Future of Information and Communication Conference (FICC) 2018
5-6 April 2018 | Singapore

Hand Gesture Authentication using Depth Camera

Jinghao Zhao
Graduate school of IPS
Waseda University
Kitakyushu, Fukuoka, Japan
Email: jinghao_zhao@126.com

Jiro Tanaka
Graduate school of IPS
Waseda University
Kitakyushu, Fukuoka, Japan
Email: jiro@aoni.waseda.jp

***Abstract*—Nowadays humans are concerned more about their privacy because traditional text password becomes weaker to defend from various attacks. Meanwhile, somatosensory become popular, which makes gesture authentication become possible. This research tries to use humans dynamic hand gesture to make an authentication system, which should have low limitation and be natural. In this paper, we describe a depth camera based dynamic hand gesture authentication method, and generate a template updating mechanism for the system. In the case of simple gesture, the average accuracy is 91.38%, and in the case of complicated gesture, the average accuracy is 95.21%, with 1.65% false acceptance rate. We have also evaluated the system with template updated mechanism.**

***Keywords*—Gesture authentication; three-dimensional hand gesture; depth camera**

I. Introduction

In recent years, how to protect human's privacy attracted the attention of people. Text password is the most frequently way that we used in our delay life, but recent research shows that such kind of password become weaker and need to be longer and more complex to be safety [1], which make it very inconvenient and unnatural. Besides, even we used such kinds of text password, it also can be easily stolen or copied by smudge attack or shoulder surfing attack [2], [3].

In order to make up for the shortage of traditional text password, biometric information is used as a new kind of authentication methods. Biometric information consists of physical and behavioral characteristics. Physical characteristics are referred to the static features of human body, such as iris, fingerprints or DNA, behavioral characteristics means dynamic information from human's behave, including gait, gesture, typing and so on. According to the source of biometric information, it also can be divided into congenital features and acquired features.

Current biometric authentication technologies are mainly based on computer vision technology with RGB camera or wearable sensor such as accelerometer embedded in wristbands or smart phone, those methods have already achieved good results, but both of them still have demerits. The computer vision technology can get high accuracy, but has high demands on equipments, users are always need to perform in special environment. The wearable sensor method is portable and usually does not have environment limitation, but it has some accuracy lost because of sensor itself, and some sensors have difficulty in power charging. Hence, how to combine the good points of each method to make a more natural authentication system become more important.

II. Background

We will introduce the basic processes in biometric authentication: first, we discussed some biometric techniques and application. Then introduce some authentication using hand gesture. Finally, we will give some research for evaluating an authentication system.

A. Biometric Techniques

Biometric techniques are used in authentication, such as iris [4] and fingerprint [5], other researchers also used face or palm print as authentication methods.

B. Hand Gesture Authentication

As an unique biometric information for individuals, hand gesture is proved based on several research that it can have good performance in authentication system. Liu et al. presented an accelerometer-based personalized gesture recognition methods, in their paper, acceleration data of hand gesture is sensed by mobile phone, and they used Dynamic Time Warping (DTW) to analysis and match the input gesture. Kratz et al. [6] presented an air-auth hand gesture authentication system using Intel Creative Senz3D, they also made some improvements and assessment methods. Chahar et al. used Leap Motion and Neural Network built a dynamic hand gesture authentication system [7]. Those researches shows that dynamic hand gesture could obtain high accuracy in authentication.

C. Evaluation methods

Jain et al. have proposed seven basic rules for an authentication system [8], which are universality, uniqueness, permanence, measurability for any biometric system and performance, acceptability, circumvention in a practical biometric system. Aumi et al. [9] discussed security performance of their system by the test of shoulder-surfing attacks and smudge attacks. Yang et al. [10] discussed how to evaluate the permanence and memorability of their research.

III. Goal and Approach

In our research, we are aimed to find a robust authentication methods by using human dynamic hand gesture. Since traditional text password are not friendly to users, while some biometric authentication methods such as iris and fingerprint methods need high requirements on equipments and can not be pervasive. Therefore, we try to find a natural and acceptable authentication method for users, also, it should have relative high performance.

