Dual-Minimax Probability Machines for One-class Mobile Active Authentication

Pramuditha Perera and Vishal M. Patel
Department of Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, MD, USA.
pperera3@jhu.edu, vpatel36@jhu.edu *

Abstract

Active Authentication(AA) systems operating on mobile devices are expected to continuously authenticate the enrolled user based on enrolled sensor observations. Due to unavailability of training samples from negative classes, AA can be viewed as a one-class classification problem. In this work we introduce a Single-class Minimax Probability Machine(1-MPM) based solution called Dual Minimax Probability Machines(DMPM) for AA applications. In contrast to 1-MPM, proposed method has two notable differences. 1) We learn an additional hyper-plane to separate training data from the origin by taking into account maximum data covariance. 2) We consider the possibility of modeling the underline distribution of training data as a collection of sub-distributions. Intersection of negative half spaces defined by the two learned hyper-planes is considered to be the negative space during inference. We demonstrate the effectiveness of the proposed mechanism by performing evaluations on three publicly available face based AA datasets.

1. Introduction

Both security and technology community witnessed a rapid paradigm shift in mobile device authentication in recent years. During this time, traditional passive device authentication mechanisms such as passwords and swipe-patterns have been largely replaced by biometric based authentication methods (such as fingerprint and face based systems). Although biometrics based passive authentication provide more security compared to traditional means, it is still vulnerable to intrusions occurring between two consecutive authentications. To this end, Active Authentication(AA) has been proposed as an alternative solution.

The goal of active authentication is to continuously monitor sensor data of the device and to perform continuous user authentication [25]. In recent works, various biometric modalities including frontal face image [9], swipe patterns [22], key-pad typing patterns and gait [40], has been proposed for AA. In particular, frontal-face based AA has emerged as a strong candidate largely owing to advances in face recognition technology [23]. Although users find face-based authentication convenient, sharing captured face images (or its representations) with a centralized system (or potentially with other devices) could arise complications in terms of data privacy. Therefore, in general, AA is treated as a one-class classification problem where a classifier is trained solely on device owner's biometric samples as shown in Figure 1.

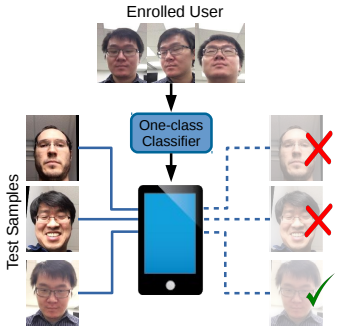


Figure 1. Active Authentication(AA) as a one-class problem. Device owner provides a sample set of observations (face images in this case) during enrollment. Based on enrolled data, goal of AA is to continuously obtain samples of the device user and perform authentication. Since the device performs authentication solely based on enrolled observations it is a one-class problem.

Due to constraints in the problem formulation, one-class classification based AA becomes a very challenging problem. In multi-class classification, discriminative classifiers are learned by taking into account relative spread of different classes. Such a discriminative approach is not possible in one-class classification due to the absence of negative classes during training. As a result, one-class classification is traditionally treated as a quantile estimation problem

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

- where a compact boundary for the class is sough in the feature space. One-class support vector machine(OCSVM) [32] and Support Vector Data Descriptor(SVDD) [38] are two of widely used quantile estimation methods for this purpose in the literature. As shown in Figure 2, these two formulations try to find the tightest half-space and the tightest hyper-ball that contain training data respectively.

In our work, We view AA problem as an equivalent problem to novelty detection, where we treat data of enrolled user as *nominal data* and data of intruders as *novel data*. We build our solution on the foundations of a classical novelty detection method - Single class Mini-max Probability Machines(1-MPM)[21]. Similar to one-class SVM, 1-MPM tries to find the tightest half space that contains training data. But, different from one-class SVM, MPM strives to arrive at better separating hyper-plane by taking second order statistics of data into consideration as shown in Figure 2. Building on foundations of 1-MPM, we make following three contributions in this work:

- We learn an additional hyper-plane considering projection that maximizes covariance of data. By combining the learned hyper-plane with the decision boundary obtained through 1-MPM, we narrow down the negative space in to a polyhedron as illustrated in Figure 2.
- We extend one-class classification based on 1-MPM for the case where underline data distribution is not homogeneous - i.e. when there exists a collection of sub-distributions which is often the case with face data.
- We obtain state of the art one-class classification results on three publicly available face datasets.

2. Related Work

In this section we briefly discuss work related to one-class Active Authentication. In addition, we briefly discuss developments in face representation, one class learning and open-set learning.

