Image and Vision Computing 28 (2010) 1098–1105

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Biometric classifier update using online learning: A case study in near infrared face verification

Richa Singh [a], Mayank Vatsa [a,*], Arun Ross [b], Afzel Noore [b]

[a] Indraprastha Institute of Information Technology (IIIT), New Delhi, India
[b] Lane Department of CSEE, West Virginia University, Morgantown, WV, USA

ARTICLE INFO

Article history:
Received 22 September 2008
Received in revised form 18 January 2010
Accepted 26 January 2010

Keywords:
Biometrics
Online learning
Support vector machine
Near infrared face verification

ABSTRACT

The performance of a large scale biometric system may deteriorate over time as new individuals are continually enrolled. To maintain an acceptable level of performance, the classifier has to be re-trained offline in batch mode using both existing and new data. The process of re-training can be computationally expensive and time consuming. This paper presents a new biometric classifier update algorithm that incrementally re-trains the classifier using online learning and progressively establishes a decision hyperplane for improved classification. The proposed algorithm incorporates soft labels and granular computing in the formulation of a 2ν-Online Granular Soft Support Vector Machine (SVM) to re-train the classifier using only the new data. Granular computing makes it adaptive to local and global variations in data distribution, while soft labels provide resilience to noise. Each time data is acquired, new support vectors that are linearly independent are added and existing support vectors that do not improve the classifier performance are removed. This constrains the size of the support vectors and significantly reduces the training time without compromising the classification accuracy. The efficacy of the proposed online learning strategy is validated in a near infrared face verification application involving different covariates. The results obtained on a heterogeneous near infrared face database of 328 subjects show that in all experiments using different feature extraction and classification algorithms the proposed online 2ν-Granular Soft Support Vector Machine learning approach is 2–3 times faster while achieving a high level of accuracy similar to offline training using all data.

1. Introduction

A carefully designed biometric system should be stable and robust to environmental dynamics and variations in the data distribution. However, in real world applications, these systems are affected by several factors including template aging, improper interaction of the user with the sensor, and noise. Furthermore, as the size of a biometric database increases, the system has to re-train itself in order to handle the variations introduced due to the newly enrolled subjects. For example, an identification system may have to re-compute its thresholds to handle the increased number of subjects, a verification system may have to re-compute its decision boundaries based on the new set of genuine and impostor scores, and a face verification system may have to update its templates to address the aging effects of the face biometric. The re-training process can be time consuming and may not be pragmatic in large scale applications. A generic biometric system, as shown in Fig. 1, has five modules that require regular update or re-training. Template update is required to address the issue of template aging, sensor update is necessary to accommodate advancements in sensor technology, re-training of the preprocessing and feature extraction algorithms is necessary to handle variations in data, and classifiers have to be updated to account for changes in the intra-class and inter-class dynamics of the subjects.

The paper focuses on developing online learning algorithm at the classifier level when the number of subjects is increased. Currently, biometric classifiers such as support vector machine (SVM) and neural network are trained *offline* with the available training data and domain specific knowledge. However, large scale biometric applications such as US-VISIT and FBI-IAFIS continuously enroll new individuals. Due to the high computational complexity required for re-training in such applications, it is not feasible to regularly update the classifier knowledge and decision boundary, thereby affecting the verification performance of the biometric system. Online learning presents an efficient alternative to offline learning and classification by updating the classifier knowledge upon the arrival of new data. While it has been extensively used in problems related to machine learning [1–7], *online learning is not very well studied in biometrics*, though some of the existing online learning algorithms show experimental results on face recognition [6,7]. Existing research

* Corresponding author. Tel.: +91 9654653404.
E-mail addresses: rsingh@iiitd.ac.in (R. Singh), mayank@iiitd.ac.in (M. Vatsa), arun.ross@mail.wvu.edu (A. Ross), afzel.noore@mail.wvu.edu (A. Noore).

0262-8856/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2010.01.009

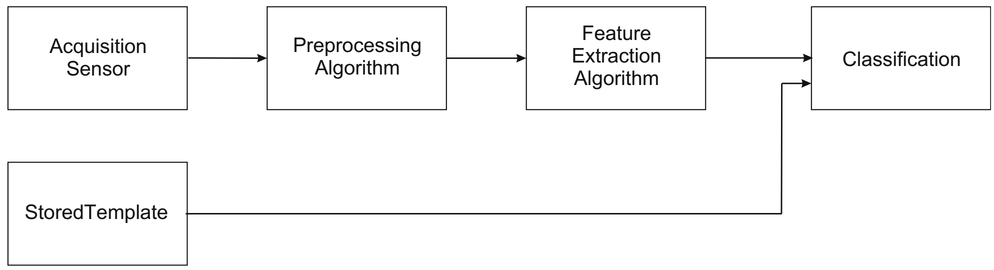

Fig. 1. Block diagram representing the modules of a biometric system that may require regular update or re-training.

present incremental learning for face recognition algorithms such as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA), and results have shown the usefulness of incremental algorithms. Considering the high applicability of this concept in biometrics, this paper presents the concept of online learning for biometric classifier training and update.

The contribution of the research lies in incorporating soft labels [9] and granular computing in the formulation of a 2 ν-Support Vector Machine (2ν-SVM) [10] to design a 2 ν-Granular Soft Support Vector Machine (2ν-GSSVM) and applying the online learning concept for classifier update. The introduction of soft labels and granular computing increases the classification performance while online learning enables the classifier to update its knowledge as database enrollment increases. The performance of the algorithm is evaluated in the context of a *near infrared (NIR) face verification application*. Facial features are computed using Principal Component Analysis [11], Linear Discriminant Analysis [12] and modified C2 features [13]. Experiments performed on a heterogeneous NIR face database indicate that the proposed online learning based classification algorithm not only improves the accuracy but also reduces the training time significantly.

2. Why online learning is applicable in biometrics?

Online learning is a concept from machine learning and inspired by human cognition. Humans are very good in adapting to the environment and learning concepts based on new information. Starting from childhood, a natural learning process constantly trains the human mind to develop new abilities and sustain old (useful) behavior. It also learns to remove behaviors or patterns that become extraneous and redundant over time. Thus, online learning has two components namely incremental and decremental learning.

