A framework for liveness detection for direct attacks in the visible spectrum for multimodal ocular biometrics

4 authors, including:

Abhijit Das
Inria Sophia Antipolis - Méditerranée
39 PUBLICATIONS   **210** CITATIONS

SEE PROFILE

Umapada Pal
Indian Statistical Institute
396 PUBLICATIONS   **6,051** CITATIONS

SEE PROFILE

Miguel A. Ferrer
Universidad de Las Palmas de Gran Canaria
273 PUBLICATIONS   **2,451** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Face analysis View project

HOCR in Bangla View project

Chapter 1

An online learning-based adaptive biometric system

Abhijit Das,[1**] Rituraj Kunwar,[1**] Umapada Pal,[2] Miguel A. Ferrer Ballseter[3] and Michael Blumenstein [1]

[1]Institute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia, abhijit.das@griffithuni.edu.au, kunwar.rituraj@gmail.com, m.blumenstein@griffith.edu.au.
[2]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India, umapada@isical.ac.in
[3]IDeTIC, University of Las Palmas de Gran Canaria, Las Palmas, Spain, mferrer@dsc.ulpgc.es

Abstract. In the last decade adaptive biometrics has become an emerging field of research. Considering the fact that limited work has been undertaken on adaptive biometrics using machine learning techniques, in this chapter we list and discuss a few out of many potential learning techniques that can be applied to build an adaptive biometric system. In order to illustrate the efficacy of one of the incremental learning techniques from the literature, we built an adaptive biometric system. For experimentation we have used multimodal ocular (sclera and iris) data. The preliminary results have been reported in the results section, which are very promising.

Keywords. Adaptive biometrics, online incremental learning.

1.1 Introduction

Since the last few decades, intensive research work has been performed in the biometrics arena. Although various accurate systems have been proposed, surprisingly the adaptiveness of such systems to environmental changing conditions or change in biometric traits is low. Change in biometric traits or variation in the traits is a big challenge for the biometrics domain as it can lead to misidentification. The main reasons for misidenti-

fication or rejection of the correct individual by biometric systems are scarcity of training samples, presence of substantial intra-class variation during testing and lack of standardization of the data acquisition environment. Biometric characteristics can change either temporarily or permanently, perhaps due to ageing, diseases or treatment to diseases. In order to handle these issues; it would be preferable to have a biometric system which adapts well to the changing problem. These drawbacks of the present biometric systems have stimulated the interest of researchers in this domain. Recent developments in the adaptive system research area have opened up a new research field and that is "Adaptive Biometrics".

The ideal case of an adaptive biometric system is expected to handle the intra-class variations which changes with time (in many cases). These changes can happen for various reasons like aging, variations in pose and lack of standardization of data acquisition rules. The advantages of such a system are: learner need not get trained from scratch every time new data is available (as the learning happens continuously with time) and no need to store old data. This aspect of learning will significantly reduce the maintenance cost of biometric system. These are the characteristics which makes this research area so attractive and suitable for real-time scenario.

The existing automated adaptive biometric systems have adopted semi-supervised learning to create an adaptive system. A semi-supervised learning or online learning semi-supervised learning is a machine learning scheme which uses both labelled and unlabelled data. In such a machine learning systems the input samples are assigned labels using existing references and the positively classified samples are incorporated into existing references improve the adaptability of the system. The commonly adopted adaptation procedure is to augment the reference set with the newly classified input samples. The efficacy of such systems can be estimated by comparing the obtained performance gain with traditional biometric schemes which does not have any adaptation mechanism. The performance gain of such systems depends on the accurate labelling of the input samples. And this is because misclassifications will introduce impostor samples into the updated reference set, the result of which can be counterproductive and the inter class gap may get affected dramatically.

An adaptive biometric system can also operate in supervised or off-line mode in which biometric samples are manually labelled and updated. The supervised method represents the best-case performance as all the available positive (genuine) samples are used for adaptation. However, manual

intervention makes this process time consuming and expensive. Therefore it is generally infeasible to manually update references regularly. Despite of these above highlighted advantages, a large amount of limitations are also associated with the existing offline adaptive biometric systems. First of all capturing substantial amount of samples for such systems is quite time consuming. Commonly adopted self-updating system captures only limited amount of available samples during enrolment. As a result, a large number of input samples with informative and significant variations remain unenrolled and consequently results in limited performance gain over the baseline system. Secondly present adaptive biometric systems are vulnerable to impostor attacks or spoofing. As a consequence, such system bears the risk of getting adopted by imposter samples. Such consequence can affect the performance of the system dramatically. Even distinguishing between informative, redundant and noisy input samples is not possible in such system, not even in instances of supervised system. Hence as a result, these occurrences can not only affect the performance of the system but also affect the interclass variance space of the system.

In this chapter we discuss several existing online/adaptive learning techniques which can be applied in the biometric domain to build an adaptive biometric system which will overcome above discussed drawbacks. And this will bridge the existing gap in the literature. In order to justify the applicability of the discussed existing online theories theory multi-modal ocular (sclera and iris) biometric trait was used for experimentation with one of the existing techniques.

