faisal Shafait, Ralph Kricke, Islam Shdaifat, and Rolf-Rainer Grigat. 2006. Real time lip motion analysis for a person authentication system using near infrared illumination. In *Image Processing, 2006 IEEE International Conference on.* IEEE, 1957–1960.

Tim Wark, Sridha Sridharan, and Vinod Chandran. 1998. An approach to statistical lip modelling for speaker identification via chromatic feature extraction. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conf.*, Vol. 1. IEEE, 123–125.

Di Wu and Qiuqi Ruan. 2014. Lip reading based on cascade feature extraction and HMM. In *Signal Processing (ICSP), 2014 12th International Conference.* 1306–1310.