978-1-5386-2056-4/18/$31.00

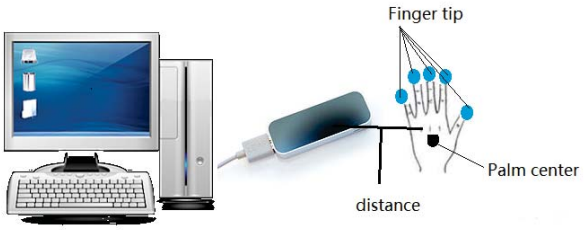


Fig. 1. System platform.

To achieve our goals, we take a series of approaches. We used new generation depth sensor - Leap Motion, which has lighter weight but keep a high accuracy. In order to use it with good result, we designed a series of preprocessing methods such as smoothing and normalization. Then, we came up a feature extraction method based on clustering and filtering. Finally, we used Hidden Markov Model (HMM) to solve the classification of identities.

Instead of using static hand shape or other static biometric information, we used dynamic hand gesture as our core elements for authentication. Because dynamic gesture contains more information than static posture, it is harder to be copied or imitated by others, meanwhile, it has lower equipment requirements than traditional fingerprint or iris authentication methods, which will make it can be used in more situation.

IV. System Implementation

To implement hand gesture authentication, we designed our system by using Leap Motion and a series algorithm including data processing, feature extraction and classification.

A. Prototype

Leap Motion is used as our hardware platform. We built our system with following prototype in Fig. 1.

The system will works in following process:

1) First, users are asked to perform their gesture over Leap Motion camera to register their account and "gestural password", all of registered data will be stored as user's own template.
2) Then, when a user want to access his account, just choose his account and perform the registered gesture, the system will automatically return whether current user can pass or not. Passed verification gesture will be stored as a temporary template waiting to be used to update current template.
3) After a period of time, the accuracy of current template will dropped, therefore, the system will merge the temporary template with current template and make it become a new template.
4) Some times users may frequently access their account, the system will also update the template when successful access times achieved the threshold.

And the flowchart in Fig. 2 also shows the sequence of system work.

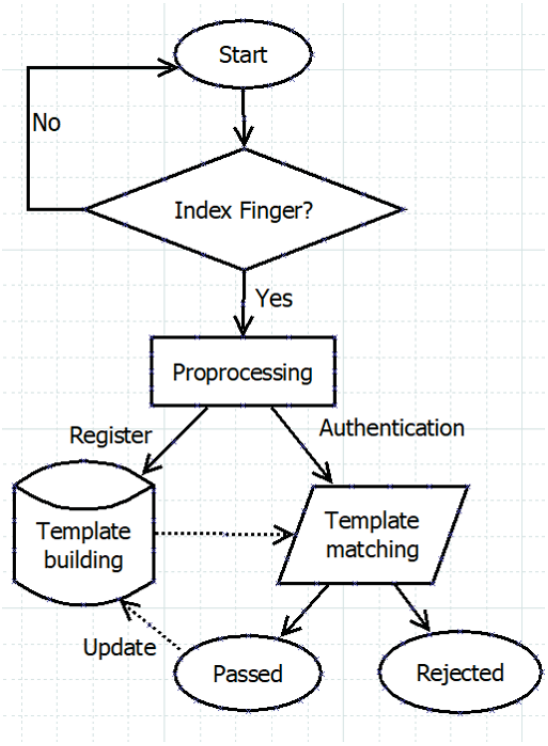


Fig. 2. Process of System.

