Face-based Multiple User Active Authentication on Mobile Devices

Pramuditha Perera, *Student Member, IEEE* and Vishal M Patel, *Senior Member, IEEE*

***Abstract*—Multiple user active authentication, in contrast with single user active authentication, requires verification of identity of multiple subjects. Both traditional verification and identification based solutions fail to address the specific challenges presented in this problem. We introduce Extremal Openset Rejection (EOR), a two fold mechanism with a sparse representation-based identification step and a verification step for this purpose. In the verification step, concentration of the sparsity vector and the overlap between matched and non-matched distributions are considered for decision making. We introduce a semi-parametric model based on Extreme Value Theory (EVT) for modeling the distributions, and an algorithm to estimate the parameters of extreme value distributions. Effectiveness of the proposed method is demonstrated using three publicly available face-based mobile active authentication datasets.**

***Index Terms*—Open-set recognition, mobile biometrics, multiuser authentication, active authentication, continuous authentication.**

I. Introduction

The problem of person recognition has been traditionally studied under two main domains - identification and verification [2]. Given a probe, the role of identification is to find the best match from a gallery set. On the other hand, verification is typically performed between a probe and a single class, where the objective is to verify whether the probe belongs to the same class. Mobile Active Authentication (AA) [1] has recently received considerable interest in the research community as an alternative for traditional explicit authentication [4],[8],[16]. The purpose of AA is to continuously monitor biometric and behavioral traits of legitimate users of a device, and thereby perform authentication continuously. In exception of [23], previous works on AA have only addressed the case where only a single legitimate user is associated with the device. In this simplistic case, AA boils down to the problem of user verification.

In reality, mobile devices can be used by multiple users interchangeably [23]. For instance, different family members may want to share a mobile device. In this case, the number of such users generally varies between two to four; at most ranging to seven. This is also seen in workstations where a common device is often shared by different staff members. For instance, in hospitals or clinics, multiple medical professionals may want to share the same mobile device without switching between users. Therefore, for deployment of AA as a technology, it is essential that it has the capability to perform multiple user verification. However, in this specific use-case, producing the identity of a given probe is not required. Key characteristics of multiple user AA are as follows:

- **Low number of multiple users**: Number of enrolled users are usually low, typically between one to seven (i.e. normal family size). It is also possible to have just a single enrolled user.
- **Low focus on identity**: The goal is to verify whether the probe belongs to one of the enrolled users. Producing the exact identity of the user is not important.

With this background, when there exists multiple legitimate users for a mobile device, multiple AA problem can be formulated as follows. When the device is configured initially, each legitimate user is first enrolled into the device. Later on, in the course of normal operation of the device, sensor observations of the device user are continuously obtained. The goal of multiple AA is to determine whether the device user belongs to one of the enrolled subjects based on the obtained sensor observations.

Given the nature of the problem, it is tempting to treat multiple user AA as a single openset-identification task or as a series of verification tasks. However, this problem is unique in nature as it inherits different forms of challenges faced both in verification and identification. In verification, the class under consideration is compact due to relatively low variations of intra-class samples. Taking this fact into account, it is possible to generate a model that sufficiently

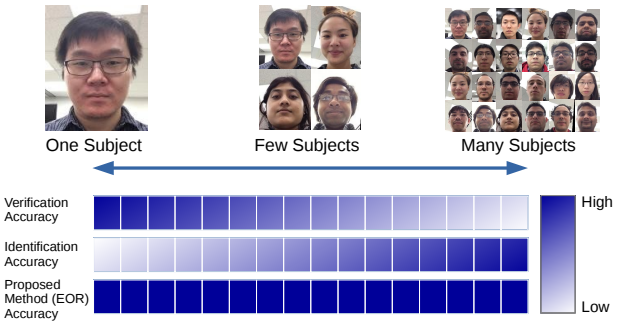


Fig. 1: Verification and identification algorithms are specifically designed to deal with single class and many classes, respectively. Therefore, their extensions fails to perform well consistently in multiple user AA where there are small number of classes. The proposed EOR method yields higher performance across the different protocols consistently.

Pramuditha Perera is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA. Email: pramuditha@live.com

Vishal M. Patel is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA. Email: vpatel36@jhu.edu.

[1]Active authentication is also known as continuous authentication, implicit authentication and transparent authentication [20].

encapsulates the features of the class under consideration. On the other hand, in identification, large number of classes are present. Therefore, for a given class, union of other classes produces a reasonable estimate of the open set negative classes. Hence, a discriminative model can be produced which yields identification with high accuracy. In contrast in multiple user AA, where generally 2 to 7 classes are present, there is a huge variation among the enrolled classes; this variation however is not representative enough to characterize the open world. Therefore, algorithms specifically designed for either verification or identification (or their extensions) do not comply well with the requirements of multiple user AA as the number of users vary (See Figure 1).

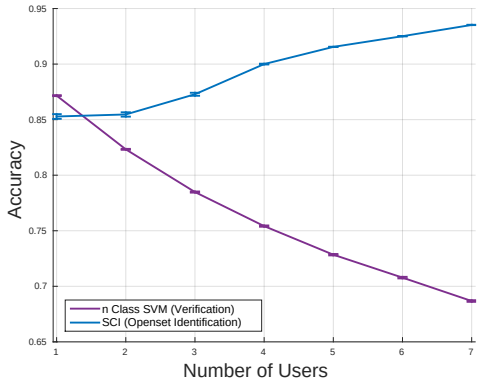


Fig. 2: Challenge of multiple user verification. Verification based algorithms are more suited when few classes are present. Identification based algorithms works better when large number of classes are present. Otherwise, there is a significant performance drop in both approaches. For the case with a few classes none of these methods are consistent in terms of accuracy.

Consider the scenario where n users are enrolled into a AA system. First let us consider a solution based on series of verifications. For each user it is possible to train a set of SVM classifiers $C_1, C_2, \ldots C_n$ to detect the presence of a legitimate user. When a probe $\mathbf{y}$ is present, using the maximum SVM classifier score $\max C_i(\mathbf{y})$, the presence of a legitimate user of the device can be queried via hard thresholding. However, as shown in Figure 2, the recognition accuracy of such a scheme drops rapidly as the number of users increase. Here, recognition accuracy for a single user had dropped from 87% to 82% with the introduction of an additional user. This would further drop down to 62% in the extreme case where seven users are using the device. Alternatively, it is possible to seek a solution to this problem based on an identification algorithm. If the identification algorithm associates scores $S_1, S_2, \ldots, S_n$ for each class with the given probe, similar to before, $\max S_i(\mathbf{y})$ can be used to determine whether the probe belongs to an intruder. With such an approach, the trend of detection accuracy seems to invert. Accuracy of the identification-based solution is initially low at 85% when a single user is present. This seems to increase as the number of users are increased. When the number of users increase, the number of known negative classes with respect to each user also increases. Since this allows better modeling of the negative population, intruder detection accuracy increases in identification-based methods. This empirical observation suggests that neither verification nor identification based methods seems to be a robust solution for multiple user AA. This outcome advocates the need for an alternative formulation to tackle the multiple user AA problem.

