Proc. of IEEE 3rd Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS 09), Washington DC, September 2009

Online Learning in Biometrics: A Case Study in Face Classifier Update

Richa Singh, Mayank Vatsa, Arun Ross, and Afzel Noore

Abstract— **In large scale applications, hundreds of new subjects may be regularly enrolled in a biometric system. To account for the variations in data distribution caused by these new enrollments, biometric systems require regular re-training which usually results in a very large computational overhead. This paper formally introduces the concept of online learning in biometrics. We demonstrate its application in classifier update algorithms to re-train classifier decision boundaries. Specifically, the algorithm employs online learning technique in a 2ν-Granular Soft Support Vector Machine for rapidly training and updating face recognition systems. The proposed online classifier is used in a face recognition application for classifying genuine and impostor match scores impacted by different covariates. Experiments on a heterogeneous face database of 1,194 subjects show that the proposed online classifier not only improves the verification accuracy but also significantly reduces the computational cost.**

I. INTRODUCTION

In the literature, it is well understood that a carefully designed biometric system should be stable and not susceptible to environmental dynamics, variations in data distribution, and increasing database size. However, in large scale applications such as National ID/passport, US VISIT and FBI IAFIS, scalability and variations in data distribution play an important role. In these applications, hundreds of individuals can be enrolled on a regular basis. As the size of the biometric database increases, the classifier may have to be re-trained in order to handle the variations introduced due to the newly enrolled subjects. A generic biometric system, as shown in Fig. 1, has five stages where regular update or re-training is required. Template update is required to address the issue of template aging; sensor update is required to keep up with advancements in sensor technology; updating the preprocessing and feature extraction modules is necessary to handle changes in sensors or ambient conditions; and classifier update is needed to modify the linear/non-linear decision boundary that is used to determine if image pairs belong to the genuine or impostor classes. For example, in a face recognition system, template update is required to address facial aging while classifier update is needed to update the decision boundary based on changes in intra-class and inter-class variations due to new enrollments. It is computationally complex to re-train a large scale biometric system where the size and contents of the database size are regularly changing. Without re-training, the disparate characteristics of additional biometric data can cause the performance to degrade.

Online learning, a well defined concept in machine learning, provides a solution to the high computational complexity associated with offline learning/training process. It is inspired by the learning process of the human mind that continually adapts to its environment. A natural online learning process constantly trains the human mind to (1) acclimatize to the environment, (2) learn new concepts, (3) retain useful behavior, and (4) remove redundant or superfluous information. These properties illustrate that online learning consists of incremental (add new knowledge) and decremental (remove spurious/redundant knowledge) learning.

Online learning has been used in problems related to machine learning [2], [7], [14], but very limited research has been conducted in biometrics. Considering the high relevance, we introduce the concept of online learning for biometric classifier training and update. The contribution of this research lies in incorporating the online learning concept in a 2ν-Granular Soft Support Vector Machine [12] for classifier update. Online learning technique enables the classifier to continually update its knowledge with the increase in database enrollments. The performance of the proposed algorithm is evaluated in the context of face recognition. Facial features are computed using Kernel Fisher Discriminant Analysis (KFDA) [9]. Experiments performed on a heterogeneous face database indicate that the proposed algorithm not only improves the classification performance but also reduces the training time significantly.

II. FORMULATION OF ONLINE LEARNING FOR 2ν-GSSVM

In general, SVM is used for a two-class classification problem. Let $\{\mathbf{x}_i, y_i\}$ be a set of N data vectors where $i = 1, ..., N$, $\mathbf{x}_i \in \Re^d$ and y_i is the hard label such that $y_i \in (+1, -1)$. The basic principle behind 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,

$$\begin{array}{l} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \psi_i \\ \text{subject to } y_i\left(\mathbf{w}\,\varphi(\mathbf{x}_i) + b\right) \geq (1 - \psi_i), \quad \psi_i \geq 0 \end{array} \tag{1}$$

where b is the offset of the decision hyperplane, $\mathbf{w}$ is the normal weight vector, and $\varphi(\mathbf{x})$ is the mapping function used to map the data space to the feature space and provide generalization for the decision function. C is a 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 to allow classification errors [15]. The optimal