Active Authentication. As described in the preceding section, AA aims at producing a classifier solely based on sample observations obtained from device users. Although some works have investigated the problem of multiple-user AA [27], AA is conventionally studied for the case of a single device user. Effectiveness of AA has been studied under various biometric modalities including gait [40], [17] , keystroke, voice, swipe patterns [22], [34], [12], [19], and face images [9], [14], [23], [26]. However, majority of previous works focus on either devising new system architectures [9] or better features [22], [34] for recognition. In contrast, in our work, we focus on developing a better one-class classifier for the purpose of AA.

Face representation. Face representation has evolved dramatically in recent years with the advent of deep-learning. Hand crafted features such as Gabour, LBP [37] and HOG [10] have been replaced with more efficient deep-learning based descriptors [36], [7]. It was later shown that performance in face recognition, particularly in verification, can be improved when both verification and classification tasks are learned jointly [35]. Both VGG-Face [24] network and FaceNet [33] architectures take advantage of this fact by integrating triplet loss in the learning procedure. Although it is possible to learn one-class features through autoencoder networks [5] or Generative Adversarial Networks [30], such introverted approaches produce less discriminative features compared to representations obtained using aforementioned methods. Therefore, it is the norm in one-class face based AA applications to use classical one-class learning tools on top of deep-features for recognition [23].

One-class learning. One-class learning is a tool used in many biometric applications [13], [1], [28]. It is traditionally treated either as a density estimation problem or a quantile estimation problem. In the former case, distribution of the given class is approximated using a predetermined distribution - where distribution is commonly chosen to be Gaussian or a mixture of Gaussians [18], [29]. In comparison to density estimation, quantile estimation methods have been applied more commonly for various one-class learning applications. One-class SVM[32] is one of the earliest, and most commonly used one-class learning schemes proposed. Here, the best hyper-plane that separates a high portion of data from the origin is sought. In contrast, Support Vector Data Descriptor(SVDD) [38] tries to construct a hyper-ball that contains a high proportion of data in the interior of the hyper-ball. Single class MiniMax probability machines (1-MPM) [21], seeks to find a hyper-plane similar to that of one-class SVM by taking second order statistics of data into consideration. As a result, 1-MPM ends up learning a decision boundary that generalize well to the underline concept. Further, one-class classification can be viewed as an extreme case of open-set classification. In order to reduce false positives due to open-set samples, one-vs-set machine introduced in [31] constructs a slab parallel to the SVM decision boundary. A similar strategy is employed in one-class slab SVM introduced in [15].

3. Proposed Method

We build the proposed method using single class MPM(1-MPM) as a foundation with two noticeable improvements. In this section, we first review 1-MPM in detail. Initially we consider the non-robust version of 1-MPM proposed in [21], where we set both robust parameters to zero. Then we introduce the proposed modifications to the 1-MPM to arrive at Dual-minimax probability machines(DMMP). Finally we introduce the robust version of DMMP.

Single-class MPM. For given set of data $\mathbf{x} \sim (\hat{\mathbf{x}}, \mathbf{\Sigma_x})$

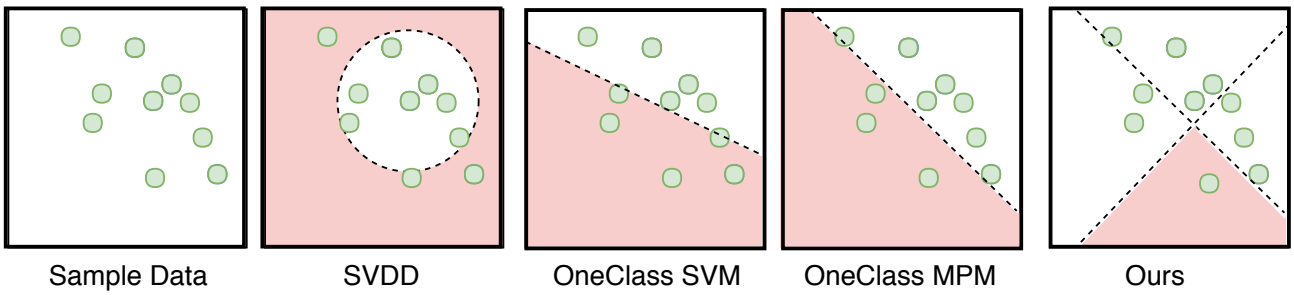


Figure 2. Comparisons of one-class classification methods. Classification planes obtained for data denoted by green circles in the first picture are shown. Negative spaces learned by both one-class SVM and SVDD overlaps with regions where positive samples appear. Comparatively, 1-MPM has learned a tighter hyper-plane agreeing with structure of data. Negative space learned by considering directions that minimizes and maximizes data variance in the proposed method provides a very low overlap with sample data.