Mathematically, we have the instance space I consisting of patterns, label space Y indicating the classes of patterns, and a learning based classifier $\Im$. The task of the classifier $\Im$ is to map I–Y using training examples. In general, there are two modes of learning: (1) offline/batch learning mode and (2) online learning mode. In batch mode, a set of training samples $(x_i, y_i) \in I \times Y$ $(i = 1, \ldots, N$ and N is the number of training samples) are used as input. The learning classifier, $\Im_B$ (B denotes batch mode), attempts to obtain a robust/optimal solution such that $\Im_B(x_i) = y_i$. Unlike batch learning, in the online learning mode, training samples become available in a sequential manner. In other words, at time j, training sample (x_j, y_j) and previously processed samples $(x_1, y_1), \ldots, (x_{j-1}, y_{j-1})$ are used to train the learning classifier $\Im_O$ (O denotes online mode) by predicting $y'_j = \Im_O(x_j)$ and comparing it with the true label y_j. The objective of an online learning classifier is to accommodate new training samples and minimize the error over the whole training sequence. It is intuitive that online learning is pertinent to real world biometric applications where training samples are available sequentially (one at a time) and data representation must be updated for optimal classification performance. Another important aspect of online learning is that for a large scale application, it is computationally efficient to solve the offline/batch training process in an online manner [2].

In large scale real world biometrics applications, it is challenging and computationally complex to train a classifier in advance. Without re-training, the disparate characteristics of additional biometric data can cause the performance to degrade. However, training the classifier dynamically online facilitates updated learning in real-time by reducing the computational cost. In a typical biometric system, online learning can be applied to any module that requires training and learning. *In this paper, we propose a* 2ν*-Online Granular Soft Support Vector Machine (*2ν*-OGSSVM) and apply it in the classification stage of the biometric system.*

3. Formulation of 2ν-Online Granular Soft Support Vector Machine

In general, an SVM is used for binary classification problems. Let $\{\mathbf{x}_i, y_i\}$ be the set of N data vectors where $i = 1, \ldots, N$, $\mathbf{x}_i \in \Re^d$ and y_i is the hard label such that $y_i \in (+1, -1)$. The basic principle of SVM is to find the hyperplane that separates the two classes with the widest margin, i.e. $\mathbf{w}\varphi(\mathbf{x}) + b = 0$ to minimize,

$$\begin{aligned} &\tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \psi_i \\ &\text{subject to } y_i(\mathbf{w}\varphi(\mathbf{x}_i) + b) \geqslant (1 - \psi_i), \quad \psi_i \geqslant 0 \end{aligned} \tag{1}$$

where b is the offset of the decision hyperplane, $\mathbf{w}$ is the normal weight vector, $\varphi(x)$ is the mapping function used to map the data space to the feature space and provide generalization for the decision function. C is the regularization factor between the total distance of each error from the margin and the width of the margin, and ψ_i is the slack variable used for classification errors [14]. The optimal SVM parameters are obtained by manually setting the parameters until an optimal error rate is achieved. This heuristic process is very time consuming. Dual ν-SVM (2ν-SVM), originally proposed by [10], is a computationally efficient variant of SVM. It is more flexible in the training and overcomes the issues when the training class sizes are not same. In Eq. (1), additional class dependent parameters (ρ, ν and C_i) are introduced such that the formulation becomes,

$$\begin{aligned} &\min \quad \left\{\tfrac{1}{2}\|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i)\right\} \\ &\text{subject to } \; y_i(\mathbf{w}\varphi(\mathbf{x}_i) + b) \geqslant (\rho - \psi_i), \quad \rho, \; \psi_i \geqslant 0 \end{aligned} \tag{2}$$

where ρ is the position of the margin and ν is the error parameter that can be calculated using ν_+ and ν_- which are the error parameters for training the positive and negative classes, respectively.

$$\nu = \frac{2\nu_+\nu_-}{\nu_+ + \nu_-}, \quad 0 < \nu_+ < 1 \quad \text{and} \quad 0 < \nu_- < 1 \tag{3}$$

$C_i(\nu\rho - \psi_i)$ is the cost of errors and C_i is the error penalty for each class which is calculated as,

$$C_i = \begin{cases} C_+, & \text{if } y_i = +1 \\ C_-, & \text{if } y_i = -1 \end{cases} \quad (4)$$

where

$$C_+ = \frac{\nu}{2n_+\nu_+}$$
$$C_- = \frac{\nu}{2n_-\nu_-} \quad (5)$$

Here n_+ and n_- are the number of training points for the positive and negative classes respectively. Further, the 2ν-SVM objective function can be formulated as (Wolfe Dual formulation),

$$L = \sum_i \alpha_i - \left\{ \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \quad (6)$$

where $i, j \in 1, \ldots, N$, $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function [14], α_i, α_j are the Lagrange multipliers such that $0 \leqslant \alpha_i \leqslant C_i$, $\sum_i \alpha_i y_i = 0$, and $\sum_i \alpha_i \geqslant \nu$.