B. Data Preprocessing

1) Data acquisition: After the platform is built in our research, we first got the raw data of hand gesture. Leap Motion will return gesture information automatically. The basic gesture information will be recorded frame by frame, and will be shown in the format below:

$$frame = \{frameid, timestamp, palmposition, fingertipposition, fingertipspeed, etc.\} \quad (1)$$

where, position or speed part consists of data from x, y and z-axis. For each feature, we extracted data and make them in time series as below:

$$f = \{(x_1, y_1, z_1), ..., (x_n, y_n, z_n)\} \quad (2)$$

Here, n means frame length of the gesture, and each tuple (x, y, z) is feature data in frame. After that we will get the raw data of each gesture. And the raw data will be the format as Fig. 3 shows:

2) Data smooth: Because there are noises' remains in the raw data, we used Kalman filter to eliminate those noises. The basic idea of Kalman filter is that current state f at time t is related to previous state at time t-1 [11], [12]. In processing control, suppose we have a system which can be described in a following linear stochastic difference equation,

$$s_t = As_{t-1} + BU(t) + C \quad (3)$$

where A and B is system parameter of this system, s_t is the feature data of time t, $U(t)$ is the external control, and C

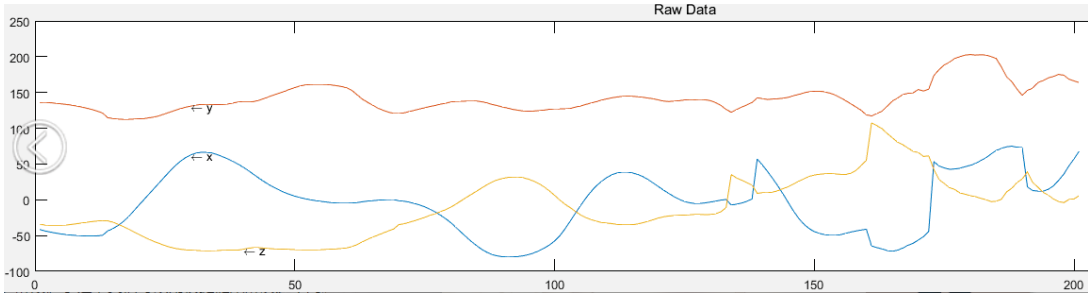


Fig. 3. Raw data.

is the measurement noises from devices. We only can measure the output of device.

$$y_t = Hs_t + N_t \tag{4}$$

where, y is the output value, H is parameter of system, and N is the noise in reading data from device. Obviously, the output value contains two part of noises, one is measurement error C, another is reading error N. We use Q and R to represent the covariance of C and N.

1) Set $s_{t|t-1} = As_{t-1} + BU(k)$ to estimate current $x_{t|t-1}$ value based on previous optimal x_{t-1} Since there are no external control in our system, the parameter B can be 0, and parameter A, H in (3) and (4) can be treated as 1.
2) Set $P_t = AP_{t-1}A^T + Q$ to calculate the covariance of estimated x_t, Here, P_t is the covariances of current state.
3) Used $Kg_t = P_tH^T(HP_tH^T + N)^{-1}$ to calculate the Kalman Gain Kg_t of the current time.
4) $s_t = s_{t|t-1} + Kg_t(y_t - Hs_{t|t-1})$ to get the current optimal x_t by estimated value $x_{t|t-1}$ and observed value y_t.
5) Finally, we updated the covariance by $P_t = Kg_tHP_t$;

And the smooth result is as Fig. 4 shows

3) Data normalization: To make authentication data easy to match the template data, we used

$$nf^k = \frac{\hat{f}_i^k - \bar{\hat{f}}^k}{\sigma\hat{f}^k} \tag{5}$$

to normalize (x_k, y_k, z_k) of feature data in frame k, Where $\bar{\hat{f}}^k$ means the mean of smoothed feature f, and $\sigma\hat{x}^k$ means the standard deviation of smoothed x sequence.

After that, the smoothed data will be changed into normalized sequence

$$nx^k = \{nx_1^k, ny_1^k, nz_1^k, .., nx_n^k, ny_n^k, nz_n^k\} \tag{6}$$

The normalized data is shown in Fig. 5.

C. Feature Extraction

In order to collect suitable features, a kind of cluster method is used in our system.