In this work, we motivate the need of a specialized method for performing multiple user AA. We present Extremal Openset Rejection (EOR), a sparse representation-based multiple user AA scheme based on semi-parametric extreme value analysis. In conjunction, we introduce a parameter selection scheme for extreme value distributions to make it feasible for an automated mechanism. Finally, we propose an extension to sparse representation-based recognition algorithm to make it feasible to be used even in single user applications.

II. Related Work

In recent years, AA has been studied based on a number of different biometric modalities including gait [32], [14] , keystroke, voice, swipe patterns [9], [27],[15], and face images [19], [4], [8],[16]. Most of these works encompass a single user scenario and therefore verification is performed using a generic verification algorithm. In [23], single user AA has been extended to the multiple user scenario. Here, authors have proposed a SVM based solution where scores of each SVM output is fused using a new fusion rule. In speaker recognition, the need to have multiple user systems have been previously discussed [18],[7]. In [22] multiple user authentication is formulated as a conjunction between a classification task and a verification task. This formulation is used as the basis for the proposed EOR multiple user AA algorithm.

In [29], sparse representation-based classification (SRC) was introduced for face recognition. This algorithm has been successfully adopted for mobile applications in previous works [17]. The motivation of SRC is to represent enrolled classes in terms of a dictionary and to query the expressiveness of each probe in terms of the constituents of the dictionary. Work in [31] has utilized the SRC algorithm to perform AA assuming the availability of some intruder classes at the time of testing. However, this assumption is impractical due to the privacy concerns. Even though not intended at AA, work in [29] introduces a method based on Sparsity Coefficient Index (SCI) to reject open set classes. A similar measure has been introduced in [21] for open set rejection. However, in their work they train n number of separate dictionaries when n number of users are enrolled. This is not a feasible practice for mobile AA due to the multiplied order of operations involved with each additional user.

The closest theoretical work related to multiple user authentication from an identification perspective is found in the open set recognition literature [12],[26],[25]. Here, the motivation is to perform identification in the presence of classes unseen during the testing phase (open set samples). The initial work in open-set recognition was the 1-vs-set machine introduced in [26]. This algorithm was later enhanced by PI-SVM [12] and WSM [25] algorithms. Both PI-SVM and WSM aim at

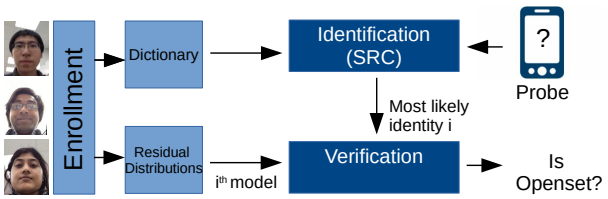


Fig. 3: An overview of the proposed EOR method.

modeling the score distributions of each enrolled class. The state of the art work in open set recognition presented in [30] out performed WSM by considering the matched and the sum of non matched score distributions. Proposed method in [30] models the tail portions of the said distributions using the statistical Extreme Value Theory (EVT). This approach yields good results only when the overlap between the distributions are restrained to the tail portions. When this is not the case, as in most AA applications, this method does not guarantee good detections.

III. Extremal Openset Rejection (EOR)

Using the work presented in [22] as a conceptual platform, we propose a two fold process to solve the multiple verification problem as shown in Figure 3. This two-fold process contains an identification step and a verification step. In the proposed system, a set of images of each legitimate user of the device is obtained during an enrollment phase and a dictionary is constructed. In addition, matched and non-matched residual score distributions are constructed for each user. Given the enrolled classes, the identification process determines the most likely identity of a probe based on the SRC model using the enrolled dictionary. The verification step determines whether the estimated identity of the probe is true or not based on corresponding matched and non-matched residual distributions. If the estimated identity is found to be true at this step, authenticity of the user is declared to be true.

The proposed method essentially utilizes matched and non-matched distribution information on top of the identification criterion to make a better decision. This additional processing has a significant gain particularly when identification criterion is poor (that is when a low number of users are enrolled). If a large number of classes are present, the additional verification step does not introduce a significant improvement. EOR is expected to perform on par with the identification method in such scenarios. Therefore, the proposed framework is particularly suited for multiple-user authentication problems.

In what follows, a detailed explanation of this process is presented.

A. Enrollment Phase

Consider the case where n number of legitimate users are identified for a device. Given a set of k enrolled images $\mathbf{V}_i = [\mathbf{v}_{1,i}, \mathbf{v}_{2,i}, \ldots, \mathbf{v}_{k,i}]$, where $\mathbf{v}_{j,i} \in \mathbb{R}^m$ for each ith user, a dictionary $\mathbf{D}$ can be constructed by stacking images of all users as columns (atoms) as in, $\mathbf{D} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n] \in \mathbb{R}^{m \times nk}$. If $\mathbf{D}$ is sufficiently expressive, a new image $\mathbf{y}$ belonging to the ith enrolled user can be expressed as a linear combination of columns of $\mathbf{D}$ as in, $\mathbf{y} = \mathbf{D}\mathbf{x}$. Here, vector $\mathbf{x}$ will be predominantly zero except for the entries associated with the ith column block of $\mathbf{D}$. For a given validation image $\mathbf{y}_v$, the corresponding sparse representation vector $\mathbf{x}_v$ can be found by solving the following optimization problem [29],

$$\hat{\mathbf{x}}_v = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}_v - \mathbf{D}\mathbf{x}\|_2 < \epsilon, \tag{1}$$

where ϵ is the noise energy and $\|\mathbf{x}\|_1$ denotes the ℓ_1-norm of $\mathbf{x}$ defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. If δ_i is the characteristic function that selects the coefficients of $\mathbf{x}_v$ that are only associated with the ith class, then the residual error of representation of the ith class can be obtained as

$$r_i(\mathbf{y}_v) = \|\mathbf{y}_v - \mathbf{D}\delta_i(\hat{\mathbf{x}}_v)\|_2, \quad i = 1, 2, , \ldots, n. \tag{2}$$

Due to the self-expressiveness property of SRC, an image belonging to the ith class is likely to be represented well in terms of the atoms of the dictionary. Therefore, the residual $r_i(\mathbf{y}_v)$ of such a probe is likely to take a lower value. Based on the obtained residual errors, the most likely class of the probe $\mathbf{y}_v$ can be determined using,

$$\hat{i} = \arg\min_i r_i(\mathbf{y}_v). \tag{3}$$

For a given probe $\mathbf{y}_v$, the classifier results in a single matched residual $r_{\hat{i}}(\mathbf{y}_v)$. In addition, the classifier results in $n-1$ non-matched scores $r_j(\mathbf{y}_v)$, where $j \neq \hat{i}$. Using a set of sample images, matched and non-matched residual scores can be obtained for each class accordingly. Let $F_{M,i}$ and $F_{N,i}$ be cumulative distributions of the matched and non-matched residuals obtained for the i^{th} class.

B. Classification Phase

During testing, for a given probe $\mathbf{y}_p$, the most corresponding sparse vector can be found according to (1). The obtained sparse vector can be used to evaluate the residual error associated to the each class using (2). Then, the most likely class associated with the probe can be obtained using (3). A probe that belongs to one of the enrolled classes is expected to produce a relatively low residual for the corresponding class. In such a scenario, (3) would provide the identity of the probe.