During training, it is possible that some of the data points may be noisy or incorrectly labeled. In such cases, like any classifier, 2ν-SVM performs erroneous classification. To address this limitation, the formulation of 2ν-SVM is extended to include soft labels [9]. Tao et al. have shown that the use of soft labels not only reduces the classification error but also decreases the number of stored support vectors. Let z_i be the soft label for the ith training data x_i. 2ν-SVM is transformed into 2ν-Soft SVM (2ν-SSVM) as follows:

$$\begin{aligned} &\min \quad \left\{ \tfrac{1}{2} \|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i) \right\} \\ &\text{subject to} \quad z_i(\mathbf{w}\varphi(\mathbf{x}_i) + b) \geqslant z_i^2(\rho - \psi_i) \end{aligned} \quad (7)$$

Even with 2ν-SSVM, training a large database is still time consuming. Granular computing [15] is based on a divide-and-conquer approach and is aimed at reducing the computational time as well as increasing the adaptability to data distribution both locally and globally. In the proposed 2ν-Granular SSVM, the data space is divided into c subspaces with one 2ν-SSVM operating on each subspace. This subspace division is performed using standard fuzzy C-means clustering that generates clusters of similar patterns. Let 2ν-SSVM$_i$ represent the ith 2ν-SSVM, and $2\nu\text{-SSVM}_i :\rightarrow L_i$ represent the 2ν-SSVM operating on the ith subspace ($i = 1, 2, \ldots, c$). The compound margin width W is computed using (8).

$$\begin{aligned} W &= \left| \sum_{i=1}^{c} \frac{t_i}{t} (2\nu SSVM_i :\rightarrow L_i) - L_0 \right| \\ t &= \sum_{i=1}^{c} t_i \end{aligned} \quad (8)$$

where t_i is the number of training data in the ith subspace. 2ν-SSVM learning yields L_i at the local level and L_0 is obtained by learning another 2ν-SSVM on the complete feature space at the global level. Eq. (8) provides the margin width associated with the 2ν-GSSVM hyperplane.

3.1. 2ν-Online GSSVM

Support Vector Machines, including the proposed 2ν-GSSVM, are trained using a training database and evaluation is performed using a test database. Several applications including biometrics, that use SVM as a classifier, require re-training at regular intervals to accommodate the changes in data distribution. Re-training the SVM every time is computationally expensive and may not be feasible for real-time applications. In this paper, we propose an online learning scheme for 2ν-GSSVM termed as 2ν-OGSSVM. The main concept behind the proposed approach is to first construct the decision hyperplane using an initial training dataset and then re-train the classifier by incorporating the new training data points into the decision hyperplane. It also removes unnecessary and irrelevant data so that the number of support vectors does not increase drastically with the increase in training samples. *Thus, the proposed online learning algorithm includes both incremental and decremental learning*. In this process, the Karush–Kuhn Tucker conditions [16] are maintained so that the 2ν-OGSSVM provides an optimal decision hyperplane. The training procedure of 2ν-OGSSVM is as follows:

(1) 2ν-GSSVM is trained using an initial training database and a decision hyperplane is obtained with m support vectors.
(2) For each new training data $\bar{x}_i$,
   (a) $\bar{x}_i$ is classified using the trained 2ν-GSSVM.
   (b) The classification output is compared with the associated label $\bar{z}_i$; if the classification is correct then re-training is not required.
   (c) Otherwise,
      (i) The decision hyperplane is re-computed using the m trained support vectors and $\{\bar{x}_i, \bar{z}_i\}$ via a standard batch training model.[1]
      (ii) After recomputing the hyperplane, the number of support vectors increases. If the number of support vectors is more than $m + \lambda$, where λ is a threshold that controls number of support vectors, then a support vector that is farthest from the current decision hyperplane is selected.
      (iii) The selected support vector is removed from the list of support vectors and stored in the list, l. The classifier with $m + \lambda - 1$ support vectors is used for validation and testing.
(3) The support vectors in the list l are used to test the new classifier. If there is any misclassification, Step 2(c) is repeated to minimize the classification error.
(4) The least recently included support vectors are removed from the list, l, in the final classifier.

It is worthwhile noting that for some cases, it may be required to have an exit condition at Step 3 to avoid any deadlock. One possible exit condition is to include the sample causing the deadlock as a support vector. However, in our case study we did not encounter any deadlock. Fig. 2 shows an example of the decision boundary generated using the proposed 2ν-OGSSVM approach. The two classes in the example represent the genuine and impostor match scores obtained from a face verification algorithm. Fig. 2a shows the result with 200 training samples. In this case, the 2ν-SVM uses all 200 training samples as input for offline or batch learning, while the proposed 2ν-OGSSVM uses the 200 training samples to perform the initial training for online learning. Fig. 2b compares the performance when the number of training samples is increased to 400. In this case, the 2ν-SVM uses all 400 training samples for offline or batch learning, while the 2ν-OGSSVM updates the previously trained classifier one sample at a time with the remaining 200 samples for online learning. The figure shows that the proposed 2ν-OGSSVM efficiently performs binary classification.

4. Application of 2ν-OGSSVM to Near Infrared Face Verification

Face verification is a long standing problem in computer vision and researchers have proposed several algorithms to address the challenges of pose, expression and illumination [17–19]. Recent

[1] Number of samples in this recalculation is small and hence computational time is relatively less compared to computing hyperplane with old and new training data combined.

