Learning a Policy for Gesture-Based Active Multi-touch Authentication

Raquel Torres Peralta, Anton Rebguns, Ian R. Fasel, and Kobus Barnard

Universidad de Sonora, Departamento de Ingenieria Industrial,
The University of Arizona, Department of Computer Science,
Tucson, AZ 85721-0077
{rtorres,anton,kobus}@cs.arizona.edu,
ianfasel@gmail.com

Abstract. Multi-touch tablets can offer a large, collaborative space where several users can work on a task at the same time. However, the lack of privacy in these situations makes standard password-based authentication easily compromised. This work presents a new gesture-based authentication system based on users' unique signature of touch motion when drawing a combination of one-stroke gestures following two different policies, one fixed for all users and the other selected by a model of control to maximize the expected long-term information gain. The system is able to achieve high user recognition accuracy with relatively few gestures, demonstrating that human touch patterns have a distinctive "signature" that can be used as a powerful biometric measure for user recognition and personalization.

1 Introduction

Tabletop devices allow the interaction by touch using a comfortable interface highly visible to those individuals close to the display. This particular characteristic has made the authentication of a user a challenging task.

Currently, one solution is to use a password. However, a relevant issue with touch devices is the lack of privacy while typing characters. The same problem applies to a gesture-based authentication systems. The alternative then is to use special "signature" a user leaves unconsciously on the touch of the surface and within the gesture itself. But, does this special signature exists? To answer this question, we examined basic one-stroke gestures of 8 different users using different representations based on speed and shape with the goal of finding the features that differentiate one user from the rest when drawing a set of gestures.In this paper, we propose a new gesture-based authentication system for touch devices, which does not consist on a secret combination, but rather in the shape and speed when drawing a particular gesture. The authentication can be made after a series of gestures requested systematically and strategically by the system with no need of privacy. Although no high-precision equipment was used, it was possible to achieve high recognition rates after a few gestures.

L. Marinos and I. Askoxylakis (Eds.): HAS/HCII 2013, LNCS 8030, pp. 59–68, 2013.

2 The Experiment

The samples were collected at a business office, using a 36in x 22in multi-touch tablet in horizontal position (as a table). Gestures were restricted to an area of 20 x 20 inches on the surface. Within that constraint, participants freely performed the gesture the size they preferred using either hand. The data was captured using the Touchlib library. Eight participants had three sessions scheduled at different days and times (one participant at a time).

2.1 Procedure

The participants were between 24 and 33 years old. Four were male and four were female. They were asked to reproduce eight different gestures (Figure 1) over the tablet, including breaks to avoid fatigue. The samples were asked to be provided sequentially (The participant performed the same gesture a number of times before passing to the next one) and randomly (The gestures were performed in an unordered sequence).

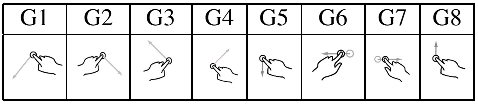

Fig. 1. Vocabulary of gestures. The gestures were selected considering the possibility of using them in different devices of different size.

The samples were collected using a multi-touch tablet using a frustrated total internal reflection (FTIR) technology, which is very sensitive to light pollution such as reflections from shiny clothes. Given the characteristics of the device, the trajectories captured are more susceptible to noise than other non-optical devices, exhibiting more detection of false-blobs not part of the gesture. In some cases the number of false blobs detected as part of the trajectory provided a noisy sample even after the noise reduction process. In general, we did not attempt to exclude such cases from the dataset, since the intention was to test the approach using the data as obtained from a real setting.

3 Data Representation

In this study, a gesture is a series of touches belonging to the same trajectory. Trajectories with less than 3 blobs are considered “orphan blobs” and were removed from our data to reduce noise. All samples were then resized to 30 points long using interpolation and smoothed with local regression using weighted linear least squares and a 2nd degree polynomial model. The smoothing was done using the rloess function in Matlab.

Our vocabulary of gestures consists of simple one-stroke gestures to identify the characteristics of basic movements. In order to extend the potential application of this study beyond FTIR multi-touch tablets, we worked only with x and y coordinates which represent 2D trajectories. We worked with two basic representations based on shape and speed.

1. Angle Representation. The trajectories were converted to a vector of 29 points, each one representing the angle between successive points. Giving a pair of points *(x_1, y_1), (x_2, y_2)*:

$$Angle = \frac{arctan(y_2 - y_1, x_2 - x_1) \times 180}{\pi} \tag{1}$$

2. Speed. Since the system captures the data of the touch in equal time intervals, the speed at each time is represented as the distance between consecutive points along the trajectory.