However, even when a probe is from a class outside the enrolled classes (i.e. from an intruder), there will be a lowest residual value and a corresponding most likely class. In [29], Sparsity Coefficient Index (SCI) defined as,

$$\text{SCI}(\mathbf{x}_p) = \frac{n \times \max_i \|\delta_i(\mathbf{x}_p)\|_1 / \|\mathbf{x}_p\|_1 - 1}{n-1} \tag{4}$$

is introduced to differentiate such intruder classes from the enrolled classes. However, as will be shown later, this measure performs poorly when the number of enrolled classes are small. In what follows, we propose a more robust, alternative method for this purpose.

C. Verification Phase

Upon determination of the most likely class of a probe, verification is carried out to confirm the identity obtained in SRC. If the probe $\mathbf{y_p}$ belongs to an intruder, verification process is likely to fail and an intrusion will be declared. On the other hand, if the probe belongs to the identity declared by the SRC algorithm, verification is likely to be successful.

If the probe belongs to class d, resulting residual score(r) should satisfy two conditions simultaneously - it should be from the matched score region $\mathcal{M}_d$ and it should not fall on to the non-matched score region $\mathcal{N}_d$ of class d. Mathematically, probability of probe belonging to class d can be formulated as,

$$P(d) = P(r \in \mathcal{M}_d \cap r \notin \mathcal{N}_d).$$

Using addition rule of sets, this expression can be expanded as,

$$P(d) = P(r \in \mathcal{M}_d) + P(r \notin \mathcal{M}_d) - P(r \in \mathcal{N}_d \cup r \notin \mathcal{N}_d).$$

Since residual scores take lower values for the matched probes, $\mathcal{M}_d$ can be expressed as $F_{M,d}(r)$ [13]. In contrary, $\mathcal{N}_d$ can be obtained from $1 - F_{N,d}(r)$, since non-matched scores take larger values. Here, $F_{M,d}(.)$ and $F_{N,d}(.)$ are the matched and non matched cumulative residual score distributions, respectively. In addition, $P(r \in \mathcal{M}_d \cup r \notin \mathcal{M}_d) = 1$. Therefore,

$$P(d) = F_{M,d}(r) + (1 - F_{N,d}(r)) - 1, \tag{5}$$

and,

$$P(d) = F_{M,d}(r) - F_{N,d}(r), \tag{6}$$

where $F_{M,d}(.)$ and $F_{N,d}(.)$ are the cumulative distributions of the matched and non-matched residual scores for the dth class, respectively. Hence, $P(d)$ can be used as an indicator to determine the authenticity of the probe.

In addition, we use the coefficient concentration of the sparsity vector as another indicator to aid decision making. When the probe belongs to one of the legitimate users, sparsity coefficients are likely to be concentrated on to a single class. In contrary, for an intruder, the coefficients are likely to spread across various classes. This reasoning was motivated by the rationale behind the SCI method. Therefore, we propose $\|\delta_i(\hat{\mathbf{x}_p})\|_2 / \|\hat{\mathbf{x}}_p\|_2$ as an intrusion indicator $P(I)$, where $\hat{\mathbf{x}_p}$ is the corresponding sparse vector. However, the discriminative power of the latter is only high when there are larger number of enrolled classes present. Therefore, we use a linear fusion scheme, which we call EOR, to fuse the two indicators to obtain a matched score for class d based on the residual score r with variable weights as in,

$$EOR = (1 - w)P(I) + wP(d)$$

$$EOR = (1 - w)\frac{\|\delta_d(\hat{\mathbf{x}}_p)\|_2}{\|\hat{\mathbf{x}}_p\|_2} + w(F_{M,d}(r) - F_{N,d}(r)), \tag{7}$$

where $w = \exp(-\alpha n^2)$ with α being a constant and n being the number of enrolled users. The definition for w was selected to ensure the weights are always below one. The α value needs to be determined empirically. In our experiments, it was set equal to 0.2. The score obtained for EOR, is thresholded to arrive at a detection decision.

D. Modeling of Cumulative Distribution Functions

For the proposed method to yield results of higher accuracy, it is essential that the models for matched and non-matched residual distributions are of higher precession. The trivial choice would be to use an empirical distribution obtained through a kernel density estimation process.

However, due to low occurrences of extreme observations, sufficient number of extremal samples are often not present to accurately model the tail portion of distributions empirically [6]. This issue is more pronounced in intruder detection problems such as in AA, where the tail region contains vital information that can be used to reject intruders. Therefore, modeling the tail region using a better technique is necessary. In this light, we propose the use of EVT to model the tail portion of each distribution.

Picklands, Balkema and de Haan Theorem: Consider a set of samples from Z that exceeds a sufficiently high threshold u. If F_U is the cumulative distribution of the excess of Z over u such that

$$F_U(z) = P(Z - u \leq z | Z > u), \tag{8}$$

where $z > 0$, the CDF of F_U can be approximated using a Generalized Pareto Distribution (GPD)

$$G(z; \xi, \beta) = \begin{cases} 1 - (1 + \frac{\xi z}{\beta})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-z/\beta} & \text{if } \xi = 0, \end{cases} \tag{9}$$

such that $-\infty < \xi < \infty, 0 < \beta < \infty, x > 0$ and $\xi x > -\beta$ [3], [10], [1]. When $\xi = 0$, GPD reduces to an exponential distribution with mean β. When $\xi > 0$ and $\xi < 0$, the GPD takes the form of an ordinary Pareto distribution and a Pareto II distribution, respectively. This result of EVT provides us a way to model the tail portion of a distribution given a sufficiently high threshold u.

Semi-Parametric Modeling of Distribution: Given a set of observations, the empirical cumulative distribution function of the data $H(z)$ can be evaluated using Kernel density estimation. For a given parameter u, the tail portion of the distribution (values above u) can be modeled using GPD to obtain $G(z|\xi, \beta)$. Then, the following semi-parametric model can be used to represent the cumulative density function,

$$F(z; \xi, \beta, \phi_u) = \begin{cases} (1 - \phi_u) + \phi_u G(z; \xi, \beta), & \text{if } z > u \\ \frac{(1-\phi_u)}{H(u)} H(z), & \text{if } z \leq u. \end{cases} \tag{10}$$

It was shown in [6], that the maximum likelihood estimate of ϕ_u is the sample proportion of excess where $\hat{\phi} = t_u / t$. Here, t_u is the number of in excess of u and t is the total sample size. With this formulation, the CDF of the matched and the non-matched distributions of all considered classes can be modeled.