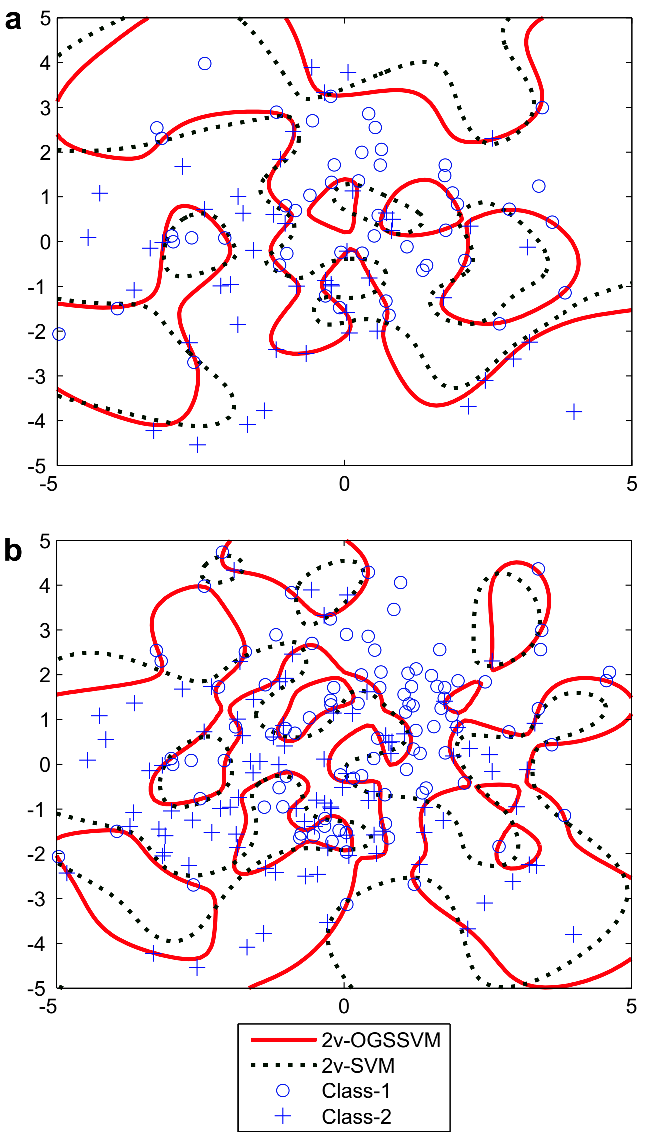

Fig. 2. Comparison of non-linear decision boundary generated with offline learning using 2ν-SVM and online learning using the proposed 2ν-OGSSVM. (a) No. of training samples = 200. (b) No. of training samples = 400.

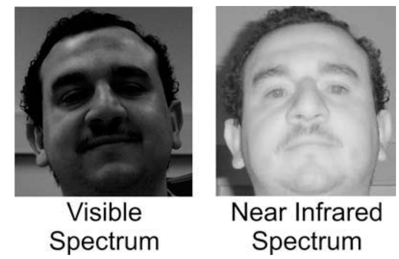

Fig. 3. Visible spectrum and NIR spectrum face images of the same individual under same illumination conditions.

studies have shown the usefulness of near infrared face images for verification [20,21]. Previous research results also suggest that the performance of face verification with NIR images can be better than visible spectrum images in some cases [21]. This is because, as shown in Fig. 3, the near infrared spectrum, generally, provides resilience to varying image quality, changes in illumination, and minor variations in expression. Therefore, in this case study, NIR face verification is used to evaluate the performance of the proposed 2ν-OGSSVM classifier.

Fig. 4 shows the steps involved in NIR face verification. Since the proposed 2ν-OGSSVM is a binary classifier, it can be used for biometric verification. Firstly, the classifier is trained for binary classification (genuine and impostor). Next, at the probe level, facial features are extracted from the NIR face image using feature extraction algorithms (PCA/LDA/C2). Match scores are computed using the Mahalanobis distance measure by comparing the extracted features against the stored features. Finally, classification is performed using the trained classifier.

4.1. Training 2ν-OGSSVM

Let the input training data be $\{\mathbf{x}_i, z_i\}$ where $i = 1, \ldots, N$. N is the total number of training samples and $\mathbf{x}_i$ is the ith match score. z_i is the soft class label and labeling is performed using the density estimation approach [22] similar to Tao et al.'s method [9]. For the ith training sample, likelihood ratio $P_i = \frac{g_{gen}(x_i)}{g_{imp}(x_i)}$ is computed where g_{gen} and g_{imp} represent the marginal densities of genuine and impostor scores, respectively. If $P \geqslant \epsilon$ (where ϵ is a very small value and samples belong to the genuine class if the condition holds) then the value of 'P' is assigned to the soft label with a '+' sign otherwise, it is assigned to the soft label with a '$-$' sign. 2ν-OGSSVM is trained using the radial basis function (RBF) kernel (=$exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2)$). The output of the trained 2ν-OGSSVM is a non-linear decision hyperplane that can classify genuine and impostor match scores.

4.2. Probe classification and decision making

At the probe level, the trained 2ν-OGSSVM is used to classify the match scores. The match score obtained by matching the probe and gallery facial features, x_p, is provided as input to the mixture model and the likelihood ratio $P_p = \frac{g_{gen}(x_p)}{g_{imp}(x_p)}$ is computed. Finally, the trained

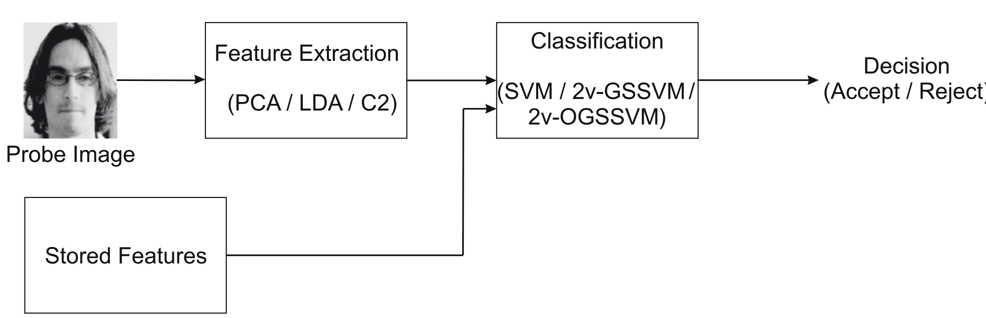

Fig. 4. Illustrating the steps involved in the NIR face verification case study.

Fig. 5. Sample NIR face images from the WVU database [13].

Table 1
Composition of the heterogeneous NIR face database.

Face database	Number of subjects	Number of images per subject	Number of images
CBSR	197	20	3940
Equinox	91	8–16	1307
WVU	40	14	560
Total	**328**	–	**5807**

2ν-OGSSVM is used for classifying the likelihood ratio and a decision of accept or reject is made.

4.3. Experimental protocol

The images from three NIR face databases are combined to create a relatively large database with heterogeneous characteristics. As shown in Table 1, the CBSR NIR face database contains 3940 images pertaining to 197 subjects, the WVU database [13] contains 560 images belonging to 40 subjects, and the Equinox database has 1307 images from 91 subjects. Fig. 5 shows sample NIR face images from the WVU database. The CBSR and Equinox databases contain images that are captured under varying illumination and expression whereas the WVU database contains images with variations in pose, illumination and minor expression. Combining these databases provides a wide range of inter-class and intra-class variations that typically occur in real world face verification applications. Size of images in these databases vary significantly, therefore, we detect face images using OpenCV face detector and normalize the detected face images to a size of 320×240. The database is divided into training database and testing database. The training database contains four images of each individual, i.e. a total of 1312 images, and the remaining 4495 images are used as the testing or probe dataset.