First, we put all features that is given by Leap Motion into a feature set X, including speed, track, and other information of fingertip and palm and bones.

$$X = \{F_1, F_2, ..., F_n\} \tag{7}$$

Here, F represents one kind of features, such as Fingertip location, speed and so on. Then, we build a empty set Y to store optimal feature combination. The feature extraction steps are:

1) Test each feature of X and find the feature F_i that performed best in X, put it into optimal set Y, and remove it from X.
2) Then, for remaining features in X, pick up one feature F_i each time and combine it with features of Y, test the new set and get result. If the performance is better than using current optimal set Y, the new set will replace Y set, otherwise Y will not be changed and F_i will be discarded.
3) Repeat Step 2 until feature set X become empty.

Table I shows the accuracy when only using one feature to classify. The position of fingertip perform best above all features, we put it into the optimal set Y at first, and remove it from feature set X. Then for features remain in X, we picked up one feature each time and made the union for Y and this feature, testing the accuracy and got the result in Fig. 6.

Therefore, the best combination of two features is fingertip position and fingertip speed. Since the accuracy of feature set with three features is not better than accuracy of fingertip position and speed, we decided to use the combination of these two features.

D. Classification

We used trajectory and speed as our features to be used in classifier.

1) First, we make the gesture information for k^{th} frame in following format:

$$f^k = (p_x^k, p_y^k, p_z^k, v_x^k, v_y^k, v_z^k) \tag{8}$$

while p is the position of one fingertip, and v is the speed of this fingertip.
2) Then we put the sequence into the classifier, and use Hidden Markov Model to build the model for each features.

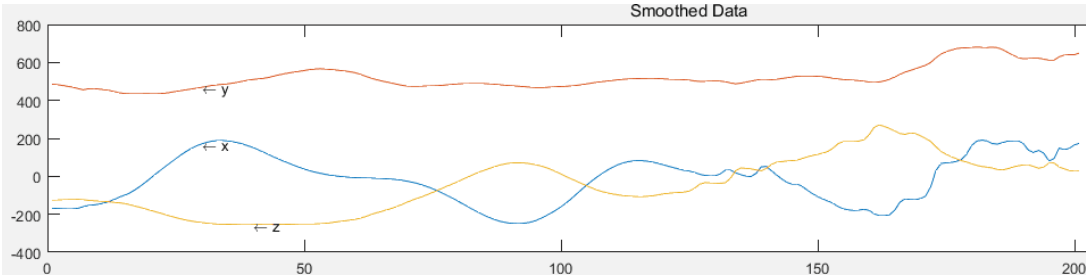


Fig. 4. Smoothed data.

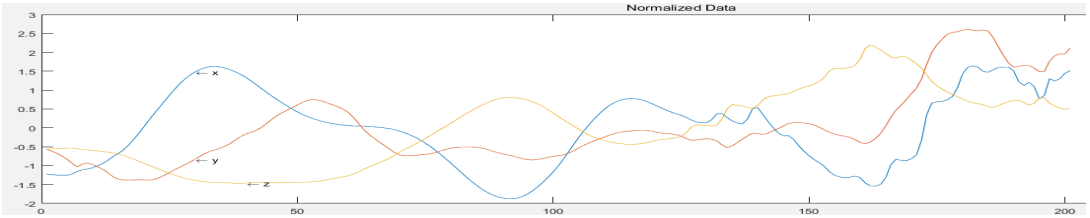


Fig. 5. Normalized data.

TABLE I. ACCURACY OF USING DIFFERENT FEATURES

Feature	palm center position	fingertip position	fingertip speed	wrist position
Accuracy	86.2%	88.4%	84.3%	83.2%
Feature	metacarpal bone	proximal bone	intermediate bone	distal bone
Accuracy	84.5%	81.6%	84.3%	81.2%

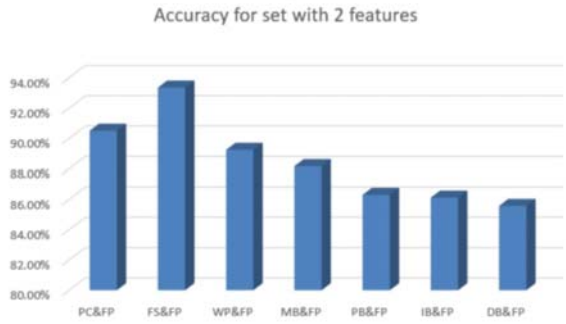


Fig. 6. Accuracy of using two features.