(where data is from any distribution with mean $\hat{\mathbf{x}}$ and covariance $\mathbf{\Sigma_x}$), 1- MPM seeks to find a hyper-plane $(\mathbf{a}, b)$ with $\mathbf{a} \in \mathbb{R}^n \setminus \{0\}, b \in \mathbb{R}$ at least with a probability of α, such that data lies in the positive half space defined by $\{\mathbf{z}|\mathbf{z} \in \mathbb{R}^n, \mathbf{a}^T\mathbf{z} \geq b\}$ even in the worst case scenario. Similar to in one-class SVM, 1-MPM tries to maximize the distance between the origin and learned hyper-place with the objective of arriving at a tighter lower bound to the data. With this background, the objective function of single class MPM can be formulated as in,

$$\max_{\mathbf{a}\neq 0, b} \frac{b}{\sqrt{\mathbf{a}^T\mathbf{\Sigma_x}\mathbf{a}}} \text{ s.t. } \inf_{\mathbf{x}\sim(\hat{\mathbf{x}},\mathbf{\Sigma_x})} \mathbb{P}(\mathbf{a}^T\mathbf{x} \geq b) \geq \alpha,$$

when the distance between the origin and the hyper-plane is measured in terms of Mahalonabis distance with respect to $\mathbf{\Sigma_x}$. Since this problem is positively homogeneous in $(\mathbf{a}, b)$ and because $a \neq 0$ is always satisfied when $b > 0$, value of b is taken to be one without loss of generality. Then, a equivalent optimization problem can be obtained by minimizing the reciprocal of Mahalonobis distance as in,

$$\min_{\mathbf{a}} \sqrt{\mathbf{a}^T\mathbf{\Sigma_x}\mathbf{a}} \text{ s.t. } \inf_{\mathbf{x}\sim(\hat{\mathbf{x}},\mathbf{\Sigma_x})} \mathbb{P}(\mathbf{a}^T\mathbf{x} \geq 1) \geq \alpha. \quad (1)$$

Using the core MPM theorem in [20], where it is stated that $\inf_{\mathbf{x}\sim(\hat{\mathbf{x}},\mathbf{\Sigma_x})} \mathbb{P}(\mathbf{a}^T\mathbf{x} \geq b) \geq \alpha$ is equivalent to $b - \mathbf{a}^T\hat{\mathbf{x}} \geq K(\alpha)\sqrt{\mathbf{a}^T\mathbf{\Sigma_x}\boldsymbol{a}}$, where $K(\alpha) = \sqrt{\frac{\alpha}{1-\alpha}}$, equation 1 can be re-written as,

$$\min_{\mathbf{a}} ||\mathbf{a}^T\mathbf{\Sigma}_{\mathbf{x}}^{\frac{1}{2}}||_2 \text{ s.t. } 1 - \mathbf{a}^T\hat{\mathbf{x}} \geq K(\alpha)||\mathbf{a}^T\mathbf{\Sigma}_{\mathbf{x}}^{\frac{1}{2}}||_2. \quad (2)$$

Here we note that for a real symmetric covariance matrix $\mathbf{\Sigma}_{\mathbf{x}}^{\frac{1}{2}}$ always exists. Since optimization problem shown in equation 2 is a second order cone program it can be efficiently solved using convex optimization tools.

Maximum variance MPM (MVMPM). When the derived optimization problem in equation 1 is solved, the solver is effectively searching for a feasible vector $\mathbf{a}$ that minimizes $\sqrt{\mathbf{a}^T\mathbf{\Sigma_x}\mathbf{a}}$. However, we note that $\mathbf{a}^T\mathbf{\Sigma_x}\mathbf{a}$ is the covariance of data projected on vector $\mathbf{a}$. Therefore, by solving equation 1, 1-MPM is effectively searching for a feasible direction along which data will have a very low variance. In fact, in an unconstrained setting, optimal solution of equation 1 would be $\mathbf{v}_{min}(\mathbf{\Sigma_x})$, the Eigen vector corresponding to the lowest Eigen value of $\mathbf{\Sigma_x}$, as $\min_{\mathbf{a},||\mathbf{a}||_2=1} \mathbf{a}^T\mathbf{\Sigma_x}\mathbf{a} = \mathbf{v}_{min}(\mathbf{\Sigma_x})$.