For feature extraction and evaluation, three different algorithms are used. The first facial feature extraction algorithm is the appearance based PCA algorithm [11] and the second algorithm is LDA algorithm [12]. The third algorithm is the modified C2 feature extraction algorithm [13] which is a local texture feature based approach. This algorithm uses 2D log-polar Gabor features and a visual cortex model to extract facial features. The training database is used to train these feature extraction algorithms and verification is performed using a two-class classifier such as SVM, 2ν-GSSVM and 2ν-OGSSVM. To evaluate the efficacy of the proposed approach, the experimental comparison of SVM, 2ν-SVM, and 2ν-GSSVM with online classification (2ν-OGSSVM) is performed. For SVM, 2ν-SVM, and 2ν-GSSVM, the complete training database is used to train the classifier (batch learning mode). On the other hand, the to evaluate the performance of the proposed 2ν-OGSSVM classifier, it is initially trained with 100 subjects and then online learning is performed with training samples from the remaining 228 subjects, i.e. online training with one sample at a time. Further, the train-test partitioning is performed 20 times for cross validation and ROC curves are generated by computing the false reject rate (FRR) over these trials at different false accept rates (FAR). Finally, verification accuracies are reported at 0.01% FAR.

4.4. Experimental results

In all the experiments, RBF kernel with $\gamma = 6$ was used for the SVM, 2ν-SVM, 2ν-GSSVM and 2ν-OGSSVM classifiers.[2] In the first case, dimensionality of PCA subspace and PCA coefficients were extracted using a standard approach as described in [12] and matching was performed using the Mahalanobis distance measure. The match score was further classified using these classifiers separately. The verification performance of PCA with non-linear classification was also compared with the traditional PCA algorithm [11] (that uses Euclidean distance and a linear decision threshold for classification) and incremental PCA (IPCA) algorithm [6]. This experiment provides the baseline performance to facilitate systematic comparison with other algorithms. Furthermore, accuracies were also computed when IPCA was combined with the proposed 2ν-OGSSVM (i.e. IPCA + 2ν-OGSSVM, both feature extraction and classification algorithms were online). Similar to the PCA experiments, in the second case, LDA [12] and incremental LDA (ILDA) [7] were used for feature extraction and SVMs were used for classification. Here also, performance of ILDA + 2ν-OGSSVM was compared with their batch mode counterpart. Finally, in the third case, modified C2 features were classified using the three classifiers separately. The ROC plots in Figs. 6–8, and Tables 2–4 show the experimental results for comparison. The key results and analysis are summarized below:

- Baseline PCA (batch/offline learning mode) and IPCA (online learning mode) provide similar verification performance, but the advantage of IPCA is reduced computational cost[3] (Table 5). Specifically, the training time of IPCA is significantly lower compared to PCA. Experimentally, this shows that the online learning approach is a faster alternative for biometrics applications.
- PCA with SVM classifier yields the verification accuracy of 62.23%, whereas with 2ν-GSSVM, the verification accuracy improves by 8.9%. This suggests that incorporating granular computing and soft labels improves the classification performance. Granular computing makes it adaptive to variations in data distribution and soft labels provide resilience to noise.
- If the feature extraction algorithm and classification algorithm are in online learning mode, i.e. IPCA + 2ν-OGSSVM, verification accuracy is about 0.1% greater than when both the algorithms are in batch learning mode (PCA + 2ν-GSSVM). However, the training time in online learning mode, as shown in Table 5 is about half of the batch learning mode which shows the suitability of online learning algorithms for large scale applications.
- Similar to PCA experiments, LDA experiments also show advantages of online learning (Tables 3 and 5). Here the important point is that in our case study on NIR face verification, accuracies in LDA experiments are better than PCA experiments. This strengthens existing analysis that, in general, LDA yields better accuracy than PCA.
- Similarly, improvements are observed in the case of the modi-

[2] In this case study, we observed that $\gamma = 6$ provided the best verification accuracy for all variations of SVM across all cross validations.

[3] Time was computed on a 2 GHz Pentium Duo Core processor with 2 GB RAM under MATLAB environment.

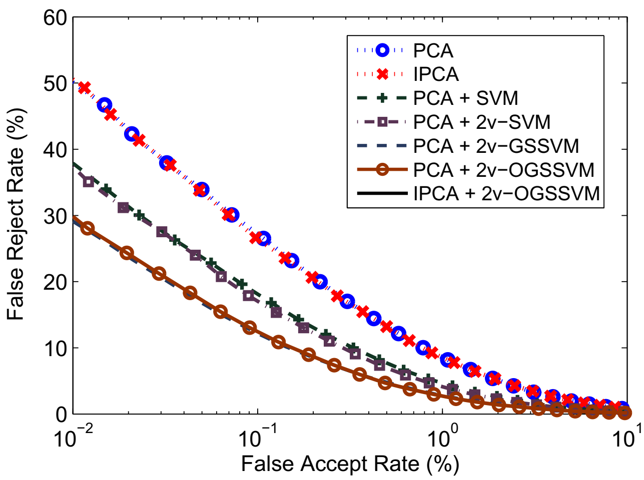

Fig. 6. Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using appearance based PCA/IPCA algorithm.