A total of 34 samples per user per gesture were collected from the sequential and random sets. The trajectories included in this work were all restricted to five points long, leaving 25 to 29 samples per user per gesture after filtering. For training, 768 samples were randomly selected (12 per user-gesture), leaving the rest for the testing set, having a total of 1000 samples (13 to 17 per user-gesture). The results and procedures presented in this paper assume the accurate recognition of the corresponding gesture for every sample.

4 User Recognition Using One Gesture

We trained a multi-class SVM for each gesture (eight classes, one per user). The models were trained using LIBSVM [2], with RBF kernels with soft margins, with the kernel degree and margin slack variables determined empirically through 5-fold cross validation using a grid search, independently for each model. The model prediction is the Platt's probability output [8] of the sample belonging to each user. The predictions were 30% accurate for the angle representation, while speed achieved 34.3%. Assuming independence, the angle and speed probability distributions were multiplied to create the third representation of angle and speed combined, increasing the accuracy to 38%, which was still not acceptable.

Table 1. User recognition average results using speed representation for user 1 performing gesture 1. This user is predicted 70.5% as user 1, 17.6% as user 4, 5.8% as user 2 and 5.8% as user 7.

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8
User 1	70.6%	5.9%	0.0%	17.7%	0.0%	0.0%	5.9%	0.0%

The results suggest that some users share certain similarities in the way they draw the gestures. In Table 1, the results for user recognition using the speed representation for gesture 1 shows that user 1 is correctly predicted 70.5% of the time, but also is predicted as user 4 (17.6%), user 2 (5.8%), user 7 (5.8%), but never as user 3, user 4, user 5 or user 8.

5 Combining Gestures for Authentication

The results above suggest that user authentication can not be accurately achieved using just one sample of one-stroke gestures. Thus, we investigate a strategy of combining several gestures to reduce uncertainty over the user's identity. We observed that most users tend to be predicted as one or two specific users, and some gestures are more informative than others depending on the individual, which makes the order they are obtained an important factor to reduce the number of samples needed for authentication.

6 Multi-action Category Model

The SVM model outputs a probability distribution[1] over all classes (8 different users) for a gesture drawn by a user (similar to the sample provided by Table table:SpeedG1). When combining gestures, the probabilities could be multiplied to provide a result assuming independence. However, the independence assumption may be too strong. For instance, if a gesture is repeated by a user, then taking products of the SVM Platt's probability estimates would usually result in an overly high confidence for one user even if that user only gets slightly higher probability on each individual trial. These problems suggest an intermediate representation to deal with the lack of independence that allow the combination of several samples to increase accuracy in the authentication process.

A similar problem is presented by Regbuns et al [9], where an InfoMax controller has been implemented on a robot to identify objects.The acoustic similarities are represented by a Dirichlet distribution using a Latent Dirichlet Allocation (LDA) model [1]. This approach is implemented in the authentication problem to reduce the number of gestures needed during the process.

6.1 Semi-latent Dirichlet Allocation for an Intermediate Representation

In the user recognition problem, the similarities in the way users draw a gesture was the main reason of a low accuracy rate (users were predicted as one or two users who used to draw the gesture in a very similar way). Thus, it is important to explicitly model the fact that some users look somewhat like others under certain actions/gestures.In our framework we use the Platt's probability outputs obtained by the SVM multiclass model for user recognition to represent the correlations between one user and the rest when drawing a specific gesture. Then, a LDA samples a Dirichlet distribution of shape-speed correlations that can be used by an Infomax controller to learn the best policy per user. The Dirichlet distribution describes how latent clusters mixing proportions ϕ vary among a collection of samples. Originally, the number of underlying latent clusters is unknown and determined in the training phase according to the observations. If the number of clusters is known in advance, that can be used to an advantage, as done, for example, by Wang et al. [12]. The same approach is used in this work to train the model using the already known number of classes, specifically, the number of users.

[1] The multiclass model was trained and tested using the LIBSVM library which provides the Platt's probabilistic outputs [8], following the improved version provided by Lin et al [5].

In this model, ϕ represents the Platt's probability distribution of a sample from a specific gesture belonging to the different users.

The model parameters were obtained using the set of outputs of the user recognition model for the training set, with the angle-speed combined representation (section 4).The probability of generating probabilities ϕ by drawing gesture a by user i is

$$p(\phi|a,i) = \frac{\Gamma(\sum_{j=1}^{N} \alpha_{aij})}{\prod_j \Gamma(\alpha_{aij})} \prod_{k=1}^{N} \phi_k^{\alpha_{aik}-1} \tag{2}$$

where $\alpha_{ai} = (\alpha_{ai1}, ..., \alpha_{aiN})$ are the Dirichlet distribution parameters over probabilities for user i under gesture a, and N is the number of classes, in this case, the number of users.