Parameter Selection for GPD: Even though extreme value analysis is not new to computer vision applications [28],[10],[30], to date, selecting the parameter u has been

done experimentally. Since this parameter is specific to the distribution, selecting a universal parameter for all the distributions is not reasonable. Therefore, we propose a mean squared error-based approach to select the parameter u based on Hall's bootstrap method [11]. Here, for a series of candidate values of u, the tail portion of the distribution is extracted and modeled using a GPD. Then, a series of random values are generated using the estimated semi-parametric distribution. A second GPD is fitted to the tail portion of the obtained random variables and its GPD parameters are compared with the original GPD to arrive at a mean squared error (MSE) figure. The value u that generated the lowest MSE is selected as the appropriate threshold. This parameter selection method is outlined in Algorithm 1.

input : Set of training data S of size t
output: GPD distribution threshold parameter u

Fit empirical distribution to data;
CDF_empirical = fitDist(S);
for j *in* $[0.6 : 0.05 : 0.95]$ **do**
  Calculate Threshold for the percentile j ;
  S = Sort(S);
  $T = S[tj]$;
  Select entries over the threshold T;
  $E = \{S : S > T\}$;
  [ξ,β] = GPDfit(E) ;
  Perform bootstrapping;
  for $i \leftarrow 1$ **to** 10 **do**
    Generate random numbers from the estimated parameters;
    $R[1 : 1000j]$= RandGPD(ξ,β);
    $R[1000j : 1000]$= Rand(CDF_empirical) ;
    Estimate GPD parameters from random numbers;
    $A = \{R : R > T\}$;
    [$\hat{\xi}[i]$,$\hat{\beta}[i]$] = GPDfit(A) ;
  end
  MSE = $\mathbb{E}[(\xi - \hat{\xi})^2]$;
end
u = argmin$_T$(MSE);

Algorithm 1: Threshold Selection for GPD.

Extending to the Case of a Single User: In the outlined two-fold procedure, initial classification step uses the SRC algorithm to recover the most likely identity of the probe. When only a single enrolled user is present, this step is not necessary as the most likely identity is trivial. However, with only a single user, the outlined verification step is infeasible to be carried out due to unavailability of a non-match residual distribution. For this particular case we propose the following modifications.

1) For the legitimate user, create dictionary $\mathbf{D} \in \mathbb{R}^{m \times k}$ as explained in Section III-A. Append a set of random numbers of size $m \times k$ to the dictionary. This set of random numbers is expected to represent openset classes.
2) In the classification process, calculate the residuals for the two dictionaries as before. However, most likely identity now is always associated with the only enrolled user.
3) Residual error corresponding to the random dictionary is used to calculate the non-match residual score. Verification process is carried out as outlined in Section III-C.

Using a set of random entries to represent the open-set class subjects is motivated by the result obtained in Section IV-G. There, we empriricaly show that non-matched distrubution generated by an arbitary user has a high probability of been similar to the true non-matched distribution with respect to a given subject. This result however is only justified when the SRC matching rule is used.

IV. Experimental Results

In order to investigate the effectiveness of the proposed method, experiments were conducted using three publicly available mobile face datasets. Seven test cases were considered where the number of users were varied from one to seven. For each test case, 50 trials of experiments were considered. In a single testing trial, legitimate users and 10 intruder users were randomly picked. Images of selected users were divided in half to form a training and a test set. Training set was used to construct the dictionary and to obtain matched residual scores. Dictionary size was kept fixed at 50 atoms per subject. During testing, equal number of positive and negative images were used to prevent accuracy paradox. Detection accuracy defined as

$$\text{Detection Accuracy} = \frac{TP + TN}{2}, \tag{11}$$

where TP and TN are true positive rate and true negative rate, respectively, is used to measure the performance of different methods. When evaluating detection accuracy, we considered all possible thresholds and selected the highest resulting detection accuracy for comparison.

The following methods were used for comparison:

1) n class SVM : n binary SVMs used in parallel (Implemented using LibSVM)
2) PISVM [12] : Used the published code for [12].
3) WSVM [25]: Used the published code for [25].
4) Calibrated SVM : Operating point of an n class SVM been calibrated using matched and non matched distributions.
5) MUSVM : State of the art multiple user AA method based on binary SVM [23].
6) MUWSVM : State of the art multiple user AA method based on binary WSVM [23].
7) SCI : Intruder rejection based on sparsity distribution of the sparsity vector as described in (4) [29].
8) SROSR : Intruder rejection based on matched and sum of non matched distributions [30].

In implementing these methods, whenever selection of a parameter was necessary, it was done using a validation set. We considered 10% of the legitimate user images when constructing the validation set. In addition, we used images from the remaining users (users excluding legitimate and

(a) (b) (c)

Fig. 4: Sample images from datasets used for testing. (a) UMDAA-01 (b) MOBIO (c) UMDAA-02. Each column represents sample images obtained for the same user.

	UMDAA-01						
	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.73 (0.0327)	0.74 (0.0342)	0.73 (0.0340)	0.72 (0.0340)	0.72 (0.0341)	0.71 (0.0350)	0.71 (0.0347)
WSVM	0.80 (0.0309)	0.77 (0.0319)	0.74 (0.0328)	0.73 (0.0336)	0.72 (0.0339)	0.72 (0.0344)	0.72 (0.0346)
n-Class SVM	0.87 (0.0204)	0.82 (0.0204)	0.78 (0.0223)	0.75 (0.0238)	0.73 (0.0252)	0.71 (0.266)	0.69 (0.272)
C-SVM	0.88 (0.0207)	0.83 (0.0258)	0.78 (0.0291)	0.74 (0.0313)	0.71 (0.0333)	0.69 (0.0337)	0.67 (0.0335)
MUSVM	0.88 (0.0212)	0.84 (0.0273)	0.82 (0.0371)	0.80 (0.0404)	0.78 (0.0340)	0.76 (0.0447)	0.75 (0.0477)
MUWSVM	0.88 (0.0215)	0.84 (0.0274)	0.81 (0.0369)	0.80 (0.0402)	0.78 (0.0335)	0.77 (0.0448)	0.76 (0.0480)
SCI	0.85 (0.0461)	0.85 (0.0441)	0.87 (0.0367)	0.90 (0.0187)	0.92 (0.0118)	0.93 (0.0144)	0.94 (0.0125)
SROSR	0.88 (0.0505)	0.92 (0.0294)	0.90 (0.0350)	0.90 (0.0279)	0.90 (0.0256)	0.89 (0.0279)	0.91 (0.0257)
EOR	**0.90** (0.0303)	**0.93** (0.0249)	**0.92** (0.0202)	**0.93** (0.0121)	**0.93** (0.0102)	**0.94** (0.0104)	**0.94** (0.0093)

TABLE I: Variation of mean detection accuracy against different number of users for UMDAA01 dataset (standard deviation in brackets).