The classifier will work when f^k is input. First it calculates the state transition probability for each f^k, then we used the result of training set and calculated the output probability for each state. By calculating probabilities for each state, the classifier algorithm will build a whole hidden markov model for gesture data.

E. Template Updating

We used a double threshold method to update our gesture template, one threshold is the successfully accessed times, another is the build time for a template. In our research, the template updating mechanism work in following sequence.

1) First, the system will build the template for each user and record the last access date and access times.
2) If a user's accepted times achieved the threshold, the new accepted gesture will be attached to the existed template.
3) If the access time does not achieved the threshold, but it has been a long time after template built or last update times, the system will also update new gesture to the template.
4) Finally, system will update timestamp and access times to be used in next time.

V. EXPERIMENTS

We mainly did some experiments with the accuracy of original system and template updating mechanism.

A. Evaluation Standard

To evaluate an authentication system, first thing to think is the False Acceptance Rate (FAR) because the performance usually depends on the ability of defending. It should also keep good performance on False Rejection Rate (FRR).

False Acceptance Rate is the probability that the system accepted a non-authorized person, and also it is the probability that the system incorrectly rejects a genuine person.

In an identification system, false rejection rate is always concerned to be more important and should be decreased, but in a verification system, in order to have high security performance, False acceptance rate should be emphasized to prevent from attacks. In such case, false rejection rate can be relatively high to reduce false acceptance rate.

B. Accuracy of General Authentication

In this part, we invited 4 users aged 22-24, each of the user have no experience with such kind of system but have related

knowledge about how to use Leap Motion. The experiment step are as follows:

1) Each user registers a gesture and perform it 20 times, then the system will store the gesture as their template. And they can see the gesture of other users.
2) After built the template, each user try to pass the authentication with their own hand gesture for 20 times, this step will give the false rejection rate of the system.
3) Then, each user try to imitate each others gesture and try to attack their "account", then the false acceptance rate will be recorded.
4) Finally, calculate the total error using all data from Steps 2 and 3.

We did experiments with simple and complicated gesture separately, Fig. 7 and 8 show some examples of two kinds of gesture in 2D vision.

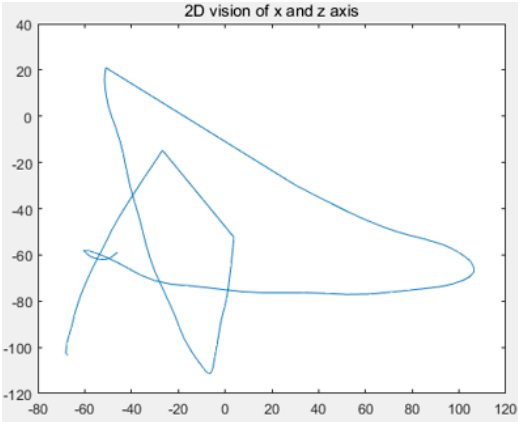


Fig. 7. Simple gesture.