R. Singh and M. Vatsa are with the Indraprastha Institute of Information Technology (IIIT) Delhi, India {mayank, rsingh}@iiitd.ac.in

A. Ross and A. Noore are with Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, USA {arun.ross, afzel.noore}@mail.wvu.edu

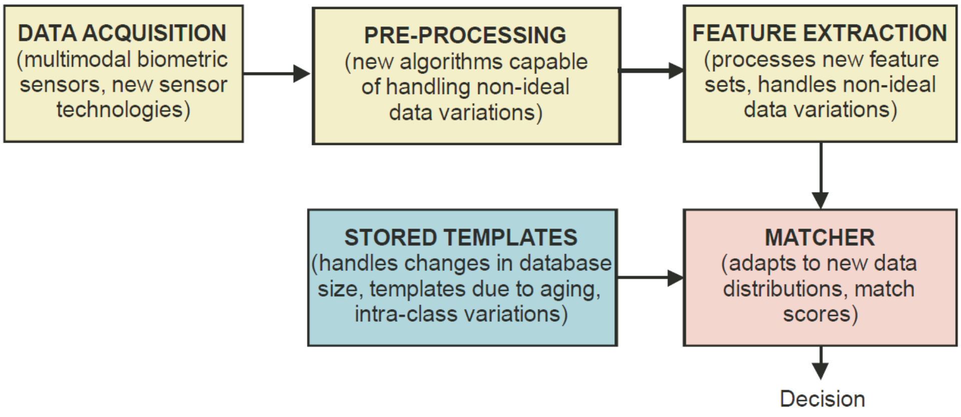
DATA ACQUISITION
(multimodal biometric sensors, new sensor technologies)
PRE-PROCESSING
(new algorithms capable of handling non-ideal data variations)
FEATURE EXTRACTION
(processes new feature sets, handles non-ideal data variations)
STORED TEMPLATES
(handles changes in database size, templates due to aging, intra-class variations)
MATCHER
(adapts to new data distributions, match scores)
Decision

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

SVM parameters are obtained by manually setting the parameters until an optimal error rate is achieved. This is a very time consuming heuristic process. Dual ν-SVM (2ν-SVM), originally proposed by [3], is a computationally efficient variant of SVM. It is more flexible in the training process and overcomes the issues when the training class sizes are not same. In (1), additional class dependent parameters (ρ, ν and C_i) are introduced such that the formulation becomes,

$$\begin{array}{l} min\{\frac{1}{2}\|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i)\} \\ \text{subject to } y_i\,(\mathbf{w}\,\varphi(\mathbf{x}_i) + b) \geq (\rho - \psi_i), \quad \rho,\ \psi_i \geq 0 \end{array} \tag{2}$$

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,

$$C_i = \left\{ \begin{array}{lll} C_+, & if & y_i = +1 \\ C_-, & if & y_i = -1 \end{array} \right. \tag{4}$$

where,

$$\begin{array}{l} C_+ = \frac{\nu}{2n_+\nu_+}, \\ \\ C_- = \frac{\nu}{2n_-\nu_-}. \end{array} \tag{5}$$

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

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

where $i, j \in 1, ..., N$, $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function [15], α_i, α_j are the Lagrange multipliers such that $0 \leq \alpha_i \leq C_i$, $\sum_i \alpha_i y_i = 0$, and $\sum_i \alpha_i \geq \nu$.

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 with the use of soft labels [13]. Tao *et al.* have shown that the soft labels not only reduce the classification error but also decrease the number of stored support vectors. The formulation of 2ν-SVM with soft labels is described below.

