Active User Authentication for Smartphones: A Challenge Data Set and Benchmark Results

Upal Mahbub[1] Sayantan Sarkar[1] Vishal M. Patel[2] Rama Chellappa[1]
[1]Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742
{umahbub, ssarkar2, rama}@umiacs.umd.edu
[2]Rutgers, The State University of New Jersey, 508 CoRE, 94 Brett Rd, Piscataway, NJ 08854
vishal.m.patel@rutgers.edu*

Abstract

In this paper, automated user verification techniques for smartphones are investigated. A unique non-commercial dataset, the University of Maryland Active Authentication Dataset 02 (UMDAA-02) for multi-modal user authentication research is introduced. This paper focuses on three sensors - front camera, touch sensor and location service while providing a general description for other modalities. Benchmark results for face detection, face verification, touch-based user identification and location-based next-place prediction are presented, which indicate that more robust methods fine-tuned to the mobile platform are needed to achieve satisfactory verification accuracy. The dataset will be made available to the research community for promoting additional research.

1. Introduction

The recent proliferation of mobile devices like smartphones and tablets has given rise to security concerns about personal information stored in them. Studies show that users are more concerned about the security of their cell phones over laptops [5]. Though over 40% of users in major U.S. cities have lost their phones or have been victims of phone theft [12], industry surveys estimate that 34% of smartphone users in the U.S. do not lock their phones with passwords [1]. This contradictory behavior is due to the time-consuming, cumbersome and error-prone hassles of entering passwords on virtual keyboards or due to users' beliefs that extra passwords are not needed [12]. 76% attacks on smart phones exploit weak passwords [39], but users still prefer those over stronger passwords, as the stronger passwords are difficult to remember and type, especially since the average cell phone user checks their smartphone device 150 times per day [26].

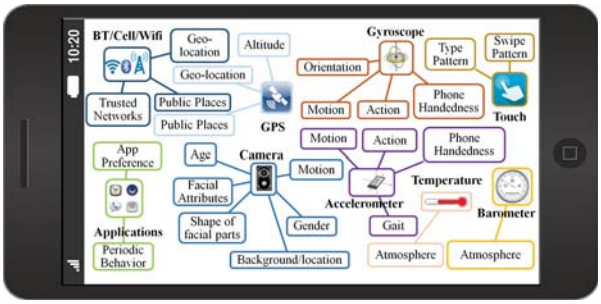

Figure 1. Association of smartphone sensors with behavioral and biometric information.

Going beyond traditional passwords and fingerprint-based one-time authentication, the concept of Active Authentication (AA) has emerged recently [29], where the enrolled user is authenticated continuously in the background based on the user's biometrics such as front camera face capture [34], [10], touch screen gesture [11], [42], typing pattern [2] etc. Conceptually, in an AA system users do not password-lock the phone at all. When a user uses the phone, the AA system compares the usage pattern with the enrolled user's pattern of use. If the system deems that the usage patterns are sufficiently similar, the phone's full functionality (including sensitive applications and data) is made available, else it blocks the current user. At present, most of the AA systems are based on face, touch and typing pattern biometrics. As shown in Fig. 1, modern smartphones provide multiple sensors associated with a variety of behavioral and physiological biometric information, however research on multi-modal authentication using multi-sensor data is lagging behind, because of paucity of datasets.

The first non-commercial dataset on smartphone us-

*This work was done in partnership with and supported by Google Advanced Technology and Projects (ATAP), a Skunk Works-inspired team chartered to deliver breakthrough innovations with end-to-end product development based on cutting edge research and a cooperative agreement FA8750-13-2-0279 from DARPA.

age containing a wide range of sensor data, namely the University of Maryland Active Authentication Dataset 02 (UMDAA-02), is introduced in this paper. Unlike task-based data collection schemes, the data collection was passive and hence is representative of the natural, regular smartphone usage by the volunteers. The data collection application ran on the Nexus-5 device, completely in the background, saving sensor data and periodically uploading the data to a secure online location.

The benchmark results of 4 experiments on the UMDAA-02 dataset are reported in this paper. Face is the most widely used biometric, but the images captured by the front-facing camera of smartphones present certain challenges such as partial face detection under occlusion and large variations in pose and illumination. The face images in the UMDAA-02 dataset are difficult to detect and the performances of traditional face detection methods are explored on a smaller annotated subset of the dataset. Next, faces of the annotated subset are verified using multiple state-of-the-art features and distance measures. On the full dataset, swipe-based user identification has been performed. Also, utilizing the user's geolocation information, the next place prediction experiment is performed which can be useful in AA research when fusing multiple modalities.