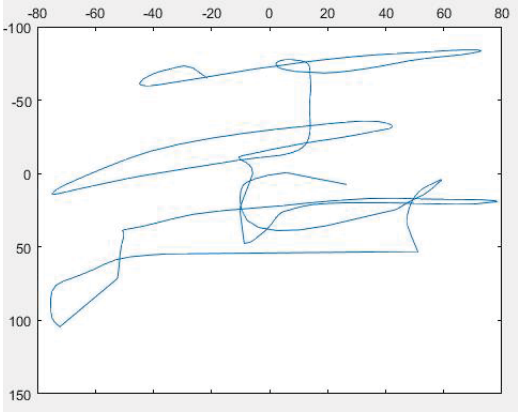
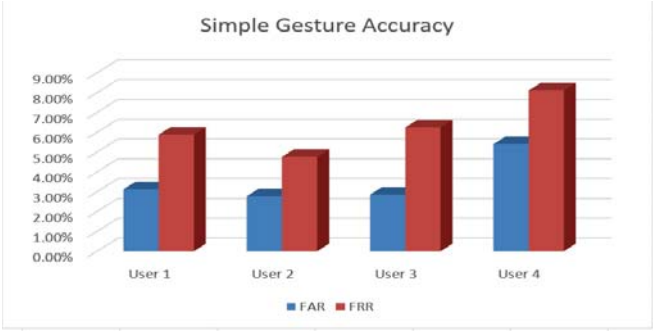


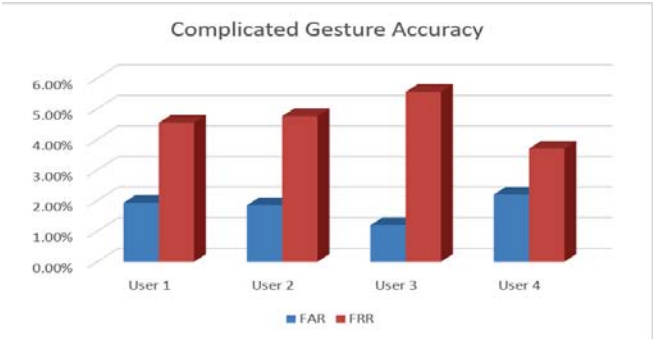
Fig. 8. Complicated gesture.

The accuracy result of two group is as Fig. 9 shows:

As Fig. 9(a) shows, in the case of simple gesture, the average accuracy is 91.38%, while false acceptance rate is 3.62%, and false rejection rate is 6.57%. In the case of complicated gesture as Fig. 9(b), the average accuracy is 95.21%, with 1.65% false acceptance rate and 4.82% false rejection rate.



(a) Simple gesture



(b) Complicated gesture

Fig. 9. Accuracy result.

C. Accuracy of Template Updating Mechanism

In the experiments of template updating mechanism, we designed steps as follows:

1) First, we invited 10 participants (aged 21-24) who are also not familiar with this system but have experience of using Leap Motion.
2) Then, 10 users are asked to perform their gesture, which will be registered as gesture password.
3) Then, 10 users were asked to access their account right after register, 5 of the users used system with template updating mechanism(Group T), others(Group O) used general system without updating template. And then we recorded the accuracy with FAR and FRR.
4) After one week, two weeks and one month, they were asked to access their account again as Step 2 works, and we recorded the accuracy.
5) Finally, we compared these three accuracy records and got the conclusion.

The accuracy comparison of template updating mechanism is shown in Fig. 10.

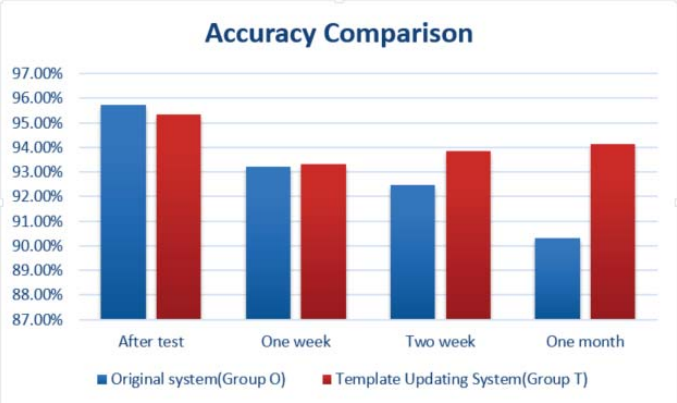


Fig. 10. Accuracy comparison of template updating mechanism.

We can see, at first, two group have almost same accuracy (T = 95.34%, O = 95.72%) after one week, the accuracy of two group are generally equal and did not changed a lot, but after two week and one month, the accuracy of group O dropped, while group T keeps a relatively high accuracy. Hence, we can prove that our template updating mechanism works in this system.