However, studies on Principle Component Analysis(PCA) suggests that selecting a projection that minimizes covariance is a poor choice for data representation [2]. PCA suggests that directions along where covariance is largest are the most sensible choices in this regard. It is shown that such directions are less effected by noisy variations of data [2]. Based on this background, we propose to learn a second hyper-plane $(\tilde{\mathbf{a}}, \tilde{b})$ such that the data projected on the hyper-plane $\tilde{\mathbf{a}}$ has the largest possible variance. Specifically we aim to solve,

$$\max_{\tilde{\mathbf{a}}} \sqrt{\tilde{\mathbf{a}}^T\mathbf{\Sigma_x}\tilde{\mathbf{a}}} \text{ s.t. } \inf_{\mathbf{x}\sim(\hat{\mathbf{x}},\mathbf{\Sigma_x})} \mathbb{P}(\tilde{\mathbf{a}}^T\mathbf{x} \geq 1) \geq \tilde{\alpha}. \quad (3)$$

Here we note that maximizing $\sqrt{\tilde{\mathbf{a}}^T\mathbf{\Sigma_x}\tilde{\mathbf{a}}}$ is equivalent to minimizing $\sqrt{\tilde{\mathbf{a}}^T{\mathbf{\Sigma_x}}^{-1}\tilde{\mathbf{a}}}$. Optimization problem in equation 3 can be transformed into another second order cone program of the form,

$$\min_{\tilde{\mathbf{a}}} ||\tilde{\mathbf{a}}^T\mathbf{\Sigma}_{\mathbf{x}}^{-\frac{1}{2}}||_2 \text{ s.t. } b - \tilde{\mathbf{a}}^T\hat{\mathbf{x}} \geq K(\tilde{\alpha})||\tilde{\mathbf{a}}^T\mathbf{\Sigma}_{\mathbf{x}}^{-\frac{1}{2}}||_2. \quad (4)$$

Solution of equation 4 gives a second hyper-plane $(\tilde{\mathbf{a}}, \tilde{b})$ which can be used for decision making.

Generalizing for Multiple Sub-distributions. In devising hyper-planes $(\mathbf{a}, b)$ and $(\tilde{\mathbf{a}}, \tilde{b})$ in preceding sub-sections, second order statistics were exploited under the assumption that all data points are sampled from a single underline distribution. However, in face representation, images of the

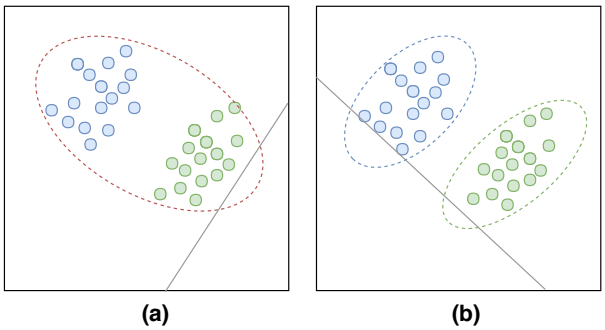


Figure 3. Effect of having sub distributions in data. Shown in Figure is representation of face images belonging to a single user. Projection vector that minimizes covariance is shown by a gray solid line (a) 1-MPM considers overall second order statistics of the data. (b) When data is clustered, two clusters with different sub-distributions can be identified. The projection vector obtained in this case is more representative of data.

same identity with variations (such as distinct poses) are likely to lie in different face sub-spaces. Therefore, for applications in face recognition, it is more reasonable to assume that data are sampled from a set of distinct sub-distributions.

When multiple sub distributions exist, approximating the data spread using a single covariance matrix is sub-optimal as illustrated in in Figure 3. Data distribution shown in Figure 3 has two identifiable clusters. Shown in sub-figures (a) and (b) are minimum variance direction with respect to the distribution when existence of one and two clusters are assumed respectively. Here, we note that minimum variance directions sough in the former case does not agree with the spread of data. This fact is valid for both minimum and maximum variance directions. In such situations, decision boundaries given by 1-MPM and maximum variance MPM too will be sub-optimal. Sub-optimality of 1-MPM in the presence of sub-distributions has been discussed in [16] in the context of binary classification. However, ours is the first work that addresses this issue with respect to one-class classification.

In order to mitigate the effect of sub-distributions, we propose to leverage both local and global level information. In particular, we view the given dataset both as a single cluster and as a collection of sub-clusters in decision making. In our approach, first we determine the number of clusters n in the dataset using Caliński-Harabasz[4] method. Then, data is clustered into n clusters using the Ward's method[39]. We calculate global mean and variance $(\hat{\mathbf{x}}_G, \mathbf{\Sigma}_{\mathbf{x},G})$ with respect to the whole dataset along with cluster-wise statistics $(\hat{\mathbf{x}}_i, \mathbf{\Sigma}_{\mathbf{x},i})$ for each i^{th} cluster. Since we require to find the direction with minimum/maximum variance, we propose to optimize over cumulative covariance of each individual cluster $\sum_i \sqrt{\mathbf{a}^T \mathbf{\Sigma}_{\mathbf{x},i} \mathbf{a}}$ while constraints are defined with respect to global statistics $(\hat{\mathbf{x}}_G, \mathbf{\Sigma}_{\mathbf{x},G})$. In the following sub-section we generalize 1-MPM and maximum-variance MPM for the case of multiple sub-distributions to arrive at the proposed algorithm - Dual MPM (DMPM).