	MOBIO						
	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.75 (0.017)	0.78 (0.0051)	0.76 (0.0126)	0.76 (0.0164)	0.76 (0.0164)	0.73 (0.0112)	0.73 (0.0137)
WSVM	0.77 (0.017)	0.79 (0.0051)	0.78 (0.0126)	0.77 (0.0163)	0.77 (0.0161)	0.74 (0.0111)	0.73 (0.0137)
n-Class SVM	0.94 (0.0192)	0.88 (0.0151)	0.87 (0.0130)	0.83 (0.0359)	0.80 (0.0101)	0.80 (0.0189)	0.77 (0.0741)
C-SVM	0.87 (0.0262)	0.79 (0.0806)	0.77 (0.0162)	0.71 (0.0270)	0.68 (0.0196)	0.68 (0.0212)	0.63 (0.0418)
MUSVM	**0.97** (0.0171)	0.92 (0.0129)	0.91 (0.0146)	0.87 (0.0148)	0.84 (0.0161)	0.85 (0.0230)	0.80 (0.0178)
MUWSVM	0.96 (0.0132)	0.90 (0.0125)	0.89 (0.0111)	0.83 (0.0224)	0.81 (0.0123)	0.83 (0.0277)	0.76 (0.0182)
SCI	0.80 (0.0697)	0.87 (0.0405)	0.91 (0.0294)	0.92 (0.0319)	0.93 (0.0214)	0.93 (0.0161)	0.95 (0.0134)
SROSR	0.90 (0.0529)	0.93 (0.0372)	0.94 (0.0279)	0.94 (0.0216)	0.93 (0.0250)	0.93 (0.0267)	0.94 (0.0153)
EOR	0.90 (0.0441)	**0.95** (0.0232)	**0.96** (0.0149)	**0.96** (0.0119)	**0.97** (0.0169)	**0.96** (0.0092)	**0.97** (0.0069)

TABLE II: Variation of mean detection accuracy against different number of users for MOBIO dataset (standard deviation in brackets).

intruders) to simulate the negative classes in the validation set.For SROSR, threshold of the distribution was selected using the procedure described in Algorithm 1. For the calibrated SVM, the operating threshold T was selected according to $T = \arg\max_x |F_m(x) - F_n(x)|$. For all datasets, PCA was performed on raw images and 50 dimensions were recorded to arrive at dictionaries.

Due to the high processing involved with deep-feature extraction, deep-feature based authentication methods are impracticable to be used in continuous processes such as Active Authentication. Therefore, state-of-the art works in AA [24] has used hand crafted features to perform authentication. Following this practice, we used dimensionally reduced raw pixel intensities as the feature in our method. However, deep-features can be easily incorporate in to our framework by using deep-features to construct dictionaries.

	UMDAA-02						
	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.53 (0.0115)	0.52 (0.0480)	0.51 (0.0052)	0.48 (0.0052)	0.50 (0.0275)	0.50 (0.0385)	0.50 (0.0385)
WSVM	0.57 (0.0207)	0.53 (0.0480)	0.52 (0.0052)	0.49 (0.0052)	0.50 (0.0275)	0.50 (0.0384)	0.50 (0.0384)
n-Class SVM	0.70 (0.0204)	0.65 (0.0356)	0.64 (0.0088)	0.63 (0.0088)	0.60 (0.0094)	0.58 (0.0100)	0.55 (0.0101)
C-SVM	0.70 (0.0456)	0.65 (0.0513)	0.64 (0.0260)	0.64 (0.0260)	0.61 (0.0262)	0.57 (0.0264)	0.54 (0.0265)
MUSVM	0.70 (0.0210)	0.67 (0.0340)	0.66 (0.0072)	0.64 (0.0076)	0.63 (0.0080)	0.61 (0.0092)	0.58 (0.0097)
MUWSVM	0.70 (0.0212)	0.68 (0.0345)	0.66 (0.0076)	0.66 (0.0080)	0.65 (0.0082)	0.63 (0.0095)	0.60 (0.0102)
SCI	0.66 (0.0469)	0.66 (0.0357)	0.67 (0.0340)	0.66 (0.0268)	0.66 (0.0234)	0.68 (0.0232)	0.68 (0.0186)
SROSR	0.63 (0.0542)	0.61 (0.0485)	0.62 (0.0500)	0.59 (0.0388)	0.56 (0.0258)	0.59 (0.0328)	0.58 (0.0230)
EOR	**0.71** (0.0420)	**0.72** (0.0347)	**0.73** (0.0340)	**0.70** (0.0308)	**0.69** (0.0252)	**0.70** (0.0205)	**0.70** (0.0188)

TABLE III: Variation of mean detection accuracy against different number of users for UMDAA02 dataset (standard deviation in brackets).

A. *UMDAA-01 Face Dataset*

The UMDAA-01 dataset [8] contains images captured using the front-facing camera of a iPhone 5S mobile device of 50 different individuals captured across three sessions with varying illumination conditions. Images of this dataset contain pose variations, occlusions, partial clippings as well as natural facial expressions as evident from the sample images shown in Figure 4(a) . For the SVM-based methods, facial attributes of the images were used [24] as features with the RBF kernel.

Shown in Figure 5(a), is the variation of accuracy of different methods for varying number of enrolled users. Eventhough SVM-based verification schemes perform reasonably well for the case of a single user, the performance drops significantly as the number of users are increased. On contrary, the accuracy of SCI is relatively lower for a single user, but increases as the number of users increase. This is expected since the

sparsity coefficients are likely to spread across all classes for openset images causing SCI to decrease as the number of classes increase. EOR yields the best performance for all considered cases. The results corresponding to this experiment are tabulated in Table I.

B. MOBIO Face Dataset

The MOBIO dataset [19] contains videos of 152 subjects taken across two phases where each phase consists of six sessions. Videos in this dataset are acquired using a standard 2008 Macbook laptop computer and a NOKIA N93i mobile phone. For the experiments on the MOBIO dataset, only videos captured using the mobile device were used as shown in Figure 4(b). For the SVM-based methods, facial attributes of each video frame were extracted using the method proposed in [24] and an SVM with the RBF kernel was used for pairwise matching.

Variation of detection accuracy against the number of users for different methods are shown in Figure 5(b) and are tabulated in Table II. Due to the less challenging nature of the dataset compared to the UMDAA-01 dataset, all the methods yielded comparatively higher accuracies on this dataset. MUWSVM method produces the best accuracy among all tested cases for the single user case. But the same as in UMDAA-01, the performance dips as the number of users increase. However, it is not as rapid as in the earlier case. In the case of SCI, the performance increases from 0.8 to 0.95 as the number of users increases from one to seven. However, the proposed method, produces the highest accuracy consistently over 0.95 across the range of different number of users.

C. UMDAA-02 Face Dataset

The UMDAA-02 Dataset [16] is an unconstrained multimodal dataset where 18 sensor observations were recorded across two month period using a Nexus 5 mobile device. Unlike the earlier datasets, there exists a huge intra-class variations in this dataset in terms of poses, partial faces, illumination as well as appearances of the users as evident from the sample images shown in Figure 4(c). For the SVM-based methods, the HOG feature [5] was used with an RBF kernel.

Detection accuracy corresponding to this dataset is plotted against the number of enrolled users in Figure 5(c). The results corresponding to this experiment are tabulated in Table III. Unlike the other two datasets, maximum accuracy for this dataset does not go beyond 0.75 due to the high intra-class variations present in the dataset. This result is reasonable as the benchmark results based on the Alexnet features and the RBF SVM has only recorded an EER of around 0.4 [16]. Even with comparatively lower accuracy figures, the trend of earlier datasets can be observed in this dataset. The SCI method yielded an accuracy of 0.66 for the single user case. That figure increased to 0.68 as the number of users were increased to seven. The proposed method produces an accuracy of 0.71 for the single user case. This figure stays more or less consistent over all considered cases.