Let z_i be the soft label for the i^{th} training data x_i. The soft label is obtained from the posterior probability $p(\cdot|x_i)$ computed using k-NN estimator such that $z_i = 2p(\cdot|x_i) - 1$. Further, for the j^{th} class, $p(j|x) = \frac{k_j}{k}$, where k_j is the number of training data belonging to class j among the k nearest neighbors. 2ν-Soft SVM (2ν-SSVM) is thus formulated as,

$$\begin{array}{l} min\left\{\frac{1}{2}\|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i)\right\} \\ \text{subject to } z_i\,(\mathbf{w}\,\varphi(\mathbf{x}_i) + b) \geq z_i^2(\rho - \psi_i). \end{array} \tag{7}$$

Even with 2ν-SSVM, training a large database is time consuming. Granular computing [1], which is a divide and conquer approach, is used to reduce the computational time as well as to increase the adaptability to data distribution both locally and globally. In 2ν-Granular SSVM (2ν-GSSVM) [12], the data space is divided into c subspaces with one 2ν-SSVM operating on each subspace. Let 2ν-SSVM$_i$ represent the i^{th} 2ν-SSVM, and 2νSSVM$_i$ $:\to L_i$ represent the 2ν-SSVM operating on the i^{th} subspace ($i = 1, 2, ..., c$). The compound margin width W is computed using (8).

$$\begin{array}{l} W = \left|\sum_{i=1}^{c} \frac{t_i}{t}(2\nu SSVM_i :\to L_i) - L_0\right|, \\ \\ t = \sum_{i=1}^{c} t_i \end{array} \tag{8}$$

where t_i is the number of training data in the i^{th} 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. Equation 8 provides the margin width associated with the 2ν-GSSVM hyperplane.

A. Online Learning for 2ν-GSSVM

Support Vector Machines, including 2ν-GSSVM, are trained using the training database and evaluated 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 easily feasible for real time applications. In this paper, we propose an online learning scheme for 2ν-GSSVM which we term 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. In this process, the Karush-Kuhn Tucker conditions [4] are maintained so that the 2ν-OGSSVM provides an optimal decision hyperplane.

The training procedure of 2ν-OGSSVM is as follows:

Algorithm 1 Training procedure of the proposed 2ν-OGSSVM classifier

```
1) $2\nu$-GSSVM is trained using an initial training database
   and a decision hyperplane is obtained.
2) For each new training data $\bar{x}_i$,
   a) $\bar{x}_i$ is classified using the trained $2\nu$-GSSVM.
   b) The classification output is compared with the
      associated label $\bar{z}_i$. If the classification is correct
      then nothing is done.
   c) Otherwise,
      i) The decision hyperplane is recomputed using
         the $m$ trained support vectors and $\{\bar{x}_i, \bar{z}_i\}$.
      ii) After recomputing the hyperplane, the number
          of support vectors increases. If the number
          of support vectors is more than $m + \lambda$, then
          a support vector that is farthest from the
          decision hyperplane is removed and stored in
          another list, $l$.
      iii) 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 are any misclassifications, Step
   2(c) is repeated to minimize the classification error.
4) The least recently included support vectors are re-
   moved from the list, $l$, in the final classifier.
```

III. Case Study in Face Recognition

Face recognition is a long standing problem in computer vision and researchers have proposed several algorithms to address the challenges of pose, expression, illumination, motion and resolution. In this research, we use face recognition as a case study to evaluate the effect of online learning on 2ν-GSSVM. To analyze the performance on a large database with challenging intra-class variations, we combined images from multiple face databases to create a heterogeneous database of more than 116,000 images pertaining to 1,194 subjects. Table I lists the databases used and the number of subjects selected from the individual databases. The CMU-AMP database[1] contains images with large expression variations while the CMU-PIE dataset [11] contains images with variations in pose, illumination and facial expressions. The Equinox database[2] has images captured under different illumination conditions with accessories and expressions. The AR face database [8] contains face images with varying illumination and accessories, and the FERET database [10] has face images with different variations over a time interval of 3-4 years. The Notre Dame face database [5] comprises of images with different lighting and facial expressions over a period of one year. The Labeled Faces in the Wild database [6] contains real world images of celebrities and popular individuals (this database contains images of more than 1,600 subjects from which we selected 294 subjects that have at least 6 images). To the best of our knowledge, there is no single database available in the public domain which encompasses such a wide range of intra-class variations. We partition the images into two sets: (1) the training dataset is used to train the KFDA and individual classifiers, and (2) the gallery-probe dataset (the test set) is used to evaluate the performance of the proposed algorithm. Four images of each subject are randomly selected to comprise the training set. The remaining images are used as the test data to evaluate the algorithms.