Robust MPM formulation. In MPM versions presented in the preceding subsections, it is assumed that sample mean and covariance is a reasonable estimate of true statistics in the underline distribution. However, in practice, this may not always be the case. Therefore, decision boundaries obtained with the current formulation will be highly susceptible to population noise. As a preventive measure [21] presents a robust version of 1-MPM.

In the robust 1-MPM formulation, it is assumed that difference between sample covariance and true covariance of the distribution will not exceed ρ and mahalanobis distance between sample mean and true mean with respect to true covariance will not exceed ν^2. Under these assumptions, it is shown in [21] that $\mathbf{a}^T \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a}$ term in equation 1 gets substituted by $\mathbf{a}^T (\mathbf{\Sigma}_{\mathbf{x}} + \rho \mathbf{I}_{\mathbf{n}}) \mathbf{a}$ while $K(\alpha)$ term is changed into $K(\alpha) + \nu$. With the same argument, assuming that difference between sample covariance and true covariance of the distribution will not exceed $\tilde{\rho}$ for $\mathbf{\Sigma}_{\mathbf{x}}^{-1}$, the robust version of MVMPM can be obtained by substituting $\tilde{\mathbf{a}}^T {\mathbf{\Sigma}_{\mathbf{x}}}^{-1} \tilde{\mathbf{a}}$ term in equation 3 by $\tilde{\mathbf{a}}^T (\mathbf{\Sigma}_{\mathbf{x}} + \tilde{\rho} \mathbf{I}_{\mathbf{n}})^{-1} \tilde{\mathbf{a}}$.

Single-class Dual MPM (DMPM). In this work we propose to use both hyper-planes $(\mathbf{a}, b)$ and $(\tilde{\mathbf{a}}, \tilde{b})$ collectively for decision making. Specifically, we treat the union of positive half spaces of the two hyper-planes as the positive space with respect to the given class. Given global and local clusters $(\hat{\mathbf{x}}_G, \mathbf{\Sigma}_{\mathbf{x},G})$ and $(\hat{\mathbf{x}}_i, \mathbf{\Sigma}_{\mathbf{x},i})$ for $i = 1 \dots c$, where c is the number of clusters , We solve the following joint-optimization problem:

$$\begin{aligned} \underset{\mathbf{a},\tilde{\mathbf{a}}}{\text{minimize}} \quad & \sum_{i=1}^{c} ||\mathbf{a}^T (\mathbf{\Sigma}_{\mathbf{x},i} + \rho \mathbf{I}_{\mathbf{n}})^{\frac{1}{2}}||_2 + ||\tilde{\mathbf{a}}^T (\mathbf{\Sigma}_{\mathbf{x},i} + \tilde{\rho} \mathbf{I}_{\mathbf{n}})^{-\frac{1}{2}}||_2 \\ \text{subject to} \quad & (K(\tilde{\alpha}) + \nu) ||\tilde{\mathbf{a}}^T (\mathbf{\Sigma}_{\mathbf{x},i} + \tilde{\rho} \mathbf{I}_{\mathbf{n}})^{-\frac{1}{2}}||_2 - 1 \leq \tilde{\mathbf{a}}^T \hat{\mathbf{x}} \\ & (K(\alpha) + \nu) ||\mathbf{a}^T (\mathbf{\Sigma}_{\mathbf{x},i} + \rho \mathbf{I}_{\mathbf{n}})^{\frac{1}{2}}||_2 - 1 \leq \mathbf{a}^T \hat{\mathbf{x}}, \end{aligned} \tag{5}$$

to find hyper-plane parameters $\mathbf{a}$ and $\tilde{\mathbf{a}}$. Since products of $\mathbf{a}$ and $\tilde{\mathbf{a}}$ do not appear in the optimization statement, this problem can be solved independently for $\mathbf{a}$ and $\tilde{\mathbf{a}}$ using two second order cone programs. Once hyper-plane parameters are obtained, given a test sample $\mathbf{y}$, identity of the sample would be assigned to be negative if $\mathbf{a}^T \mathbf{y} < 1 \cap \tilde{\mathbf{a}}^T \mathbf{y} < 1$, and positive otherwise.

4. Experiments

We evaluate the effectiveness of the proposed method over three publicly available mobile active authentication datasets with the face modality. As baseline comparisons

we considered following one-class classification methods:

1. **Linear One-class SVM (1SVM) [32].** LibSVM package[8] was used to implement one-class SVM.
2. **Linear SVDD (SVDD) [38].** SVDD extension to LibSVM provided in [6] was used.
3. **Kernel one-class SVM (k1SVM) [32].** LibSVM implementation of one-class SVM was used with a RBF kernel.
4. **Kernel SVDD (kSVDD) [38].** SVDD performed on a higher dimensional space using a RBF kernel. LibSVM extension found at [6] was used for implementation.
5. **Single class MPM (1MPM)[21].** Implementation of the original work of single class MPM was done by the authors.
6. **One-class kNFST (kNFST) [3].** A single null-space direction is found where intra-class distance of the class is low. Code released by authors was used for evaluations.
7. **One-vs-set machine (1vsSet) [31].** An additional hyper-plane is learned to reduce open-space risk. Implementation is based on the code provided by authors.