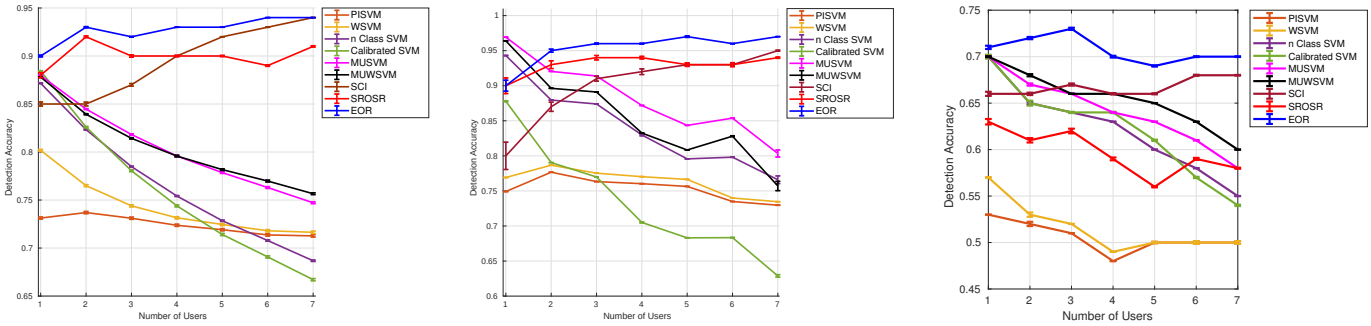
		1	2	3	4	5	6	7
PISVM	ACC	0.73	0.74	0.73	0.72	0.72	0.71	0.71
	TP	0.72	0.68	0.63	0.61	0.60	0.61	0.62
	TN	0.74	0.80	0.83	0.83	0.84	0.81	0.80
WSVM	ACC	0.80	0.77	0.74	0.73	0.72	0.72	0.72
	TP	0.71	0.67	0.64	0.61	0.61	0.60	0.62
	TN	0.89	0.87	0.84	0.85	0.83	0.84	0.82
n-Class SVM	ACC	0.87	0.82	0.78	0.75	0.73	0.71	0.69
	TP	0.88	0.88	0.89	0.90	0.94	0.97	0.99
	TN	0.86	0.76	0.67	0.60	0.52	0.45	0.39
C-SVM	ACC	0.88	0.83	0.78	0.74	0.71	0.69	0.67
	TP	0.92	0.94	0.94	0.95	0.95	0.96	0.96
	TN	0.84	0.72	0.62	0.53	0.47	0.42	0.38
MUSVM	ACC	0.88	0.84	0.82	0.80	0.78	0.76	0.75
	TP	0.86	0.83	0.82	0.82	0.83	0.84	0.85
	TN	0.91	0.85	0.82	0.78	0.73	0.68	0.65
MUWSVM	ACC	0.88	0.84	0.81	0.80	0.78	0.77	0.76
	TP	0.86	0.83	0.85	0.81	0.82	0.84	0.86
	TN	0.91	0.86	0.78	0.79	0.74	0.70	0.66
SCI	ACC	0.85	0.85	0.87	0.90	0.92	0.93	0.94
	TP	0.88	0.87	0.89	0.93	0.93	0.91	0.91
	TN	0.82	0.83	0.85	0.87	0.91	0.95	0.97
SROSR	ACC	0.88	0.92	0.90	0.90	0.90	0.89	0.91
	TP	0.92	0.97	0.91	0.88	0.89	0.86	0.89
	TN	0.84	0.87	0.89	0.92	0.91	0.92	0.93
EOR	ACC	0.90	0.93	0.92	0.93	0.93	0.94	0.94
	TP	0.92	0.95	0.93	0.96	0.95	0.96	0.95
	TN	0.88	0.91	0.91	0.90	0.91	0.92	0.93

TABLE IV: Breakdown of the mean detection accuracy in the UMDAA01 dataset corresponding to different number of enrolled users.

D. Breakdown of the Mean Accuracy

It is the usual practice to measure the performance of an AA system based on the mean detection accuracy. Nevertheless, it is interesting to investigate the breakdown of the mean detection accuracy (ACC) into true positive accuracy (TP) and true negative (TN) accuracies. In Table IV, we present this breakdown for the UMDAA01 dataset.

According to Table IV, both n-class SVM and c-SVM are able to detect own class samples with high accuracy when the number of users are increased. But in doing so, they have produced higher number of false positives as well. Since the SVM-based methods do not have any information about outside class samples, it is possible that they have opted for thresholds such that majority of the positive samples are accepted. Doing so increases the volume of the positive space as the number of users increase - thereby increasing the false positive rate. Comparatively, in openset methods (PISVM and WSVM), where the models are developed with the objective of reducing the open-space risk, both true positive rate and true negative rate decrease when the number of users are increased. In the MUSVM and MUWSVM algorithms, that are proposed specifically for the multiple AA problem, true positive rate stays more or less constant when the number of users are increased. In both SCI and SROSR, the true negative accuracy has increased with the number of users. In both of these methods, having more users improve the ability of the model to reject an unknown sample. For example, in SCI, when more users are enrolled, it is likely that sparsity coefficients of an unknown person are very well spread out across the enrolled users. However in SROSR, true positive rate drops slightly in the process. Similar to SCI, in EOR,



(a) UMDAA01 (b) MOBIO (c) UMDAA02

Fig. 5: Experimental results comparison based on mean detection accuracy. Variance of observed results are denoted using the error-bar.

gradual increase in both true positive rate and false positive rate can be observed.

E. Impact of EVT

A key contribution of our work is that we introduce an EVT-based tail distribution modeling and an automated threshold selection mechanism based on Hall's bootstrap method. In this section, we investigate the impact of these proposals on decision making using the MOBIO face dataset. Detection accuracy results of residuals over the 95%th percentile of the matched residual distribution was considered for this study. Since P_d captures the probabilistic information of the matched scores, it was considered for evaluation of results. It should be noted that the tail region, where the maximum amount of confusion is occurred, yields considerably lower detection accuracies as compared to that of Table II.

Shown in Figure 6 is the histogram of automatically detected thresholds from the proposed algorithm. According to Figure 6, 0.95% had been detected as the threshold in majority of the time. Based on this information, we used a fixed threshold of 0.95% to test the impact of accuracy in our experiment. We also considered the case where 0.7% is used as the threshold. Finally, we considered the case where the tail of distributions are represented using empirical distributions for comparison.

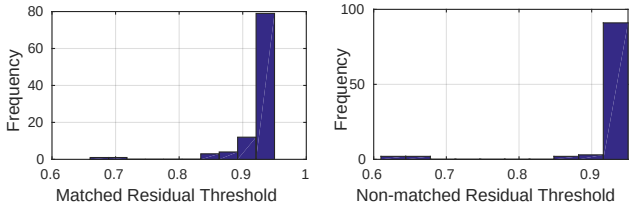


Fig. 6: Histogram of automatically detected thresholds.

Tabulated in Table V are the experimental results obtained after 50 iterations. Obtained results suggest that EVT-based modeling of the tail has improved performance by 1.12% compared to modeling the tail using an empirical distribution. For the case of modeling the tail with a fixed threshold of 0.95%, this improvement is 0.84%. This result shows that the few cases where the automated threshold selection has picked a different threshold has had a significant impact on the detection accuracy. This effect is more apparent when a generic threshold such as 0.7% is selected without any informed knowledge.