VI. Related Work

Chahar et al. [7] used Leap Motion and hand gesture built a Leap password system. In their work, they proposed a aLPhabet framework which used hand and finger static shape data such as length and width information, combined with time information for each user to perform a whole gesture and verify user's identity. In their work, they used Levenshtein Algorithm to estimate the similarity for gestures,and gave weight to each kind of feature to ranking the importance of features, finally, they used Naive Bayes, Neural Network and Random Decision Forest classifier and get an average score for each possibility and get the classification result. In their work, they kept a 1% FAR, and got an accuracy about 81%.

Compared with Chahar's work, our system got a relative high accuracy with some lost on FAR, meanwhile, we used dynamic hand gesture in our system, which is more unique and more difficult to be copied by tools, and we did some attack and safety experiments to prove the stable of our system.

Aumi et al. [9] used dynamic hand gesture and Intel Senze 3D in their work, and they used Dynamic Time Warping (DTW) as their classification method. They also did some threat experiments, such as shoulder-surfing threat, and got a high accuracy if they set a very low threshold for DTW. Besides, they designed a template updating mechanism by counting successful access times.

Compared with Aumi's work, we used Leap Motion, which is lighter and easier to use. Meanwhile, we proposed a double threshold template updating mechanism based on period and access times, which can reduce the threaten of false acceptance.

VII. Conclusion and Future Work

A. Conclusion

Our research generated a high accuracy dynamic hand gesture based human authentication system. From the general accuracy result we can see our system got an over 95% accuracy with 1.65% false acceptance rate with complicated dynamic hand gesture, compared with previous research [13], we made some improvement on the accuracy performance. Besides, we introduced template updating mechanism of our system, by using this mechanism, the system have keep its authentication accuracy with time passing, which proved the good permanence and robust of our system.

From the experiment result, it is obvious that complicated gesture perform better than simple gesture, which has same reason with current text password but gesture is easier to remember. Hence, we recommend to use a complicated gesture such as personal sign in our system.

After the experiments, most of participants think our system is easy to use and have better security performance than traditional password, which proved the usability of our methods.

B. Future Work

In our research, we required every user perform over 20 times to build their own template, which is not convenient in practice, what we should do next is to find a method that can reduce the repetition while keep the accuracy and robust of template.

Another problem is the template updating method. In our design, we used double threshold mechanism which consists of time and template length to reduce the influence of false acceptance situation and improve the permanence. But such kind of methods can not completely eliminate the threaten of false acceptance. We will think about how to improve the template updating mechanism to make the system safer.

References

[1] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek, "Usability and security of text passwords on mobile devices," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 527–539.

[2] A. J. Aviv, K. L. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens." *Woot*, vol. 10, pp. 1–7, 2010.

[3] A. H. Lashkari, S. Farmand, D. Zakaria, O. Bin, D. Saleh *et al.*, "Shoulder surfing attack in graphical password authentication," *arXiv preprint arXiv:0912.0951*, 2009.

[4] J. Daugman, "How iris recognition works," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 21–30, 2004.

[5] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[6] S. Kratz and M. T. I. Aumi, "Airauth: a biometric authentication system using in-air hand gestures," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2014, pp. 499–502.

[7] A. Chahar, S. Yadav, I. Nigam, R. Singh, and M. Vatsa, "A leap password based verification system," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–6.

[8] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[9] M. T. I. Aumi and S. Kratz, "Airauth: evaluating in-air hand gestures for authentication," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 2014, pp. 309–318.

[10] Y. Yang, G. D. Clark, J. Lindqvist, and A. Oulasvirta, "Free-form gesture authentication in the wild," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 3722–3735.

[11] R. Faragher, "Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]," *IEEE Signal processing magazine*, vol. 29, no. 5, pp. 128–132, 2012.

[12] G. Welch and G. Bishop, "An introduction to the kalman filter," 1995.

[13] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.