2. Previous Works

Among the AA techniques, the most explored are based on faces [10], [34], touch/swipe signature[37], multi-modal fusion [42], gait [8] and device movement-patterns/accelerometer [30], [6]. Face-based authentication, though most accurate, requires more computational power and can cause faster battery drain if the images are captured frequently. On the other hand, swipe and accelerometer data alone are not discriminative enough. Among the other AA approaches, in [14], the authors fused stylometry with application usage, web browsing data and location information.

Various protocols for AA with and without multi-modal fusion have been suggested over the years. In [32], the authors explored the idea of progressive or risk-based authentication by combining multiple verification signals to determine the users level of authenticity. The AA system surfaces only when this level is too low for the content being requested. In [18], the authors proposed context aware protocols for more flexible yet robust authentication. In [33], the authors discuss three possible levels of fusion (a) fusion at feature level, (b) fusion at score level, and (c) fusion at decision level. Different fusion algorithms based on k-Nearest Neighbour classifiers, Support Vector Machines, decision trees, Bayesian methods, Gaussian Mixture Models (GMM) have been employed. [33], [7].

The MOBIO dataset [25] is a well-known dataset for face-based AA research. It contains 61 hours of audio-visual data from a NOKIA N93i phone (and a 2008 Mac-book laptop) with 12 distinct sessions of 150 participants spread over several weeks. However, since users were required to position their head inside a certain elliptical box within the scene while capturing the data, the face images of this dataset do not represent real-life acquisition scenarios.

Faces captured by the front camera (and also screen touch data) of University of Maryland Active Authentication Dataset (UMDAA-01) [42][34] of 50 users are unconstrained and hence presents a more realistic and challenging scenario for face-based continuous authentication where partially visible, frontal and non-frontal faces under various illumination conditions are available. In [23] and [36], the authors introduced facial segment-based face detection (FSFD) method and deep feature-based face detection for UMDAA-01-FD which is a small annotated subset of the UMDAA-01 dataset, respectively, and showed that the partial face detection capabilities of these methods make them suitable candidates for mobile front-camera face detection.

The MIT Reality Dataset [9] consists of call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status (such as charging and idle) information from 100 Nokia-6600 smart phones users collected over 450,000 hours. Since it focused on analyzing social behavior of the subjects, it does not contain vital biometrics such as face and touch. The Rice Livelab dataset [38] consists of information on application usage, wifi networks, cell towers, GPS readings, battery usage and accelerometer output of 35 users, collected from iPhone 3GS devices over durations ranging from a few days to less than a year.

The largest known dataset on smartphone usage is the Google's Project Abacus data set consisting of 27.62 TB of smartphone sensor signals collected from approximately 1500 users for six months on Nexus 5 phones [27]. Data was collected for the front-facing camera, touchscreen and keyboard, gyroscope, accelerometer, magnetometer, ambient light sensor, GPS, Bluetooth, WiFi, cell antennae, app usage and on time statistics. Google also collected the 114GB Project Move data set, which consists of smartphone inertial signals collected from 80 volunteers over two months on $LG3$, $Nexus5$, and $Nexus6$ phones. The data collection was passive for both projects. To date, neither of these two datasets are available for the research community.

3. Description of the UMDAA-02 Dataset

The UMDAA-02 data set consists of 141.14 GB of smartphone sensor signals collected from 48 volunteers on Nexus 5 phones over a period of 2 months (15 Oct. 2015 to 20 Dec. 2015). The data collection sensors include the front-facing camera, touchscreen, gyroscope, accelerometer, magnetometer, light sensor, GPS, Bluetooth, WiFi, proximity sensor, temperature sensor and pressure sensor. The data collection application also stored the timing of screen lock and unlock events, start and end time stamps

Table 1. Significant Information for Each Modality Per Session

Modality	Information
Accelerometer	Event Time, X, Y, Z
Gyroscope	Event Time, X, Y, Z
Image	Shutter Time, Filename
Bluetooth	Developer, Paired/Unpaired Flag
Location	Event Time, Lat., Long., Accuracy
Usage	Event Time, % CPU, % Memory
Magnetic Field	Event Time, X, Y, Z
Gravity	Event Time, X, Y, Z
Connectivity	Capture Time, Flag (Bluetooth, Gps, Wifi, Cell Network), Network Name and Code
Foreground App Info	Start Time, Duration, End Time, App Name, Launched From Home Flag
WiFi	SSID, BSSID, Authentication Type, IP Address, RSSI
Ambient Light	Event Time, Value
Ambient Cells	MCC, CI, MNC, Sig. Strength, TAC
Screen	Event Time, Key
Motion/Touch	Event Time, Type, Pressure, Major-Minor Axis, Position
Call	Event Time, Key
Key	Event Time, Pressure, Type, Key Code
Screen Res	Event Time, X, Y