All baselines along with the proposed method was evaluated using a common protocol. Each dataset was split into two halves in terms of classes. Classes of the second half of the split were considered to be unknown classes for all tests. During testing, classes of the first split was considered one at a time as the known class to perform one-class classification between the known class and all unknown classes. Objects of the known class is further split into two to form a training set and a testing set. In our experiments we used 200 samples taken randomly from the training set to train the classifier. Hyper-parameters of all methods were chosen based on a validation set. For all MPM based methods we set $\nu = 0$ [21]. For all tests we used 4096 dimensional VGG-Face face descriptor [24] as the feature. Prior to testing, we first performed principle component analysis to reduce the dimension of the feature to 199 to make sure covariance matrices are full rank. In order to compare performance of each method, we used average detection accuracy(ADA) defined as, $ADA = \frac{TP+TN}{2}$,where, TP and TN are the true positive rate and true negative rate respectively. For all experiments, obtained average detection accuracy results for each method are tabulated in Table 1.

5. Results

In this section we first present experimental results carried on UMDAA01, MOBIO and UMDAA02 datasets. Then, we analyze the impact of each component of the proposed system using a ablation study. The section is concluded with a discussion about the ability of the proposed method to generalize to larger test sets.

5.1. UMDAA01 Face Dataset

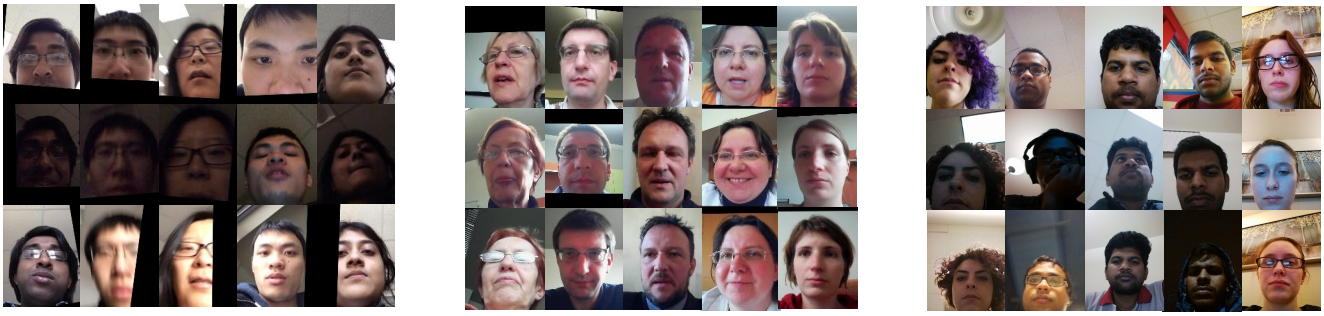
The UMDAA-01 dataset [14] contains images captured using the front-facing camera of a iPhone 5S mobile device of 50 different individuals captured across three sessions with varying illumination conditions. Images of this dataset contain pose variations, occlusions, partial clippings as well as natural facial expressions as evident from the sample images shown in Figure 4(a). For our experiments we concatenated images from all three sessions to form 50 classes. According to Table 1, Both linear versions of one class SVM and SVDD have obtained average detection results for this dataset. We note that 1-MPM, kernel SVDD and kernel one class SVM has obtained average detection accuracy in mid 70%. However, proposed method has achieved the best performance at 82.5% improving performance of 1-MPM by 6%.

5.2. MOBIO Face Dataset

MOBIO dataset contains images and voice samples of 150 individuals sourced across six locations. Data of each individual is sourced through either smartphone devices or laptops. In this experiment we only consider face images of the dataset. Sample images of the dataset are shown in Figure 4(b). Comparative to the other two datasets considered in the paper, MOBIO contains less variations and mostly front-facing face images as evident from Figure 4. Nevertheless, this dataset provides means to understand the stability of methods for significantly higher number of users. As evident in Table 1, MPM based methods have demonstrated superior performances in this dataset. Performance of other methods are comparable to that of UMDAA01. In particular, we note our method has reported an improvement of more than 5% compared to 1-MPM.