TABLE I

Composition of the heterogeneous face database.

Face Database	Number of Subjects
CMU-AMP	13
CMU - PIE	65
Equinox	90
AR	120
FERET	300
Notre Dame	312
Labeled Faces in the Wild	294
Total	**1194**

Fig. 2 shows the steps involved in this case study. Facial features are extracted from the probe face image and classification is performed against the stored features. For feature extraction and evaluation, an appearance based KFDA algorithm[3] [9] is used. The training database is used to train the feature extraction algorithm and verification is performed using a two-class classifier such as SVM, 2ν-SVM, 2ν-GSSVM and 2ν-OGSSVM. We compare the performances of SVM, 2ν-SVM, and 2ν-GSSVM with online classification (2ν-OGSSVM) in order to evaluate the efficacy of the proposed approach. For SVM, 2ν-SVM, and 2ν-GSSVM, the complete training database is used to train the classifier. On the other hand, to evaluate the performance of the proposed 2ν-OGSSVM, the training database is divided into two parts: (1) initial training dataset with 200 subjects and (2) online training dataset with 994 subjects. First the 2ν-OGSSVM classifier is trained with 200 subjects and then online learning is performed with 994 subjects i.e., *online training is done*

[1] http://amp.ece.cmu.edu/projects/FaceAuthentication/d̃ownload.htm

[2] http://www.equinoxsensors.com/products/HID.html

[3] To address the limitation of linear appearance-based algorithms, researchers have adopted kernel approaches for subspace analysis that can capture higher-order statistics for better representation.

one subject at a time. Further, the train-test partitioning is performed 10 times (trials) 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.

A. *Experimental Results*

In all the experiments, the Radial Basis Function (RBF) kernel with RBF parameter $\gamma = 6$ is used for SVM, 2ν-SVM, 2ν-GSSVM and 2ν-OGSSVM classifiers. For KFDA, RBF kernel with $\gamma = 4$ is used[4]. KFDA coefficients are extracted and matching is performed using each of these classifiers separately. The ROC plots in Fig. 3 and verification accuracies in Table II show the experimental results and comparison. The key results and analysis are summarized below:

- KFDA with SVM classifier yields a verification accuracy of 49.03% whereas with 2ν-GSSVM, the verification accuracy improves by around 20%. 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.
- From Table II, the covariate analysis with respect to variations in expression, illumination and pose shows that pose variations cause a large reduction in the accuracy of appearance based KFDA algorithm.
- Verification accuracies of the proposed 2ν-OGSSVM are better than 2ν-GSSVM classifier. However, the main advantage of 2ν-OGSSVM is computational time[5]. As shown in Table III, KFDA with SVM requires 891 minutes for training with 1194 subjects. With 2ν-GSSVM, the training time is reduced to 429 minutes because the dual ν formulation requires less time for parameter estimation and 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 (i.e. 2ν-OGSSVM), the training time is further reduced to 272 minutes. This is because for the initial training with 200 subjects, the algorithm requires only 198 minutes and an additional 74 minutes are required to train the remaining 994 subjects in the online mode. This shows that the online approach significantly reduces the computational cost of re-training the classifier without decreasing the accuracy. Further, the testing time of KFDA is also reduced significantly when 2ν-OGSSVM is used as the classifier.
- Finally, the t-test at 95% confidence shows that for this particular case study, the performance of the four variants of SVM are significantly different from each other. However, as mentioned previously, the main advantages of 2ν-OGSSVM are reduced computational time and regular classifier update.

TABLE II

COVARIATE ANALYSIS OF KFDA BASED RECOGNITION ALGORITHM WITH MULTIPLE CLASSIFIERS. VERIFICATIONS ACCURACIES ARE COMPUTED AT 0.01% FAR.