Table 2. Information on UMDAA-02 and UMDAA-02-FD Dataset

Description	UMDAA-02	UMDAA-02-FD
No. of Subjects	36M, 12F	34M, 10F
Age Range (years)	$22-31$	$22-31$
Avg. Days/User (days)	$\backsim 10$	$\backsim 10$
Avg. Sessions/User	$\backsim 248$	$\backsim 200$
Total Number of Images	600712	33209
No. of Images without Faces	–	9060
Avg. Images/User	$\backsim 12515$	$\backsim 755$
Avg. Images/Session	$\backsim 51$	$\backsim 4$
Min. no. of Image for a User	1038	64
Max. no. of Image for a User	49023	2787

of calls, currently running foreground application etc. The volunteers used the research phone as their primary device for a week and were given the option to stop data collection at will and review the stored data prior to sharing.

In Table 1, the most significant information for each modality associated with the sensor data is presented. Data for most of the modalities are stored when there is significant change in that modality. For example, the GPS data is stored at a rate proportional to the movement speed of the phone. The front camera images are captured only for the first 60 seconds for each session at a rate of 3 fps.

Some general information on the dataset is provided in Table 2. The usage information is arranged in 'Sessions' which starts when the user unlocks the phone and ends when the phone goes to the locked state. The data is stored in nested folders with the year, month, day and start time of the session embedded in the folder names.

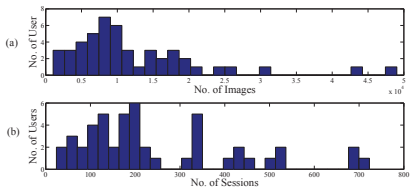

Figure 2. (a) Histogram of number of images per user, and (b) histogram of number of sessions per user.

Figure 3. Sample images from one of the users showing a wide variety of pose, illumination, occlusion and expression variations.

4. Face Detection and User Verification

In this section, we describe face detection and verification tasks from faces captured by the front-facing camera. Fig. 2 shows histograms of number of images per user and the number of sessions per user. The number of images varies between 2000 to $50,000$ per user and the number of sessions varies between 25 and 750, thus providing a large number of images for each user and session.

UMDAA-02-FD Face Detection Dataset: State-of-the-art face detection algorithms that perform satisfactorily on datasets like faces-in-the-wild [19], [20] are not suitable for detecting partially visible faces that are typically present in the UMDAA-02 dataset. Moreover, for practical implementation purposes, the algorithm must be very fast and have a high recall rate to ensure continuous authentication [23]. A few sample images are shown in Fig. 3 which shows that the faces suffer from partial visibility, illumination changes, occlusion and wide variation in poses and facial expressions.

Excluding the data of 5 users from a phone whose front camera malfunctioned during data collection phases, a set of 33209 images was selected from all sessions of the remaining 43 users at an interval of 7 seconds. The images were manually annotated for ground truth face bounding box, face orientation and five landmarks - left eye, right eye, nose, left and right corners of the mouth to create the UMDAA-02 face detection dataset (UMDAA-02-FD). Some information on the UMDAA-02-FD is provided in Table 2. The chronology and session information of all the images are also available. The histogram of face height and width distribution shown in Fig. 4 indicates that face widths vary approximately from 400 to 650 pixels, while face heights vary approximately from 300 to 700 pixels. The database contains many partial faces as can be seen from the extremities of the distribution, information from which can help tune the hyper-parameters of face detectors.

Evaluation of Face Detection Performances: Accuracy and F1-score measures are adopted as evaluation metrics

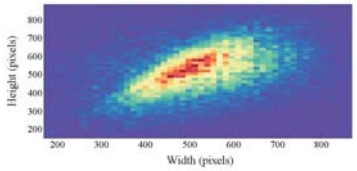

Figure 4. Distribution of bounding box width and heights

Table 3. Comparison between FD methods at 50% overlap

Method	Accuracy	F1-Score	Time/Image(s)
VJ [40]	60.24	64.50	**0.16**
DPM [43]	62.62	65.50	5.51
LAEO [24]	19.40	32.49	4.57
FSFD(C_{best})[23]	73.48	79.11	0.68
DP2MFD[31]	**76.15**	**82.83**	15.0(CPU), 0.8(GPU)

for face detection to ensure that both precision and recall performances are taken into consideration. The processing time per image is also measured to analyze the suitability for real-time operations. Prior to face detection, the images are down sampled by 4 to ensure reasonable processing time for all algorithms. 50% intersection-over-union overlap between the detection results and the ground truth bounding box is considered to be the threshold for correct detection.