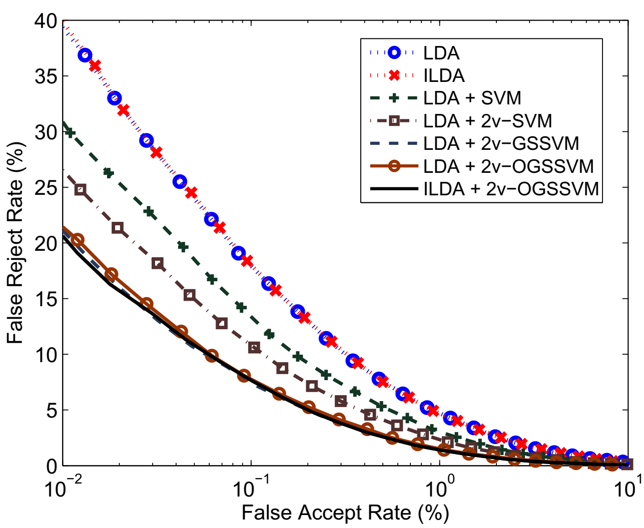

Fig. 7. Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using appearance based LDA/ILDA algorithm.

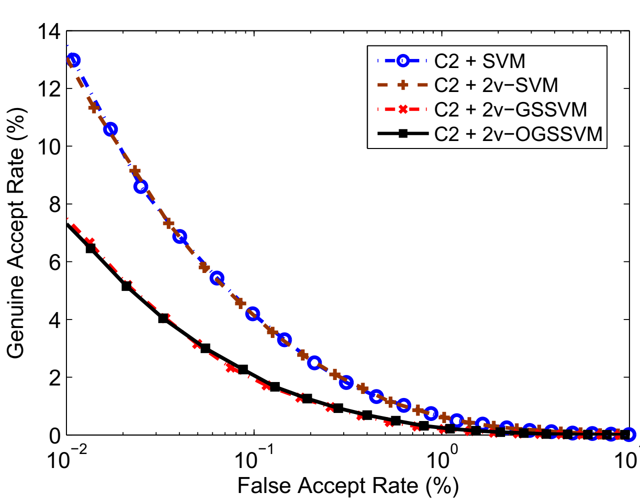

Fig. 8. Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using local texture feature based modified C2 algorithm [13].

Table 2
Covariate analysis of PCA and IPCA based verification algorithms with multiple classifiers. Verification accuracies are computed at 0.01% FAR.

Feature extraction and classification	Covariate			
	Expression	Illumination	Pose	Overall
PCA [11]	51.87	52.24	45.46	49.21
PCA + SVM	64.76	65.34	58.17	62.23
PCA + 2ν-SVM	64.98	65.83	58.41	62.95
PCA + 2ν-GSSVM	73.63	74.04	67.86	71.14
PCA + 2ν-OGSSVM	72.82	73.95	67.49	70.97
IPCA [6]	51.73	52.21	45.02	49.17
IPCA + 2ν-OGSSVM	**73.67**	**74.11**	**67.92**	**71.22**

Table 3
Covariate analysis of LDA and ILDA based verification algorithms with multiple classifiers. Verification accuracies are computed at 0.01% FAR.

Feature extraction and classification	Covariate			
	Expression	Illumination	Pose	Overall
LDA [12]	62.01	62.43	54.97	60.82
LDA + SVM	70.84	71.28	58.61	69.01
LDA + 2ν-SVM	70.97	71.58	59.08	73.85
LDA + 2ν-GSSVM	80.75	81.13	71.06	79.04
LDA + 2ν-OGSSVM	80.71	81.15	70.81	78.99
ILDA [7]	61.82	62.19	54.13	60.76
ILDA + 2ν-OGSSVM	**80.83**	**81.20**	**70.98**	**79.35**

Table 4
Covariate analysis of modified C2 based verification algorithm with multiple classifiers. Verification accuracies are computed at 0.01% FAR.

Feature extraction and classification	Covariate			
	Expression	Illumination	Pose	Overall
C2 + SVM	87.76	87.93	85.17	86.94
C2 + 2ν-SVM	87.79	88.02	85.59	87.21
C2 + 2ν-GSSVM	92.67	92.85	91.81	92.46
C2 + 2ν-OGSSVM	**92.79**	**92.88**	**92.03**	**92.71**

Table 5
Computational time analysis for the proposed 2ν-OGSSVM and comparison with other classification approaches.

Feature extraction + classification	Computation time	
	Training time (min)	Testing time (s)
PCA [11]	43.6	0.5
PCA + SVM	221.5	1.7
PCA + 2ν-SVM	194.8	1.4
PCA + 2ν-GSSVM	118.2	1.1
PCA + 2ν-OGSSVM	58.4	0.8
IPCA [6]	31.2	0.4
IPCA + 2ν-OGSSVM	**52.5**	**0.7**
LDA [12]	48.8	0.6
LDA + SVM	230.1	1.8
LDA + 2ν-SVM	198.4	1.6
LDA + 2ν-GSSVM	122.6	1.2
LDA + 2ν-OGSSVM	61.9	0.9
ILDA [7]	37.2	0.5
ILDA + 2ν-OGSSVM	**55.1**	**0.8**
C2 + SVM	336.8	2.1
C2 + 2ν-SVM	282.1	1.8
C2 + 2ν-GSSVM	192.4	1.6
C2 + 2ν-OGSSVM	**109.7**	**1.2**

fied C2 feature algorithm also. Since the modified C2 algorithm [13] originally uses SVM for classification and decision making, the comparative study is performed with different variants of SVM. Compared to the SVM classifier, 2ν-OGSSVM not only improves the verification accuracy by 5.77% but also reduces training time by three times.