Feature Extraction Algorithm	Covariate	Classifier			
		SVM	2ν-SVM	2ν-GSSVM	2ν-OGSSVM
KFDA [9]	Expression	50.53	62.17	72.11	**73.84**
	Illumination	52.71	62.01	71.98	**73.20**
	Pose	46.43	55.26	64.71	**68.93**
	Overall	49.03	61.17	69.98	**72.41**

TABLE III

COMPUTATIONAL TIME ANALYSIS FOR THE PROPOSED 2ν-OGSSVM AND COMPARISON WITH SVM, 2ν-SVM AND 2ν-GSSVM.

Feature Extraction Algorithm	Classifier	Computation Time	
		Training Time (Minutes)	Testing Time (Seconds)
KFDA [9]	SVM	891.2	1.8
	2ν-SVM	747.3	1.2
	2ν-GSSVM	429.5	1.0
	2ν-OGSSVM	**272.1**	**0.8**

- In another experiment, the performance of the "standard" LDA feature extraction and SVM classification algorithm is compared against the proposed KFDA and 2ν-OGSSVM. We observe that, under the same experimental setting, training LDA and SVM is approximately 3 times slower than the proposed algorithm. Also, verification accuracy of the proposed algorithm is around 20% better than the LDA and SVM approach. This experiment accentuates the effectiveness of the online learning concept in biometrics.

IV. CONCLUSIONS AND FUTURE WORK

Like template update, the parameters of the classification algorithms used in the biometric system also require regular update to accommodate the variations in data distribution. Current systems frequently re-train the algorithms using all the enrolled subjects. This process may not be feasible for large scale systems where the number of newly enrolled subjects can be significantly high. This paper introduces the concept of online learning in biometrics to address the problem of classifier re-training and update. A formulation of online learning for 2ν-GSSVM is proposed to train the classifier in the 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 recognition system. On a face database of 1,194 subjects, a case study using the KFDA algorithm shows that the proposed online classifier significantly improves the verification performance both in terms of accuracy and computational time. Further, it is also observed that the proposed online classifier is at least three times faster than the conventional SVM classifier.

The concept of online learning has not yet been fully explored in biometrics for different stages such as preprocessing and feature extraction. It is possible for any stage

[4] In our experiments, we found that $\gamma = 6$ for SVM and its variants, and $\gamma = 4$ for KFDA, resulted in the best verification performance.

[5] Time is computed on a 2.4 GHz Pentium Duo Core processor with 4 GB RAM under MATLAB environment.

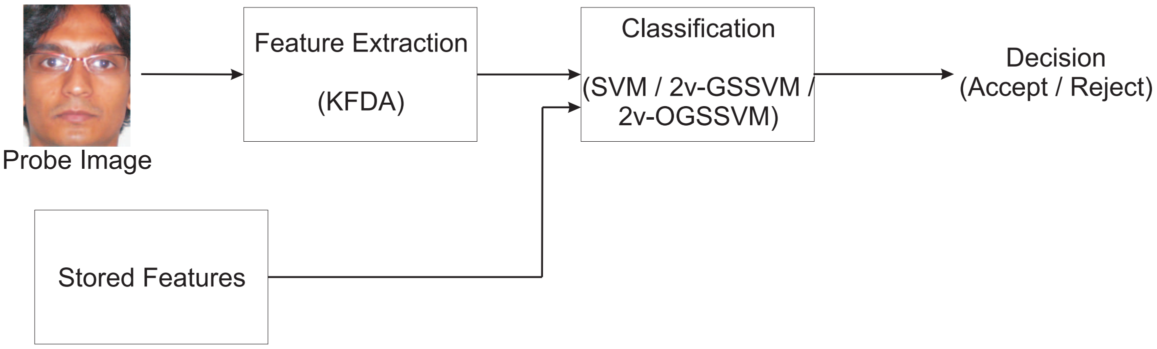

Fig. 2. Illustrating the steps involved in the face recognition case study.