The performances of four face detection algorithms on the UMDAA-02-FD dataset are presented in Table 3. The recently proposed Facial Segment-Based Face Detector (FSFD) algorithm [23] (with number of random subset $\zeta = 20$ and minimum number of segments $c = 2$), which is specifically designed for detecting partial faces, performs better than other popular non-commercial detectors like Viola-Jones (VJ) [40] and Deformable Part-based Model (DPM) [43] and in reasonable processing time. Another recent FD technique, the Deep Pyramid Deformable Part Model (DP2MFD) [31] utilizes normalized convolutional neural network (CNN) features. It outperforms all the other methods in terms of Accuracy and F1-Score but the processing time is quite long (almost 100 times more than VJ) thus making it unattractive for realtime implementation on smartphones. However, the best scores are far from satisfactory and better face detectors for AA are needed.

Face-based User Verification: Face verification is performed on the UMDAA-02-FD dataset. For each annotated face, 68 fiducial landmarks are extracted using the Local Deep Descriptor Regression (LDDR) method trained on Imagenet and FDDB datasets [22]. Feature extraction is performed after alignment, centering and cropping.

Feature Extraction from Faces: Given a face image, pixel intensity, Local Binary Pattern (LBP) [28] and Convolutional Neural Network (CNN) features using the pre-trained Alexnet network [21] and the DCNN network [4] are extracted. In total, 6 different features are extracted for each face as shown in fig. 5.

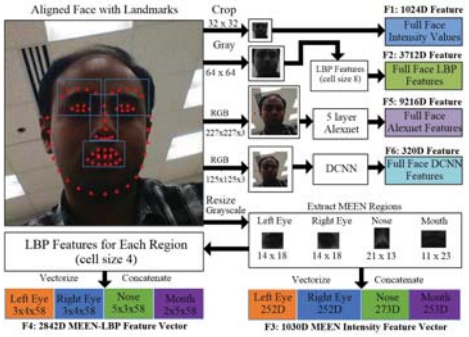

Figure 5. Flow diagram for features extraction for face verification.

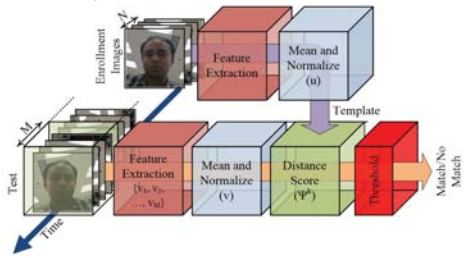

Figure 6. Block diagram depicting the face verification protocol.

- F_1: Pre-processed faces are converted to grayscale, rescaled (32×32) and vectorized (1024 dimensional vector).
- F_2: From the 64×64 rescaled grayscale image, LBP features of size $8 \times 8 \times 58$ (3712 dimensional vector) are extracted for a cell size of 8×8 pixels.
- F_3: Bounding boxes of the eyes, nose and mouth are computed from the landmarks with a 5 pixel margin for each face part from the pre-processed grayscale image. The eyes, nose and mouth bounding boxes are resized to 14×18, 21×13 and 11×23 pixels respectively, then vectorized to a 1030 dimensional MEEN feature[10].
- F_4: LBP features (2842 dimensional) are obtained from each of the resized bounding boxes of MEEN parts (F_3) with a cell size of 4×4 pixels.
- F_5: The first five convolutional layers of Alexnet are used to extract features of size $6 \times 6 \times 256$ (9216 dimensional) from resized color images of faces (227×227)
- F_6: Landmarks are input to the DCNN based face verification system [4] trained on the CASIA-WebFace dataset [41], which resizes the face to ($125 \times 125 \times 3$) and then outputs a 320 dimensional feature vector.

Evaluation Protocol: Six types of feature vectors are considered in this experiment. In the absence of any particular enrollment data, to simulate a practical AA scenario, the faces are sorted chronologically for each user and the first N faces are considered for enrollment while the rest are used for verification. The mean of the features of the enrollment set of a user followed by $L2$ normalization of the mean vector is stored as his/her template u.

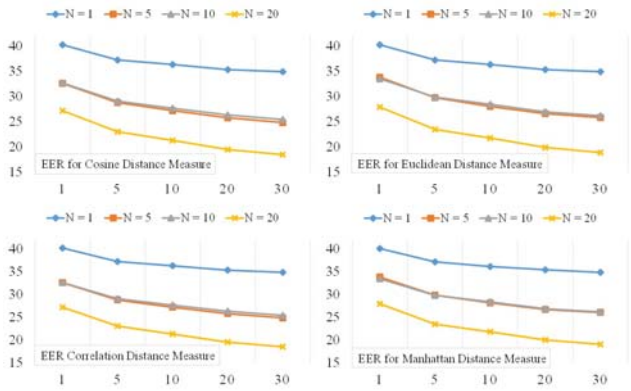

Figure 7. EER (%) vs. M for varying N using DCNN features (F_6) and four different metrics.