	0.7%	0.95%	Automated	Empirical
Accuracy	0.7529	0.7589	0.7653	0.7568
Std. Dev.	0.0226	0.0221	0.0215	0.0218
Increment	1.65%	0.84%	-	1.12%

TABLE V: Impact of modeling the tail and automated threshold selection.

F. Ablation Study

The proposed EOR score function defined in equation 7 has two contributory terms; namely $p(d)$ and intrusion indicator $P(I)$. In this subsection we analyze the contribution of each individual component in detection accuracy. In table VI we present mean detection accuracy across different number of users for the three datasets considered for each component in the loss function. In general, it is observed that performance of $P(I)$ is increasing with number of enrolled users. In contrast, performance of $P(d)$ is high even for the case of a single user and improves slightly when two users are enrolled. Performance of $P(d)$ doesn't fluctuate much when number of users are increased. In all cases, fusion of the two components have resulted in a better performance.

G. Modeling of Non-matched Distributions

In the proposed method, when multiple users are enrolled, non-matched residual distribution for a subject is obtained based on the remainder of enrolled users. Previous work [25] has also used the same technique to obtain non-matched distributions. However, since our work is based on a generative model, we argue that openset images are likely to follow a common distribution.

To investigate the validity of this assumption, we conducted an experiment using the UMDAA-01 dataset. With respect to its first subject, we generated non-matched distribution for each other subject. Shown in Figure 7(a) is the plot of obtained distributions. By observation it appears that there exists a high degree of similarity between these non-matched distributions.

	UMDAA-01							MOBIO-01							UMDAA-02						
	Number of Users							Number of Users							Number of Users						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
P(I)	0.85	0.85	0.87	0.90	0.92	0.93	0.94	0.80	0.87	0.91	92	0.93	0.93	0.95	0.66	0.66	0.67	0.66	0.66	0.68	0.68
P(d)	0.90	0.93	0.91	0.92	0.92	0.91	0.92	0.90	0.94	0.95	0.94	0.94	0.94	0.94	0.71	0.72	0.71	0.70	0.68	0.69	0.68
EOR	**0.90**	**0.93**	**0.92**	**0.93**	**0.93**	**0.94**	**0.94**	0.90	**0.95**	**0.96**	**0.96**	**0.97**	**0.96**	**0.97**	**0.71**	**0.72**	**0.73**	**0.70**	**0.69**	**0.70**	**0.70**

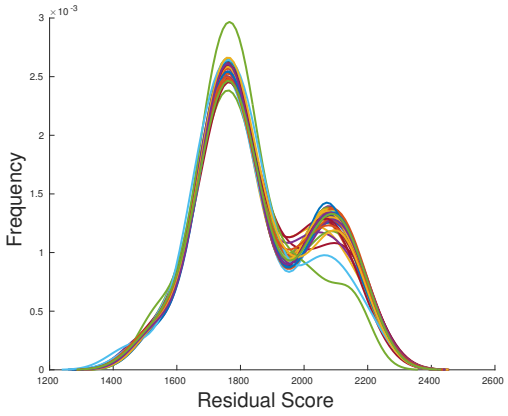
TABLE VI: Ablation study of the proposed loss function. Mean accuracy of each component against different number of users.

Subjects	1	7	10	15	20	25	30
SCI	0.80	0.95	0.98	0.97	0.98	0.98	0.99
SROSR	0.90	0.94	0.95	0.94	0.91	0.87	0.91
EOR	**0.90**	**0.97**	**0.99**	**0.98**	**0.98**	**0.98**	**0.99**

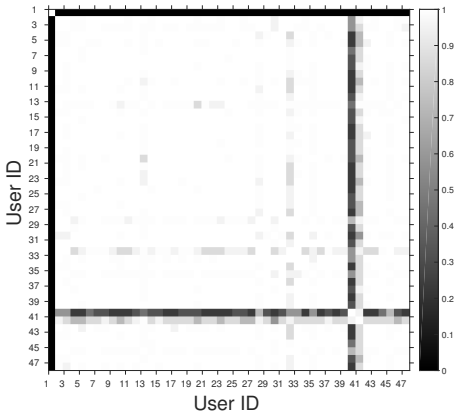
TABLE VII: Mean detection accuracy against different number of users on the MOBIO dataset.

In order to quantify this result, similarity between each pair of distributions were calculated using Kolmogorov-Smirnov test and the confusion matrix shown in Figure 7(b) was constructed using the resulting p values. Here, we have included quantified results of the first subject based on the matched distribution as a baseline for comparison.

As evident from Figure 7(b), there is a high degree of difference between the residual distribution of the first subject and the rest. In addition, except for a couple of cases, there is a high similarity between each non-match distribution. Since there is a high probability for the non-matched distribution to take a common form, it is possible to approximate this distribution by using the mean non-matched distribution of few users. In the extremal case, non-matched distribution is possible to be approximated using the non-matched distribution of a single user. It should be noted that this result may not hold true for a arbitary matching rule. However, empirically this results holds for the SRC based matching rule that we have used in this work.



(a)



(b)

Fig. 7: Effectiveness of the non-matched distribution modeling. (a) Non-matched distributions of the 1st subject of UMDAA-01 dataset. (b) Confusion matrix denoting the similarity between the non-matched distributions.

H. Effect of Having Large Number of Enrolled Users

Due to the nature of the targeted use case, proposed algorithm is intended to be used with fewer number of enrolled users. This is the reason why experiment results reported in section IV was reported upto seven users. However, it is interesting to see the sensitivity of the proposed algorithm for a relatively larger amount of enrolled users. In this sub section, we investigate the effect of the algorithm as the number of enrolled users increase.

For this purpose, we considered the MOBIO dataset and increased the number of enrolled users in several steps. The recorded detection accuracy values for SCI, SROSR and EOR are tabulated in Table VII. Acording to Table VII, as the number of users increase, the performance of EOR converges to that of SCI. This phenomenon can be explained based on the adaptive weight assignment scheme in equation (7). As the number of enrolled users increase, the first term of the equation becomes more dominating, thereby essentially capturing information about the relative sparsity same as SCI does. As a result, performance of EOR becomes comparable with SCI.

I. Impact of Dictionary Size

In all dictionary-based authentication systems, the detection performance is positively correlated with the number of atoms (number of elements in the dictionary). It is the usual practice to use over-complete dictionaries, where the number of atoms is greater than the dimensionality of data, for this purpose. In our experiments we considered the limiting case; where the number of atoms per a user is equal to the feature dimension in the extreme case (where only one user is enrolled). In this sub-section, we investigate the impact on detection accuracy when the dictionary size is further reduced. We varied the number of atoms from 10 to 50 in steps of 10 and repeated the experiments using the UMDAA01 dataset for the case of two users. The corresponding results are reported in Table VIII.

In addition to EOR, in Table VIII, we have tabulated the performance of SCI as a reference.

Atoms	10	20	30	40	50
EOR	0.85 (0.0091)	0.90 (0.0076)	0.92 (0.0012)	0.93 (0.0018)	0.93 (0.0017)
SCI	0.74 (0.0071)	0.83 0.0029	0.84 0.0018	0.84 0.0017	0.85 0.0021

TABLE VIII: Impact of the dictionary size on the detection accuracy.