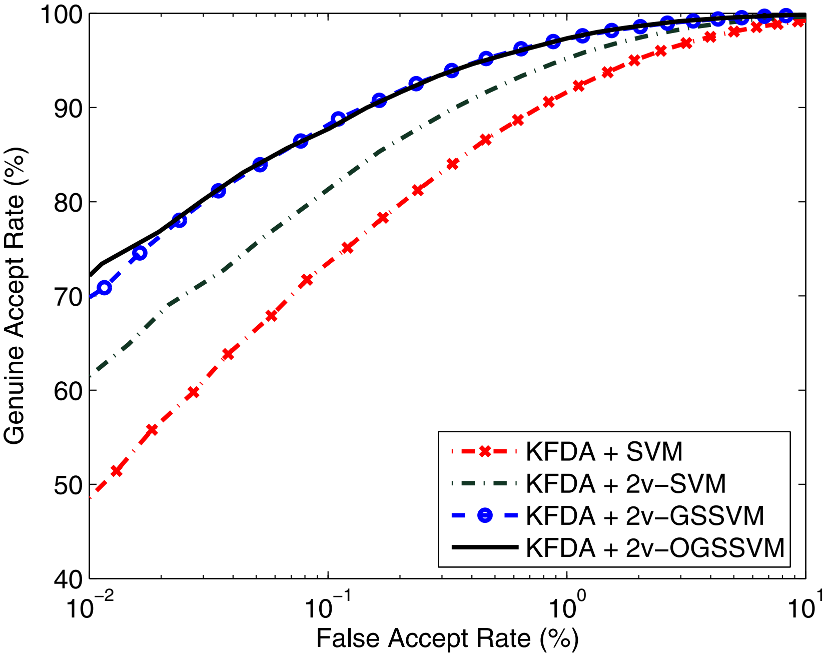

Fig. 3. Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using appearance based KFDA algorithm [9].

that requires offline training to be modified into the online learning mode. In future, we plan to develop online versions for texture- and feature-based face recognition algorithms.

V. ACKNOWLEDGMENTS

The authors would like to thank University of Notre Dame, NIST, Robotics Institute CMU, CMU AMP Research Lab, Dr. A.R. Martinez, Dr. E.G.L. Miller and Equinox Corporation 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 NSF CAREER grant number 0642554.

REFERENCES

[1] A. Bargiela and W. Pedrycz. *Granular computing: an introduction.* International Series in Engineering and Computer Science , 2002.

[2] G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. In *International Conference on Neural Information Processing Systems*, pages 409–415, 2000.

[3] H. G. Chew, C. C. Lim, and R. E. Bogner. An implementation of training dual-ν support vector machines. *Optimization and Control with Applications, Qi, Teo, and Yang (Editors)*, 2004.

[4] R. Fletcher. *Practical methods of optimization, 2nd Ed.* Wiley, 1987.

[5] P. J. Flynn, K. W. Bowyer, and P. J. Phillips. Assessment of time dependency in face recognition: an initial study. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, pages 44–51, 2003.

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

[7] J. Kivinen, A. Smola, and R. Williamson. Online learning with kernels. *IEEE Transactions on Signal Processing*, 52(8):2165–2176, 2004.

[8] A. R. Martinez and R. Benavente. The AR face database, 1998. Computer Vision Center, Technical Report.

[9] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Miller. Fisher discriminant analysis with kernels. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing IX*, pages 41–48, 1999.

[10] P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss. The FERET

evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[11] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[12] R. Singh, M. Vatsa, and A. Noore. Multiclass mν-granular soft support vector machine: a case study in dynamic classifier selection for multispectral face recognition. In *Proceedings of International Conference on Pattern Recognition*, pages 1–4, 2008.

[13] Q. Tao, G. Wu, F. Wang, and J. Wang. Posterior probability support vector machines for unbalanced data. *IEEE Transaction on Neural Network*, 16(6):1561–1573, 2005.

[14] D. Tax and P. Laskov. Online svm learning: from classification to data description and back. In *Proceedings of IEEE International Workshop on Neural Networks for Signal Processing*, pages 499–508, 2003.

[15] V. Vapnik. *The nature of statistical learning theory*. Springer, 1995.