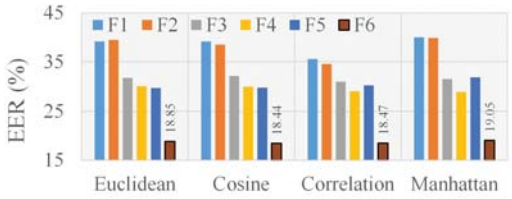

Figure 8. EER(%) for 6 feature vectors using four metrics.

Fig. 6 shows a block diagram of the verification process. A reasonable, practical assumption for robust AA is that the user is verified by the last M faces instead of a single one. Therefore features v_i $(i = 1, 2, \ldots, M)$ are extracted from each of the M faces for each location of the moving window, then averaged and $L2$-normalized to form the test vector v. The distances between v and u are calculated using four distance measures, namely, Euclidean Distance (EU), Cosine Distance (CosD), Manhattan Distance (MD) and Correlation Distance (CorrD). For the distance measure δ^k of type k the score is $\Psi^k = \frac{1}{\delta^k}$[34].

Experimental Results: In Fig. 7, the equal error rate (EER) (%) produced by using F_6 features are plotted for varying M and N values for the four distance measures. It is evident from the plots that the EER decreases with increasing N and M for all the cases. The lowest EER of 18.44% is achieved for $N = 20$, $M = 30$ using either CorrD or CosD measure.

Fig. 8 shows the EER corresponding to different features and distance measures considering $N = 20$, $M = 30$. The DCNN features (F_6) are found to be the most effective (EER of 18.44% for CosD). Since, for a reliable system the EER is expected to be at least less than 5%, this value is not satisfactory at all. The poor performance may be due to the fact that many faces in the dataset are partially visible and therefore alignment using facial landmarks fails badly for these cases. Also, matching the features from a partial face to the features of the same user's full face may result in a large distance measure. Among the other methods, the Alexnet network does not perform much better than the non-CNN features in this scenario as it is not trained particularly for faces. The LBP of MEEN face (EER of 28.83% for MD) gives the best result among non-CNN features. Note that in practice, the CNN feature extraction step is generally much slower than the non-CNN feature extraction methods without the use of a GPU. Thus, more robust yet fast verification methods are needed to produce satisfactory performance on this dataset.

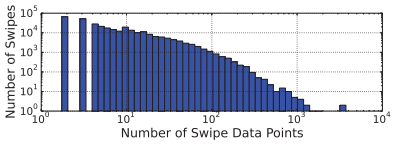

Figure 9. Histogram of the number of data points per swipe.

Table 4. General Information on Swipe Data

No. of subjects	48
Avg. Session/User with swipe data	$\backsim 196$
Total taps (finger down-finger up)	177417
Total swipes (including taps)	489723
Maximum data points in a swipe	3637
No. of Swipes/User	$\backsim 10203$
No. of Swipes/Session	$\backsim 52$
No. of Swipes (> 4 data points)	$\backsim 167126$
No. of Swipes/User (> 4 data points)	$\backsim 3482$
No. of Swipes/Session (> 4 data points)	$\backsim 18$

5. User Identification Using Swipe Dynamics

In this experiment, single finger touch sequences (swipes) on the screen are studied by considering three types of events - finger down, in-touch and finger up. The length of swipes vary between 1 to 3637 touch data points (Fig. 9). For reliable authentication using swipes, longer ones are preferable [13]. Hence, swipes with more than four data points are considered for feature extraction. Table 4 summarizes the swipe dataset, shows that it contains a large number of touch and swipe data per user and therefore can serve as a data set for practical experiments on swipe-based authentication. Since the users were not given any particular task to perform, the touch data in AA-02 is representative of how users interact with the phone through touch.

Feature Extraction: Every swipe s is encoded as a sequence of 4-tuples $s_i = (x_i, y_i, p_i, t_i)$ for $i \in 1, \ldots, N_c$ where x_i, y_i is the location coordinates and p_i is the pressure applied at time t_i. N_c is the number of data points captured during the swipe. From each swipe-action data with $N_c \geq 5$, a 24-dimensional feature vector, listed in Table 5, is extracted using the method described in [13] and [42]. Note, in the UMDAA-02 dataset, the measure of area covered by the finger is not present.