- From Tables 2–4, the covariate analysis with respect to variations in expression, illumination and pose show that the pose variations cause a large reduction in the accuracy of appearance based PCA and LDA algorithms. On the other hand, the local texture feature based modified C2 algorithm provides consistent performance for all three variations. The experiments also show that the modified C2 algorithm with 2ν-OGSSVM classifier yields the verification accuracy of more than 92% which is around 21.7% better than PCA with 2ν-OGSSVM classifier and 13.4% better than LDA with 2ν-OGSSVM classifier.
- For the three feature extraction algorithms (PCA, LDA and modified C2 feature algorithms), verification accuracies of the proposed 2ν-OGSSVM are slightly better than 2ν-GSSVM classifier. However, advantage of 2ν-OGSSVM is significant reduction in computational time. Compared to classical SVM, 2ν-GSSVM reduces the training time significantly because the dual-ν formulation requires less time for parameter estimation and the granular computing approach reduces the time by dividing the problem into subproblems and solving it efficiently both in terms of accuracy and time. With online learning approach (i.e. 2ν-OGSSVM), the training time is further reduced because initial training with 100 subjects requires limited computational time (for instance, only 36.1 min are required in the case of PCA) and then a relatively small amount of time (22.3 min in the case of PCA) is required to train the remaining 228 subjects in online mode (thus the total training time for PCA + 2ν-OGSSVM is 58.4 min). Furthermore, we observed that the computational time required for each update in online mode is small (around 6 s). On the other hand, batch mode/offline algorithms have to be re-trained using both old and new training data which increases computation cost.
- From Table 5, even though the PCA, LDA, and modified C2 feature extractors are not trained in online mode, only the online re-training of the classifier improves performance considerably. When the feature extractors are also trained in online mode, i.e. IPCA + 2ν-OGSSVM and ILDA + 2ν-OGSSVM, computational time is further decreased.
- Table 5 also shows that the testing time is considerably reduced when 2ν-OGSSVM is used as the classifier. The reason for this improved performance is same as that for the reduced training time.
- As mentioned in Section 2, another possible advantage of online learning is to efficiently solve the offline training process in an online manner. To evaluate the appropriateness of this statement in biometric classifier training, a comparative analysis of verification accuracies obtained from 2ν-GSSVM and 2ν-OGSSVM is performed. Fig. 9 shows that at the end of the learning process with 328 subjects, verification accuracies obtained in the online learning mode are similar to the accuracy obtained in the batch learning mode. Further, the time required for online learning is around half of that of offline/batch training. Therefore, it is possible to apply online learning scheme in preference to offline training scheme without affecting the verification accuracy.
- In our experiments, we analyzed the effect of number of support vectors on the verification accuracy across different cross validations. For this case study, at 0.01% FAR, we observed that the number of support vectors in classical SVM batch mode learning was around 1.6 times greater than 2ν-GSSVM and around 2.3 times greater than 2ν-OGSSVM. Specifically, across different cross validation trials, the average number of support vectors in (1) PCA + 2ν-OGSSVM case study is 316, (2) LDA + 2ν-OGSSVM case study is 292 and (3) C2+2ν-OGSSVM case study is 327.

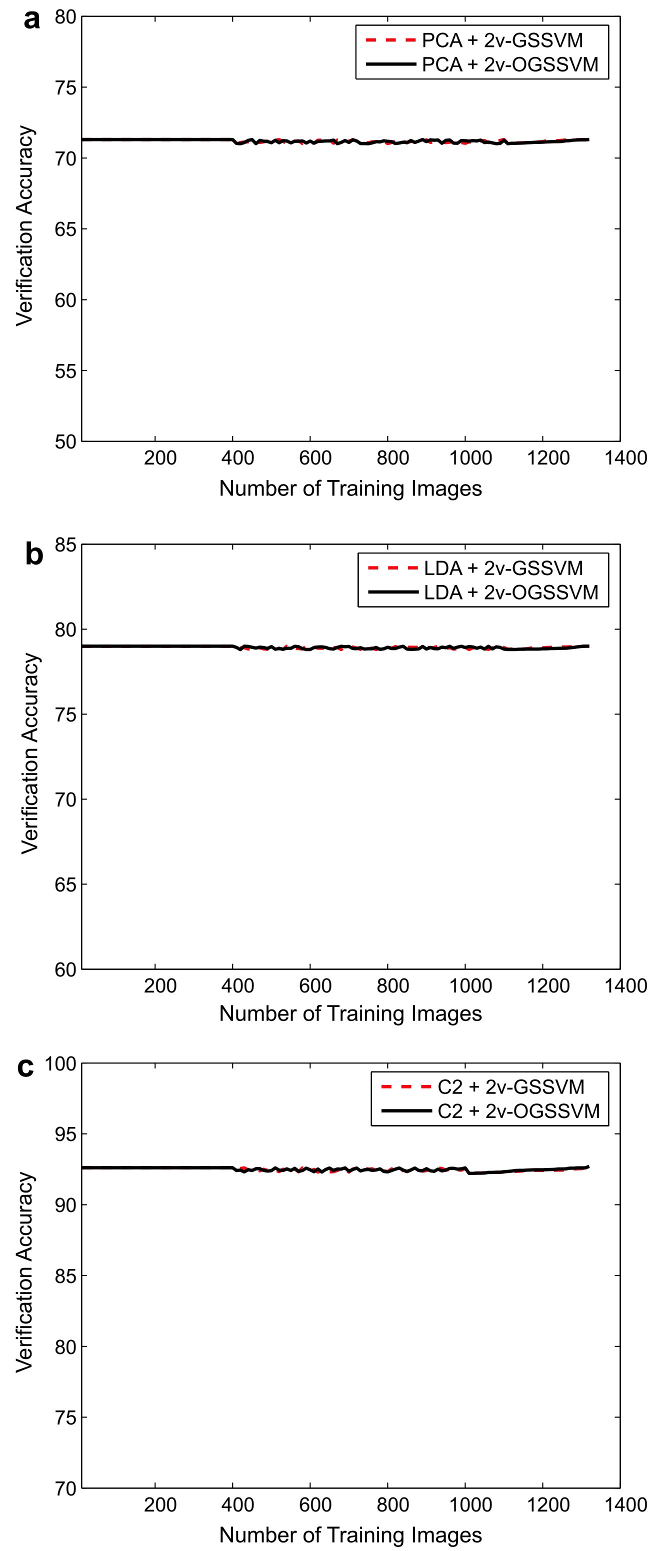

Fig. 9. Verification accuracy when adding new training samples (i.e. new classes and new class instances): (a) PCA, (b) LDA and (c) modified C2 algorithms. Note that x-axis represents 1312 training images pertaining to 328 subjects and first 400 images pertaining to 100 subjects are the initial training images.