As can be seen from Table VIII, for a very small dictionary size, there is a drop of performance in the proposed method. However, when the dictionary size is increased more than 30, the system performance does not change substantially. It should be noted that, this trend is also observable in SCI. Based on this observation, it is possible to obtain reasonable amount of performance even by selecting 30 atoms for the considered dataset. However, this number is dataset dependent and is generally determined by the amount of variation present in the dataset.

V. Conclusion

In this paper, we introduced multiple user AA problem and motivated the need of having a specific algorithm targetting the said problem. We presented a two fold recognition algorithm, EOR, which consists of an identification step and a verification step. We performed identification based on the residuals of the SRC algorithm. For verification, we modeled the matched and non-matched residual distributions using a semi-parametric distribution based on EVT. We proposed an automated threshold selection scheme to be used in extreme value distributions. In addition, sparsity concentration of the sparse coefficients was exploited to further increase the verification accuracy. In the process, we studied the impact of various introduced aspects of the EOR algorithm and validity of assumptions made.

We demonstrated the applicability of the proposed algorithm using three publicly available mobile face datasets. It was observed that verification-based algorithms generally perform well when low number of users are enrolled. On the other hand, identification-based algorithms required larger number of users to obtain good performance. However, good performance of both of these cases were confined to extremes with respect to number of users. On the other hand, introduced EOR method yielded superior performance consistently as the number of users were varied. Hence, it can be concluded that EOR is suited for multiple AA in mobile devices where the number of users may vary.

Acknowledgement

This work was supported by US Office of Naval Research (ONR) Grant YIP N00014-16-1-3134.

References

[1] Y. Bensalah. *Steps in Applying Extreme Value Theory to Finance: A Review*. Bank of Canada, 2010.

[2] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. *Guide to Biometrics*. SpringerVerlag, 2003.

[3] J. B. Broadwater and R. Chellappa. Adaptive threshold estimation via extreme value theory. *IEEE Transactions on Signal Processing*, 58(2):490–500, 2010.

[4] 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 *International Conference on Biometrics*, 2015.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005.

[6] D. K. . Dey and J. Yan. *Extreme Value Modeling and Risk Analysis Methods and Applications*. Taylor and Francis Group, LLC.

[7] R. B. Dunn, D. A. Reynolds, and T. F. Quatieri. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, 10(1-3):93–112, 2000.

[8] M. E. Fathy, V. M. Patel, and R. Chellappa. Face-based active authentication on mobile devices. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015.

[9] 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.

[10] X. Gibert-Serra, V. M. Patel, and R. Chellappa. Sequential score adaptation with extreme value theory for robust railway track inspection. In *IEEE International Conference on Computer Vision (ICCV) workshop on Computer Vision for Road Scene Understanding and Autonomous Driving (CVRSUAD)*, 2015.

[11] P. Hal. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, 1990.

[12] L. P. Jain, W. J. Scheirer, and T. E. Boult. Multi-class open set recognition using probability of inclusion. In *The European Conference on Computer Vision (ECCV)*, September 2014.

[13] L. P. Jain, W. J. Scheirer, and T. E. Boult. Multi-class open set recognition using probability of inclusion. In *Proceedings of 13th European Conference in Computer Vision - ECCV 2014*, pages 393–409, 2014.

[14] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides. Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 8–15, Sept 2012.

[15] R. Kumar and V. Phoha. Continuous authentication of smartphone users by fusing typing, swiping and phone movement patterns. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2016.

[16] U. Mahbub, S. Sakar, V. Patel, and R. Chellappa. Active authentication for smartphones: A challenge data set and benchmark results. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2016.

[17] S. R. S. B. S. Mahdi Abavisani, Mohsen Joneidi. A robust sparse representation based face recognition system for smartphones. In *IEEE Signal Processing in Medicine and Biology Symposium*, pages 1–6, 2015.

[18] A. Martin and M. Przybocki. Speaker recognition in a multi-speaker environment. In *INTERSPEECH*, 2001.

[19] 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 International Conference on Multimedia and Expo Workshops*, pages 635–640, July 2012.

[20] 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.

[21] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa. Dictionary-based face recognition under variable lighting and pose. *IEEE Transactions on Information Forensics and Security*, 7, 2012.

[22] J. Pelecanos, J. Navratil, and G. N. Ramaswamy. *Conversational Biometrics: A Probabilistic View*, pages 203– 224. Springer, 2008.

[23] P.Perera and V. Patel. Towards multiple user active authentication in mobile devices. In *Proceedings 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*, 2017.

[24] P. Samangouei, V. M. Patel, and R. Chellappa. Attribute-based continuous user authentication on mobile devices. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2015.

[25] W. J. Scheirer, L. P. Jain, and T. E. Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, November 2014.

[26] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult. Towards open set recognition. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence (T-PAMI), 36, July 2013.
[27] A. Serwadda, V. Phoha, and Z. Wang. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, Sept 2013.
[28] R. M. Walter J. Scheirer, Anderson Rocha and T. E. Boult. Robust fusion: Extreme value theory for recognition score normalization. In *The 11th European Conference on Computer Vision (ECCV)*, September 2010.
[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, Feb. 2009.
[30] H. Zhang and V. M. Patel. Sparse representation-based open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
[31] H. Zhang, V. M. Patel, M. Fathy, and R. Chellappa. Touch gesture-based active user authentication using dictionaries. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 207–214, 2015.
[32] Y. Zhong and Y. Deng. Sensor orientation invariant mobile gait biometrics. In *IEEE International Joint Conference on Biometrics*, pages 1–8, Sept 2014.

Pramuditha Perera Pramuditha Perera is a Ph.D. student at the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA. He received his bachelors degree in Electrical and Electronic Engineering from University of Peradeniya, Sri Lanka in 2014. He completed his masters degree in Electrical and Computer Engineering at Rutgers University, USA in 2018. His research interests include computer vision and machine learning with applications in biometrics. His work received the best student paper award at IAPR ICPR 2018.

Vishal M. Patel Vishal M. Patel [SM'15] is an Assistant Professor in the Department of Electrical and Computer Engineering (ECE) at Johns Hopkins University. Prior to joining Hopkins, he was an A. Walter Tyson Assistant Professor in the Department of ECE at Rutgers University and a member of the research faculty at the University of Maryland Institute for Advanced Computer Studies (UMIACS). He completed his Ph.D. in Electrical Engineering from the University of Maryland, College Park, MD, in 2010. His current research interests include signal processing, computer vision, and pattern recognition with applications in biometrics and imaging. He has received a number of awards including the 2016 ONR Young Investigator Award, the 2016 Jimmy Lin Award for Invention, A. Walter Tyson Assistant Professorship Award, Best Paper Award at IEEE AVSS 2017, Best Paper Award at IEEE BTAS 2015, Honorable Mention Paper Award at IAPR ICB 2018, two Best Student Paper Awards at IAPR ICPR 2018, and Best Poster Awards at BTAS 2015 and 2016. He is an Associate Editor of the IEEE Signal Processing Magazine, IEEE Biometrics Compendium, and serves on the Information Forensics and Security Technical Committee of the IEEE Signal Processing Society. He is a member of Eta Kappa Nu, Pi Mu Epsilon, and Phi Beta Kappa.