Experimental Setup and Evaluation: The swipe data for each user (with $N_c \geq 5$) are sorted chronologically and the first 70% swipes are considered for training-validation while the rest for testing. After extracting the 24-dimensional feature vector from each swipe, the train-

Table 5. Features Extracted From Each Swipe Event

Features	Description
1-2	inter-stroke time, stroke duration
3-6	start x, start y, stop x, stop y
7-8	direct end-to-end distance, mean resultant length
9	up/down/left/right flag
10-12	20%, 50%, 80% -perc. pairwise velocity
13-15	20%, 50%, 80%-perc. pairwise acc
16	median velocity at last 3 pts
17	largest deviation from end-to-end (e-e) line
18-20	20%, 50%, 80%-perc. dev. from e-e line
21	average direction
22	ratio of end-to-end dist and trajectory length
23	median acceleration at first 5 points
24	mid-stroke pressure

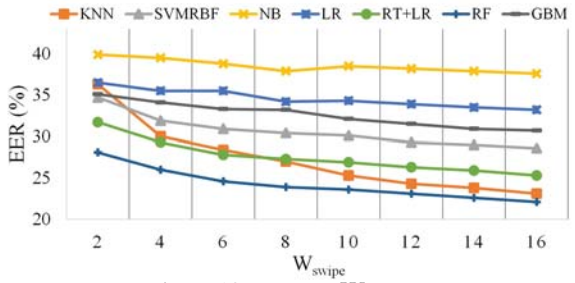

Figure 10. EER vs. W_{swipe}.

ing feature matrix is normalized to zero mean and unit variance. Then individual binary classifiers are trained for each user following the one-vs-all protocol. The classification methods considered for this experiment are k-nearest neighbor (KNN) [13], Gaussian kernel Support Vector Machine (RBF-SVM)[13], Naive Bayes (NB) [37], Linear Regression (LR) [37], Random Tree estimation followed by Linear Regression (RT+LR), Random Forest estimator (RF) [3], [11], [37] and Gradient Boosting Model (GBM) [15]. The methods are compared based on EER (%).

As proposed in [13], instead of using a single swipe for authentication, the scores of multiple, consecutive W_{swipes} number of swipes are averaged together for robustness. Since all of the methods return confidence probabilities/scores or distance from separating hyper-plane representing confidence, the score fusion is a simple average of individual scores. For the nearest neighbor-based methods, nine neighbors are considered. The parameters of RBF-SVM are tuned by 10 fold cross validation on smaller subsets of the original training data. Since the training data is very large, the SVM is trained on a reduced subset, followed by retraining on the hard negative mined error cases. For the ensemble-based methods, the number of estimators is set to 200 and the maximum tree depth is set to 10. The EER values obtained using different methods for differnet W_{swipe} values are show in Fig. 10. The random forest (RF) estimation method outperforms all the other methods and can reach an EER of 22.1%. However, for practical usage, this EER is not satisfactory and therefore achieving a better performance for this dataset is a new research challenge.

Table 6. General Information on Geo-location Data

No. of Subjects	45
Avg. No. of Sessions/User with Location Data	$\backsim$ 186
Total Number of Location Traces	8303813
Number of Location Traces Per User	$\backsim$ 184529
Number of Location Traces Per Session	$\backsim$ 993

Figure 11. Example of Geo-location Data Clustering - Analysis of the clusters reveal states of the user such as 'Home' or 'Work'.

6. Geo-location Data and Next Place Prediction

The location service of smartphones return geographical location of the user based on GPS and WiFi network. Excluding the users who kept their location service off, geolocation data, stored only if there is significant change in the location, is obtained from 45 users (summarized in Table 6). It is possible to reasonably predict the next location that a person might visit based on prior knowledge on the pattern on one's life. In this section, the next place prediction problem is approached using the geolocation data available in the UMDAA-02 dataset.

State Definition for Mobility Markov Chains: Location histories are first clustered into N_i clusters, namely $C_i^1 \ldots C_i^N$, for the i-th user using the DBSCAN algorithm [35] based on distances between data points. The maximum distance between a point from the center of the cluster in which that point belongs is set to be below a certain value R meters. Such clustering for a student (shown in Fig. 11) reveals the expected dominant regions that the user would visit - home, university, a certain shop and a restaurant. Two additional clusters, Transit (Tr) and Unknown (Unk), are also assigned for each user. If the user is traveling, causing location information to change rapidly ($\geq 2ms^{-1}$), then he/she is assigned to Tr. The remaining data points are denoted as Unk.

Data points at each cluster are assigned to six different observations based on the day and time information. Weekdays and weekend data points are flagged with WD and WE. Also, the whole day is divided into three time zones (TZs) - TZ1 (8:00 am to 4:00 pm), TZ2 (4:00 pm to 10:00 pm) and TZ3 (10:00 pm to 8:00 am). Thus, for the i-th user, there are $(N_i + 2) \times 2 \times 3$ possible observation states. However, since the location service only collects data when the phone is unlocked, there are many gaps in the data and it is possible that many of these observation states are absent in the training phase but present in the test data or vice-versa.