6.2 Handcoded and Learned Policies

A policy describes a particular order the gestures must be obtained during the authentication process. For this study, we define a handcoded policy where the samples are requested each of the eight gestures in turn, cycling if needed (G1, G2,...G8, G1, G2...). Also, we used a model of control to find the best policy to reduce the number of gestures required to authenticate each user on section 7. The LDA representation then was used for each gesture and the policies were tested separately on section 8.

7 InfoMax for Optimal Policy Learning

A handcoded policy could work well for some users, but not for all. One way to reduce the number of gestures needed to increase accuracy is to ask for the most informative gestures per individual. For this purpose we used Infomax, a model of control that maximizes the information gained about events of interest, used to model behaviors in agents [6]. In this work, Infomax is used to learn the best policy for each user. Following the approach proposed by Regbuns et al [9], an optimal policy for gesture selection can be found using the *Policy Gradients with Parameter Exploration* (PGPE) algorithm [11].

Let q_t be a d-dimensional vector combining the system's current beliefs about the user and its known internal state. Define the set of possible gestures $A = \{$*gesture 1, gesture 2, gesture 3, gesture 4, gesture 5, gesture 6, gesture 7, gesture 8*$\}$. Then let the function $F_\theta : Q \rightarrow A$ be a deterministic controller with k-dimensional parameter θ which at each time t takes as input a state-variable q_t and outputs an action (gesture to perform) a_t. Then, q_t is the representation, constructed by $(p', c', \psi(t))$s where p' a vector of 8 elements representing the current beliefs of the system for each user, c' is a vector of counters for each gesture provided by the user ignoring order and $\psi(t) = (\psi_1(t), \psi_2(t), \psi_3(t))$ is a vector of Radial Basis Functions (RBFs) of the time t. The center of the RBFs are equally distributed over the specified number of steps (gestures provided for episode) and the learned policy relies on the completition of all steps. Let a history $h = (q_1, a_1, ..., q_T, a_T)$ be a sequence of T state-gesture pairs induced by using a controller with parameters θ. The reward at time t of history h is the scaled negative Shannon entropy of the belief distribution.

$$\mathcal{R}(q_t|h) = \left(\sum_i p_i^{(t)} \log p_i^{(t)}\right) \tag{3}$$

where $p_i^{(t)}$ is the system's belief of the current user being user i at time t.

7.1 Policy Learning

To find the parameters θ that maximizes the total reward over training histories, of length L,

$$\Phi(\theta) = argmax(Hmax - (E_h[\sum_{t=1}^{L} \mathcal{R}(q_t|h)p(h|\theta)])/Hmax) \tag{4}$$

Where E_h is the expected total reward over histories and Hmax is the maximum possible entropy, the *Policy Gradients with Parameter Exploration* (PGPE) algorithm [11] is used.

InfoMax policies were trained for 500 episodes of PGPE using the gesture-user specific Dirichlet distributions. The experiments were performed in simulation by sampling sequences called for by the controller. For each episode, a set of samples from all users were sampled, and the system's beliefs about the users were initialized to uniform. At each time step, provided the state vector q_t, an action is selected using the current policy. Each full learning trial of PGPE was repeated 20 times. The results show averages across 150 experiments. The software used for these experiments is based on the library developed by the Arizona Robot Lab, available online.[2]

8 Results

We compared the accuracy of the learned policies against the handcoded policy.and ran separate experiments for the two most accurate representations for user recognition, speed and speed combined with angle, to measure the improvement for each (see Figure 2). In what follows, the number of steps represent the number of gestures required during the authentication process.

8.1 Speed

The speed representation achieved 34% (average overall gestures) for user recognition. Using this representation, the handcoded policy reaches a 50% of accuracy after the first gesture (G1) while the learned policy has 64%. After 9 gestures, the Handcoded policy has 77% of accuracy and the learned policy gets 88% (Figure 2).

[2] The package can be found at `http://code.google.com/p/ua-ros-pkg/`. For the package documentation check `http://ros.informatik.uni-freiburg.de/roswiki/doc/api/ua_audio_infomax/html/index.html`. The number of objects in the modified version is set to 1 and the entropy computation is modified as specified in this paper.