5.3. UMDAA02 Face Dataset

The UMDAA-02 Dataset [23] is an unconstrained multimodal dataset with 44 subjects where 18 sensor observations were recorded across two month period using a Nexus 5 mobile device. Authors of [23] have made face modality and touch-data modality publicly available. In our work we only consider the face modality to perform tests. A sample set of images obtained from this data set is shown in Figure 4(c). UMDAA02 is a more challenging dataset compared to UMDAA01 and MOBIO as apparent from the sample images shown in Figure 4. In particular, we note the existence of a huge intra-class variations in this dataset in terms of poses, partial faces, illumination as well as appearances of the users. In this dataset, we first sorted all 44



(a) (b) (c)

Figure 4. Sample images taken from the two Mobile AA datasets used for testing. (a)UMDAA01. (b)MOBIO. (c) UMDAA02. Each column shows multiple images captured from a single user.

Table 1. Tabulation of average detection accuracy for considered one-class classification methods for UMDAA01 and UMDAA02 datasets (variance is indicated within parenthesis).

	1SVM[32]	k1SVM[32]	SVDD[38]	kSVDD[38]	kNFST[3]	1vsSet[31]	1MPM[21]	DMPM
UMDAA01	0.632 (0.004)	0.748 (0.004)	0.582 (0.007)	0.763 (0.013)	0.560 (0.003)	0.670 (0.005)	0.768 (0.003)	**0.825** (0.007)
MOBIO	0.622 (0.002)	0.731 (0.009)	0.615 (0.018)	0.701 (0.009)	0.567 (0.012)	0.593 (0.017)	0.816 (0.003)	**0.869** (0.001)
UMDAA02	0.614 (0.008)	0.649 (0.004)	0.515 (0.007)	0.550 (0.007)	0.556 (0.003)	0.538 (0.003)	0.722 (0.006)	**0.760** (0.007)

subjects in terms of the number of images and following the same protocol as before, we considered first 22 subjects as known classes. The remaining 22 subjects were considered to be unknown. In Table 1, we notice a drop in performance in each method compared to other datasets. This is not surprising due to the challenging nature of the dataset. In this dataset, one-class SVM has performed better than SVDD both in linear and kernel versions. Proposed method is able to register best detection accuracy by improving 1-MPM by a margin of nearly 4%.

5.4. Ablation Study

The presented method proposes two modifications on top of the classical 1-MPM algorithm. In this subsection we investigate the effect of each individual proposal on performance. We carried out experiments using maximum variance MPM(MV-MPM) and 1-MPM with multiple sub-distributions(MD-MPM) introduced earlier in the text. Obtained average detection accuracy values are tabulated in Table 2. First, we note that in MOBIO dataset, MD-MPM has not introduced an improvement on top of 1-MPM in terms of mean detection accuracy. Since MOBIO contains only front-facing images, this observation is justifiable when it is assumed that data can be approximated by single covariance matrix. On other two datasets, we note that considering the existence of multiple sub-distributions have improved performance of 1-MPM marginally. In comparison, Maximum-Variance MPM have obtained nearly a 3% gain in detection accuracy in both UMDAA01 and UMDAA02. This observation validates our argument on the suitability of using maximum covariance direction for one-class classification. Dual-MPM which takes into account all contributions of the paper has achieved the best detection accuracy performances across both datasets.

Table 2. Ablation study on how each component of the proposed method influences on recognition performance. Average detection accuracy values are tabulated with variance figures shown within parenthesis. Here we note that Maximum variance MPM performs better than 1-MPM.

	1MPM	MV-MPM	MD-MPM	DMPM
UMDAA01	0.768 (0.006)	0.793 (0.005)	0.769 (0.003)	0.825 (0.007)
MOBIO	0.816 (0.006)	0.816 (0.005)	0.816 (0.004)	0.869 (0.007)
UMDAA02	0.721 (0.003)	0.742 (0.003)	0.739 (0.003)	0.760 (0.007)

5.5. Impact of Number of Unknown Classes

In all experiments conducted above, we kept the number of unknown classes fixed during testing. In this subsection, we investigate the impact of number of unknown classes has on detection accuracy. We conducted an experiment on MPM based methods using UMDAA01 dataset. We considered first four classes one at a time as enrolled classes and changed number of unknown classes from 22 to 37 in

Table 3. Impact of number of unknown classes have on performance. DMPM is less affected by when number of unknown classes are increased compared to 1-MPM.

# of unknown classes	37	32	27	22
1-MPM	0.689 (0.016)	0.690 (0.016)	0.690 (0.016)	0.725 (0.025)
DMPM	0.740 (0.029)	0.742 (0.030)	0.742 (0.030)	0.746 (0.031)

steps and evaluated classification accuracy with respect to remaining classes. Obtained results for this experiment is tabulated in Table 3. Based on the tabulations in Table 3, it is evident that detection accuracy decreases for both cases when number of unknown classes are increased. This drop is justifiable as decision ambiguity increases with number of unknown classes [11]. However, a drastic drop in performance can be observed for 1-MPM when number of unknown classes are increased from 22 to 27. Performance drop in DMPM is not as significant. This experiment empirically shows that DMPM is more robust to the change of number of unknown classes compared to 1-MPM.