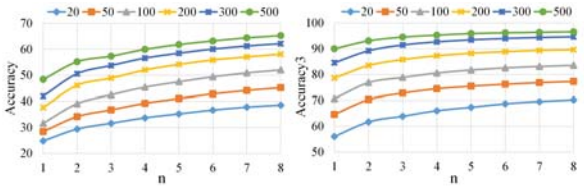

Figure 12. Next location prediction $Accuracy$ (left) and $Accuracy3$ (right) measures for increasing number of previous observations for MMC at different R.

The location service data is utilized for development and evaluation of Mobility Markov Chains (MMC) [17], [16] which is a discrete stochastic process model of the mobility behavior of an individual in which the probability of moving to a state depends only on the last visited state and the transition matrix for all probable states. Thus an MMC is composed of a set of k-states $S = s_1, s_2, \ldots, s_k$, prior probability of entering a state $p_1, p_2, \ldots, p_k$ and a set of transitions $t_{i,j}$ where $t_{i,j} = Prob(X_n = s_j | X_{n-1} = s_i)$.

Experimental Setup and Evaluation: From the chronological organization of a user's mobility traces, the first 70% are used for training while the rest for testing. Each trace of the training set is tagged with a unique tag identifying the state it belongs to. The prior and transition probabilities of each state are calculated from the chronological traces. Since, the number of states for a subject depends upon the maximum radius parameter R for the clusters, nearby small clusters get merged into bigger ones with increasing R causing a reduction in the number of states. In the training set, the average number of states per user drops to 35 from 144 if the maximum radius is increased to 500 m from 20 m.

MMC-based next location prediction results in terms of $Accuracy$ and $Accuracy3$ (percentage of times the correct next location was among the top 3 most probable locations) metrics are presented in Fig. 12. The horizontal axis represents the number of previous observations. Considering n previous observations, the MMC algorithm returns probabilities of each state to be the next. Since the day and time zone of the next location are known, states that do not belong to that day and time zone are dropped. The most probable state among the rest of the states is picked as the next predicted location. Fig. 12 indicates that knowing more prior states increases the accuracy. The accuracy also increases with increasing maximum radius R (from 20 meters to 500 meters) at the cost of localization capability. Between the two measures, $Accuracy3$ is can go much higher ($Accuracy = 65.3\%$ and $Accuracy3 = 96.6\%$ for $R = 500$ meters, $n = 8$) indicating the feasibility of location prediction.

7. Conclusion

In this paper, we presented a multi-modal challenge data set for AA problems. Benchmark results for face and touch-based active authentication are provided. Preliminary results for predicting the next location are also given. The UMDAA-02 is the first non-commercial data set on smart phone usage containing data form a wide variety of smart phone sensors. Thus this data set can provide sufficient resources to AA researchers to investigate the efficacy and performance of multi-modal fusion model for a wide variety of modalities in a practical AA scenario. The dataset will be released to the research community in due course.

References

[1] Smart phone thefts rose to 3.1 million in 2013. urlhttp://www.consumerreports.org/cro/news/2014/04/smart-phone-thefts-rose-to-3-1-million-last-year/index.htm, 2014.

[2] L. Araujo, J. Sucupira, L.H.R., M. Lizarraga, L. Ling, and J. Yabu-Uti. User authentication through typing biometrics features. *Sig. Process., IEEE Trans.*, 53(2):851–855, 2005.

[3] L. Breiman. Random forests. *Mach. Learning*, 45(1):5–32.

[4] J. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep CNN features. *CoRR*, abs/1508.01722, 2015.

[5] E. Chin, A. P. Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proc. 8th Symp. Usable Privacy and Security*, SOUPS, pages 1:1–1:16, 2012.

[6] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain. Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data. In *Int. Conf. on Biometrics (ICB)*, May 2015.

[7] I. G. Damousis and S. Argyropoulos. Four machine learning algorithms for biometrics fusion: A comparative study. *Applied Computational Intelligence and Soft Computing*, 2012(242401), 2012.

[8] M. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Intelligent Inform. Hiding and Multimedia Signal Process. (IIH-MSP), 2010 6th Int. Conf.*, pages 306–311, Oct. 2010.

[9] N. Eagle and A. (Sandy) Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, Mar. 2006.

[10] M. E. Fathy, V. M. Patel, and R. Chellappa. Face-based active authentication on mobile devices. In *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2015.

[11] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen. Continuous mobile authentication using touchscreen gestures. In *Homeland Security (HST), 2012 IEEE Conf. on Technologies for*, pages 451–456, Nov. 2012.

[12] I. T. Fischer, C. Kuo, L. Huang, and M. Frank. Smartphones: Not smart enough? In *Proc. of the 2nd ACM Workshop Security and Privacy in Smartphones and Mobile Devices*, SPSM, pages 27–32, 2012.