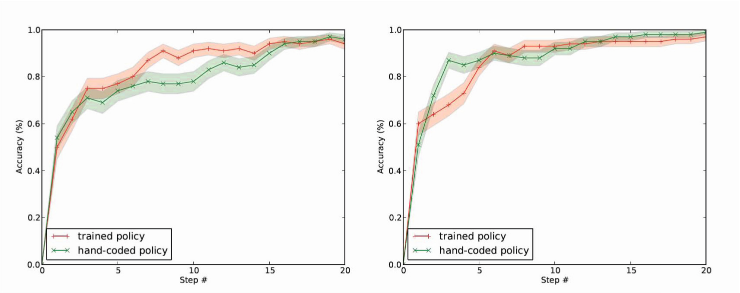

Fig. 2. Accuracy per step, hand-coded(red) vs learned policy(green) for both representations. The accuracy is averaged overall users.

Table 2. Classification Accuracy per Step for Speed Representation. The Number of Steps indicate the number of gestures provided by the user. The accuracy is expressed as an average percentage overall 8 users at the corresponding number of steps.

Number of Steps	1	4	7	9	11	13	15
Accuracy for Learned Policy	50%	75%	87%	88%	92%	92%	94%
Accuracy for Handcoded Policy	54%	69%	78%	77%	83%	84%	90%

8.2 Angle and Speed Combined

Recall that this representation had only 38% success in recognizing the user from a single gesture. In Table 3, the learned policy has 60% of accuracy while the handcoded has 50%. The diferences between both policies may not be large, but the learned policy shows a constant increment at each step.

Table 3. Classification accuracy per step for angle and speed representation. The number of steps indicate the number of gestures provided by the user. The accuracy is expressed as an average percentage overall samples from 8 users at the corresponding number of steps.

Step Number	1	4	7	9	11	13	15
Accuracy for Learned Policy	60%	73%	89%	93%	94%	95%	95%
Accuracy for Handcoded Policy	51%	85%	89%	88%	92%	95%	97%

User 3 has a high recognition rate. It is the one with a higher probability from the first gesture by the handcoded and the learned policy (see Figure 3). However, the latter has a higher rate from the first gesture following the best policy.

Contrary to the predictability of user 3, user 8 is not easy to recognize. The joint distribution following the handcoded policy favors user 8 after 11 gestures. With the learned policy, user 8 gets the higher value after only 3 gestures (G6, G1, G3) as shows Figure 4. The learned policy outperforms or ties the handcoded policy in almost every case.

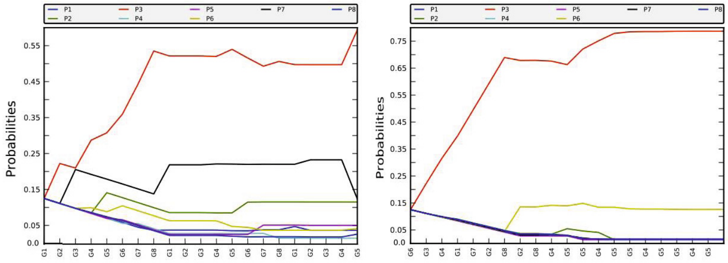

Fig. 3. Probabilities for the Handcoded policy using the Angle and Speed Combined Representation with samples provided by User 3. The X axis shows which gesture is provided while the Y axis shows the probability. The probability distribution shows User 3 as the highest after the first gesture.

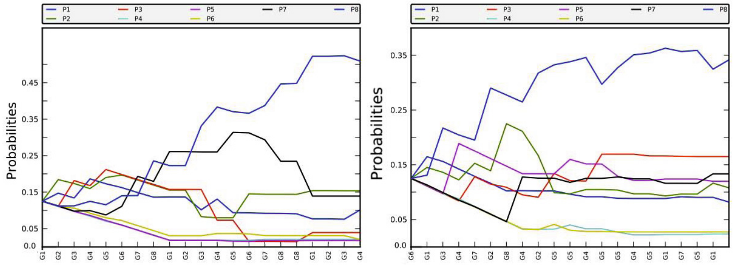

Fig. 4. Probabilities per step with the handcoded policy using the angle and speed combined representation with samples provided by user 8. The X axis shows which gesture is provided while the Y axis shows the probability. The handcoded policy (left) recognizes user 8 after 11 gestures while the learned policy (right) recognizes user 8 on the 3rd gesture.

9 Related Work

In the past, there has been interest in recognizing Tabletop devices' users. Some projects, such as the DiamondTouch table [4] and the IR Ring [10] have made use of extra devices to detect the origin (user) of the touch.

These approaches yield the problem that identifying a device is not the same as identifying a user, thus, anyone holding the device can be authorized.