Table 6
Comparison of 2ν-OGSSVM with existing online SVMs. Verification accuracies are computed at 0.01% FAR.

Feature extraction algorithm	Online classifier		
	Relaxed online SVM [8]	Incremental–decremental SVM [1]	Proposed 2ν-OGSSVM
IPCA	65.81	66.53	**71.22**
ILDA	71.48	71.97	**79.35**

- We performed another experiment in which performance of the proposed 2ν-OGSSVM algorithm was compared with existing online SVMs namely incremental–decremental SVM [1] and relaxed online SVM [8]. For this comparison we use IPCA and ILDA as online feature extraction algorithms. Under the same experimental protocol, Table 6 illustrates the results that clearly show that the proposed algorithm outperforms existing algorithm by at least 4.7%. Also, we observe that the proposed algorithm is around 1.3 times faster than existing algorithms. We observed that this improvement is mainly due to the use of granularity and soft labels in the proposed 2ν-OGSSVM classifier.
- At 95% confidence, *t*-test shows that the 2ν-GSSVM is significantly different than the SVM classifier whereas there is no statistical difference between 2ν-GSSVM and 2ν-OGSSVM. However, as mentioned previously, the main advantage of 2ν-OGSSVM is reduced computational time and online classifier update.

The experiments demonstrate that online learning approach reduces the computational cost without compromising the verification accuracy. Therefore, it is an effective alternative to traditional batch/offline learning methods.

5. Conclusion and future work

Similar to template update, the parameters of the classifiers used in biometric system also require to be updated to accommodate variations in data distribution. Current systems frequently retrain the algorithms using all enrolled subjects. This process may not be feasible for large scale systems where the number of newly enrolled subjects is significantly high. This paper introduces the concept of online learning in biometrics to address the problem of classifier re-training and update. An online learning scheme for 2ν-GSSVM is proposed to train the classifier in online mode so that it can update the decision hyperplane according to the newly enrolled subjects. This online classifier is used for feature classification and decision making in a face verification system. On a heterogeneous NIR face database, the case study using PCA, LDA and modified C2 feature algorithms shows that the proposed online classifier significantly improves the verification performance both in terms of accuracy and computational time. Indeed, it is observed that the proposed online classifier is at least three times faster than the conventional SVM classifier.

The emergence and viability of using online learning algorithms in the design of biometric classifiers addresses the real-time performance and scalability challenges; however more research is required in order to fully understand the benefits in large scale applications. It is our assertion that the promising research results of this work would stimulate further research in this important area. Currently, we are investigating possibilities of applying online learning concept for biometrics template update.

Acknowledgments

The authors would like to thank the reviewers and guest editors for their constructive and useful comments. The authors would like to thank Equinox Corporation, IEEE OTCBVS WS Series Bench, Center for Biometrics and Security Research, and AuthenMetric Co. Ltd. for granting us access to the face databases used in this research. Singh, Vatsa and Noore were supported in part through a Grant (Award No. 2003-RCCX-K001) from the National Institute of Justice, United States Department of Justice. Ross was supported by US National Science Foundation CAREER Grant No. IIS 0642554.

References

[1] G. Cauwenberghs, T. Poggio, Incremental and decremental support vector machine learning, in: International Conference on Neural Information Processing Systems, 2000, pp. 409–415.
[2] J. Kivinen, A. Smola, R. Williamson, Online learning with kernels, IEEE Transactions on Signal Processing 52 (8) (2004) 2165–2176.
[3] D. Tax, P. Laskov, Online SVM learning: from classification to data description and back, in: Proceedings of IEEE International Workshop on Neural Networks for Signal Processing, 2003, pp. 499–508.
[4] D. Skocaj, A. Leonardis, Incremental and robust learning of subspace representations, Image and Vision Computing 26 (1) (2008) 27–38.
[5] P. Auer, N. Cesa-Bianchi, C. Gentile, Adaptive and self-confident on-line learning algorithms, Journal of Computer and System Sciences 64 (2002) 48–75.
[6] B. Raducanu, J. Vitria, Online nonparametric discriminant analysis for incremental subspace learning and recognition, Pattern Analysis and Applications 11 (3–4) (2008) 259–268.
[7] M. Uray, D. Skocaj, P.M. Roth, H. Bischof, A. Leonardis, Incremental LDA learning by combining reconstructive and discriminative approaches, in: Proceedings of British Machine Vision Conference, 2007, pp. 272–281.
[8] D. Sculley, M.G. Wachman, Relaxed online SVMs for spam filtering, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007, pp. 415–422.
[9] Q. Tao, G. Wu, F. Wang, J. Wang, Posterior probability support vector machines for unbalanced data, IEEE Transactions on Neural Network 16 (6) (2005) 1561–1573.
[10] H. Chew, C. Lim, R. Bogner, An implementation of training dual-ν support vector machines, in: L. Qi, K. Teo, X. Yang, Optimization and Control with Applications, 2004.
[11] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[12] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Transaction on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228–233.
[13] R. Singh, M. Vatsa, A. Ross, A. Noore, A mosaicing scheme for pose-invariant face recognition, IEEE Transactions on Systems, Man and Cybernetics-Part B 37 (5) (2007) 1212–1225.
[14] V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.
[15] A. Bargiela, W. Pedrycz, Granular computing: an introduction. International Series in Engineering and Computer Science, 2002.
[16] R. Fletcher, Practical Methods of Optimization, second ed., Wiley, 1987.
[17] S.Z. Li, A.K. Jain, Handbook of Face Recognition, Springer, 2005.
[18] H. Wechsler, Reliable Face Recognition Methods: System Design, Implementation and Evaluation, Springer, 2006.
[19] W.-Y. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Survey 35 (4) (2003) 399–458.
[20] S. Li, R. Chu, S. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 627–639.
[21] R. Singh, M. Vatsa, A. Noore, Hierarchical fusion of multi-spectral face images for improved recognition performance, Information Fusion 9 (2) (2008) 200–210.
[22] Z. Zivkovic, F.V.D. Heijden, Recursive unsupervised learning of finite mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 651–656.