6. Conclusion

In this paper we introduced a one-class classification method targeting the application of face-based Active Authentication. We developed our solution based on single-class Minimax Probability Machines(1-MPM). We argued that a better decision hyper-plane compared to 1-MPM can be found by considering the maximum variance direction of data and we provided means to find such boundary (MVMPM). In our experiments we showed that decision boundary found in this manner is indeed better than the decision boundary produced by 1-MPM. Further, We proposed fusing positive hyper-planes of 1-MPM and MVMPM to obtain a narrower negative space. Finally, we generalized the proposed framework to the case where underline distribution has multiple sub-distributions. Proposed method was compared against standard one-class classification frameworks using three publicly available face-based AA datasets. In all datasets, proposed method demonstrated an improvement of 4%-6% compared to 1-MPM.

References

[1] S. Arashloo, J. Kittler, and W. Christmas. An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol. *IEEE Access*, 5, July 2017.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 2006.

[3] P. Bodesheim, A. Freytag, E. Rodner, M. Kemmler, and J. Denzler. Kernel null space methods for novelty detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[4] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-Simulation and Computation*, 3(1):1–27, 1974.

[5] R. Chalapathy, A. K. Menon, and S. Chawla. Anomaly detection using one-class neural networks. *CoRR*, 2018.

[6] W.-C. Chang, C.-P. Lee, and C.-J. Lin. A revisit to support vector data description (svdd). 2013.

[7] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 539–546, June 2005.

[8] C. chung Chang and C.-J. Lin. Libsvm: a library for support vector machines, 2001.

[9] 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 *2015 International Conference on Biometrics (ICB)*, pages 135–142, May 2015.

[10] 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.

[11] B. DeCann and A. Ross. Modelling errors in a biometric re-identification system. *IET Biometrics*, 4(4):209 – 219, Dec. 2015.

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

[13] Y. Ding and A. Ross. An ensemble of one-class svms for fingerprint spoof detection across different fabrication materials. In *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2016.

[14] 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.

[15] V. Fragoso, W. J. Scheirer, J. P. Hespanha, and M. Turk. One-class slab support vector machine. *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 420–425, 2016.

[16] B. Gu, S. Xingming, and V. Sheng. Structural minimax probability machine. 1:1–11, 04 2016.

[17] 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.

[18] M. Kemmler, E. Rodner, E.-S. Wacker, and J. Denzler. One-class classification with gaussian processes. *Pattern Recogn.*, 46(12):3507–3518, 2013.

[19] 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.

[20] G. R. G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan. Minimax probability machine. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, pages 801–807, 2001.

[21] G. R. G. Lanckriet, L. E. Ghaoui, and M. I. Jordan. Robust novelty detection with single-class MPM. In *Advances in Neural Information Processing Systems 15 [Neural Information Processing Systems, NIPS 2002, December 9-14, 2002, Vancouver, British Columbia, Canada]*, pages 905–912, 2002.

[22] E. M. I. M. M. Frank, R. Biedert 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, 2013.

[23] 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.

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[25] 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.

[26] P. Perera and V. Patel. Efficient and low latency detection of intruders in mobile active authentication. *IEEE Transactions on Information Forensics and Security*, 13(6):1392 – 1405, June 2018.

[27] P. Perera and V. M. Patel. Towards multiple user active authentication in mobile devices. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2017.

[28] P. Perera and V. M. Patel. Learning Deep Features for One-Class Classification. *ArXiv e-prints*, Jan. 2018.

[29] E. Rodner, E. sabrina Wacker, M. Kemmler, and J. Denzler. One-class classification for anomaly detection in wire ropes with gaussian processes in a few lines of code. In *In: Conference on Machine Vision Applications (MVA)*, pages 219–222, 2011.

[30] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli. Adversarially learned one-class classifier for novelty detection. *CoRR*, 2018.

[31] 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.

[32] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471, 2001.

[33] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 815–823, 2015.

[34] 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.

[35] Y. Sun, Y. Chen, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. In *Advances in Neural Information Processing Systems 27*, pages 1988–1996. 2014.

[36] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1891–1898, 2014.

[37] A. H. T. Ahonen and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 28:2037–2041, 2006.

[38] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Mach. Learn.*, 54(1):45–66, 2004.

[39] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[40] Y. Zhong and Y. Deng. Sensor orientation invariant mobile gait biometrics. In *IEEE International Joint Conference on Biometrics*, pages 1–8, Sept 2014.