Microsoft has implemented a gesture-based login for their TabletPC [3]. A user chooses a photo, and performs a sequence of secret getures which must be performed in the right order. Even when involving different levels of secrecy, the approach does not solve the shoulder surfing problem.

Because conventional approaches to user authentication have limited value when applied to collaborative multi-touch devices, there is a clear need to explore alternative methods. One recently introduced method [7] has been to include biometric data.

[3] Building Windows 8 - An inside look from the Windows engineering team. Signing in with a picture password. `http://blogs.msdn.com/b/b8/archive/2011/12/16/signing-in-with-a-picture-password.aspx`

The user places five fingers in the surface of the device and makes a rotation gesture. This approach is more than 90% accurate recognizing the user, and it confirms the existence of a singular signature in gestures that could make individuals differentiable, but the approach is not useful for small surfaces. A similar study using simple stroke gestures [3] showed that some users are harder to identify than others, as found in this work.

10 Advantages and Limitations

This method deals with the issue of similarities between some users on some of the gestures and the resulting lack of independence. The method also reduces the number of gestures needed based on the user. More generally, it provides a framework for this kind of problem that can be used with a number of alternatives for its components. For example, the angle and speed representations can be replaced by others. Also, the model's features could be different from outputs of an SVM as used here. Finally, other approaches to learning a good policy could be used instead of the one we chose.

One disadvantage of this method is that in some cases it could take a lot of samples to reach an acceptable probability about a user's identity (some users are harder to identify than others). One weakness is the limited number of participants, but the data is realistic, and notably samples for each user are hard to distinguish from the others. Since the number of users that share these devices is usually small, our experiments are informative for many applications. However, for deployment, data from a larger set if people is called for. In a future work, new representations must be tested and a different device, as a smart phone or a tablet, should be used to obtain cleaner data.

11 Conclusions

Authentication for Tabletop devices' passwords should not depend on a secret combination of characters or gestures, but on a private signature not subject for duplication. Our experiments suggest that authentication with a single one-stroke sample using shape, speed or both features is not enough, since some users tend to look alike from system's point of view. However, the combination of several samples can make possible the authentication of a user without the need of secret combinations. Having a Dirichlet distribution as an intermediate representation is a way to deal with the correlations between users. Also, an Infomax controller can learn the best policy to reduce the number of gestures required to obtain high recognition accuracy. The results showed that user authentication based on 2D gestures is a challenging task that might require more multidisciplinary studies (involving for instance, usability, psychology and even anatomy) to improve the results here presented.

Acknowledgements. We thank Federico Cirett and the Teknol company for their support during the data recollection process. This work was funded by ONR contract N00014-09-1-0658 and DARPA contract N10AP20008. Co-Author R.T.-P. was funded by a Conacyt Fellowship during this study.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J.: Latent dirichlet allocation. Journal of Machine Learning Research 3(2003) (2003)
2. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
3. De Luca, A., Hang, A., Brudy, F., Lindner, C., Hussmann, H.: Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI 2012, pp. 987–996. ACM, New York (2012)
4. Dietz, P., Leigh, D.: Diamondtouch: A multi-user touch technology. In: ACM Symposium on User Interface Software and Technology (UIST), pp. 219–226. ACM Press (2001)
5. Lin, H.-T., Lin, C.-J., Weng, R.C.: A note on platt's probabilistic outputs for support vector machines. Mach. Learn. 68, 267–276 (2007)
6. Movellan, J.: An infomax controller for real time detection of social contingency. In: Proceedings of the 4th International Conference on Development and Learning, pp. 19–24 (July 2005)
7. Sae-Bae, N., Ahmed, K., Isbister, K., Memon, N.: Biometric-rich gestures. a novel approach to authentication on multi-touch devices. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI 2012, pp. 977–986 (2012)
8. Platt, J.C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in Large Margin Classifiers, pp. 61–74. MIT Press (1999)
9. Rebguns, A., Ford, D., Fasel, I.: Infomax control for acoustic exploration of objects by a mobile robot. In: AAAI Workshop on Lifelong Learning. AAAI (2011)
10. Roth, V., Schmidt, P., G´uldenring, B.: The ir ring: authenticating users' touches on a multitouch display. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010. ACM (2010)
11. Sehnke, F., Osendorfer, C., Rucksties, T., Graves, A., Peters, J., Schmidhuber, J.: Parameter-exploring policy gradients. In: Neural Networks, pp. 551–559 (2009)
12. Wang, Y., Sabzmeydani, P., Mori, G.: Semi-latent dirichlet allocation: A hierarchical model for human action recognition. In: 2nd Workshop on Human Motion Understanding, Modeling, Capture and Animation (2007)