[13] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, Jan. 2013.

[14] L. Fridman, S. Weber, R. Greenstadt, and M. Kam. Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. *CoRR*, abs/1503.08479, 2015.

[15] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[16] S. Gambs, M.-O. Killijian, and M. N. n. del Prado Cortez. Show me how you move and i will tell you who you are. In *Proc. 3rd ACM SIGSPATIAL Int. Workshop on Security and Privacy in GIS and LBS*, SPRINGL '10, pages 34–41, 2010.

[17] S. Gambs, M.-O. Killijian, and M. N. n. del Prado Cortez. Next place prediction using mobility markov chains. In *Proc. the First Workshop on Measurement, Privacy, and Mobility*, MPM '12, pages 3:1–3:6, 2012.

[18] K. Halunen and A. Evesti. Context-aware systems and adaptive user authentication. In M. OGrady, H. Vahdat-Nejad, K.-H. Wolf, M. Dragone, J. Ye, C. Rcker, and G. OHare, editors, *Evolving Ambient Intelligence*, volume 413 of *Communications in Computer and Information Science*, pages 240–251. Springer International Publishing, 2013.

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, Univ. of Massachusetts, Amherst, Oct. 2007.

[20] M. Kostinger, P. Wohlhart, P. Roth, and H. Bischof. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *Comput. Vision Workshops (ICCV Workshops), 2011 IEEE Int. Conf.*, pages 2144–2151, Nov. 2011.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances Neural Inform. Process. Syst. 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[22] A. Kumar, R. Ranjan, V. M. Patel, and R. Chellappa. Face alignment by local deep descriptor regression. *CoRR*, abs/1601.07950, 2016.

[23] U. Mahbub, V. M. Patel, D. Chandra, B. Barbello, and R. Chellappa. Partial Face Detection for Continuous Authentication. *ArXiv e-prints*, 1603.09364, Mar. 2016.

[24] M. Marin-Jimenez, A. Zisserman, M. Eichner, and V. Ferrari. Detecting people looking at each other in videos. *Int. J. Comput. Vision*, 106(3):282–296, 2014.

[25] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J.-F. Bonastre, P. Tresadern, and T. Cootes. Bi-modal person recognition on a mobile phone: using mobile phone data. In *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, July 2012.

[26] M. Meeker and L. Wu. Kleiner perkins caufield and byers (kpcb): Internet trends, May 2013.

[27] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. W. Taylor. Learning human identity from motion patterns. *CoRR*, abs/1511.03908, 2015.

[28] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. and Mach. Intell.*, 24(7):971–987, July 2002.

[29] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, 33(4):49–61, July 2016.

[30] A. Primo, V. Phoha, R. Kumar, and A. Serwadda. Context-aware active authentication using smartphone accelerometer measurements. In *Comput. Vision and Pattern Recognition Workshops, IEEE Conf.*, pages 98–105, June 2014.

[31] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th Int. Conf.*, pages 1–8, 2015.

[32] O. Riva, C. Qin, K. Strauss, and D. Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *Proc. of the 21st USENIX Conf. on Security Symp.*, Security'12, pages 15–15, Berkeley, CA, USA, 2012. USENIX Association.

[33] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recogn. Lett.*, 24(13):2115–2125, Sept. 2003.

[34] P. Samangouei, V. M. Patel, and R. Chellappa. Attribute-based continuous user authentication on mobile devices. In *Int. Conf. Biometrics Theory, Applicat. and Syst. (BTAS), Arlington, VA*, Sept. 2015.

[35] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Min. Knowl. Discov.*, 2(2):169–194, June 1998.

[36] S. Sarkar, V. M. Patel, and R. Chellappa. Deep feature-based face detection on mobile devices. *ArXiv e-prints*, abs/1602.04868, 2016.

[37] A. Serwadda, V. V. Phoha, and Z. Wang. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. In *BTAS*, pages 1–8. IEEE, 2013.

[38] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum. Livelab: Measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Performance Evaluation Review*, 38(3):15–20, jan 2011.

[39] Verizon RISK Study. Data breach investigations report. url-http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf, 2013.

[40] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Comput. Vision and Pattern Recognition, CVPR 2001. Proc. IEEE Comput. Soc. Conf.*, volume 1, pages I–511–I–518 vol.1, 2001.

[41] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014.

[42] H. Zhang, V. M. Patel, M. E. Fathy, and R. Chellappa. Touch gesture-based active user authentication using dictionaries. In *IEEE Winter Conf. on Applicat. of Comput. Vision*, pages 207–214, Jan. 2015.

[43] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR)*, CVPR, pages 2879–2886, Washington, DC, USA, 2012. IEEE Computer Society.