Ocular biometrics has gained popularity due to the significant progress made in iris recognition. Various reasons advocate this trend: these patterns possess a high degree of randomness, which is never same for any two individuals, not even for identical twins and this makes it ideal for personal identification. Further the patterns remain stable throughout a person's lifetime; these patterns even differ for the right and the left eye of the same individual. Therefore it comes to no surprise that iris has been so popular among the commercially available biometric systems. However, unfortunately iris recognition is unfavourably influenced by the frontal gaze direction of the eye with respect to the acquisition device. In such scenario, additional eye trait can be used to mitigate the motion bottleneck of iris biometrics. Among the various other ocular traits available (retinal scan, peri-ocular and sclera) the sclera (the white of the eye) is believed to be the most promising one and it may be of great significance in improving biometric applicability of iris recognition. Iris patterns are better discerned in the near infrared spectrum (NIR) while vasculature patterns are better

discerned in the visible spectrum (RGB). So, the multimodal ocular biometric using iris and sclera is proposed in visible spectrum. An additional advantage of this multimodal biometric is that, the sensor can sense the biometric trait from a distance hence such systems are hygienic as well. Although this proposed multimodal biometric has achieved a high recognition accuracy but their adaption with the environmental condition is an open research issue. Additionally adaption with respect to occlusion and various gaze and angle of eye is another challenge faced by this biometrics.

Hence, in this chapter we discuss many existing online / adaptive learning techniques which can be useful to underpin an adaptive biometric system. In particular it can come to a great rescue for the biometric problem like ocular biometric which are extremely sensitive to various changes as discussed above. Therefore we applied one of the existing techniques to illustrate its effectiveness in the ocular biometric domain. The organization of the chapter is as follows: section1.2 highlights the existing works in the literature of adaptive biometrics and followed by few advance work on ocular biometrics. In section1.3 we discuss various incremental / adaptive learning techniques available in literature including some experimental results and section 1.4 draws the overall conclusion and future scope of the theory proposed.

1.2 Adaptive Biometrics Literature

In the last decade, adaptive biometric systems have been adequately investigated. Various research approaches have been explored by different researchers. To the best of our knowledge, the first adaptive biometric approach was introduced in [25]. The existing literature of adaptive biometric systems can be categorised by the key attributes of machine intelligence methods proposed or used. Supervised [33, 31, 32, 34, and 35] against semi-supervised training [32, 36, and 9], Self-train [36–38] or co-train [32, 39], and Online [36, 37] and offline [32, 40, 41].

A general study has been done in regard to adaptive biometric in [30] to address the intuitive questions like:

1.   Whether supervised adaptation better than semi-supervised?

2.   Whether co-training can outperform self-training?

3.   Whether offline adaptation is better than online?

Interesting analysis was performed further to validate the hypothesis.[28].

In [31] a two-stage classifier selection technique was proposed for automatically updating the biometric templates. In this approach a labelling scheme based on probabilistic semi-supervised learning was employed. Soft probabilistic labels were assigned to each batch of input samples by calculating the minimum energy function on the graphical representation. The harmonic function used was unique and ensured that labels were assigned to input sample using both the enrolled and nearby input data. Then the genuinely classified samples undergo the selection process based on risk minimization. The experiment was validated on a DIEE finger print dataset, an appreciable result was achieved.

In [32], effect of different threshold settings were employed for template update and novel solutions was proposed for by passing the threshold selection step. This work analyzed and inferred that template update method is better for group specific updating due to presence of different type of population. Efforts have been made in the work to give a preliminary guideline on the type of update procedures that could be followed for a specific group of population. A protocol for simulating real world situation has also been proposed for the unbiased evaluations of update methods.

Most of the adaptive biometric work in literature deals with finger prints, voice or facial traits. Ocular biometric is one the most accurate and reliable biometric; it offers various biometric traits. Adaptiveness of ocular biometrics is yet to be addressed in the literature. Among ocular biometrics, iris [26, 27] is the most promising one. Apart from iris, the human eye has an ocular white surface known as the sclera, which contains a texture pattern due to the presence of blood vessels on its surface. The sclera patterns can be acquired easily along with iris in single camera shot and it is visible, even in off-angle eye shot. Therefore by utilizing the texture pattern of sclera in addition to iris, the performance of an iris recognition system can be significantly improved even with a non-ideal or an off-angle eye data sample.

The first recognized work on sclera biometrics is recorded in [1]. In this paper, the authors discuss methods for conjunctival image pre-processing by computing a Gaussian filters and Hessian matrix. In order to derive a suitable vascular template for biometric authentication, feature extraction

and classification is performed by a minutia based template matching. Here a dataset was built to establish the experiment.

In [6], the first automatic sclera detection technique was proposed by a time adaptive active contour-based method. First automatic segmentation processes for sclera biometric was proposed in [4], Otsu's algorithm was used to segment the sclera from a gray image. Sclera segmentation based on color image was first proposed in [5], HSV model was used to segment sclera from a color eye image and a bank of Gabor filter was used to enhance the sclera. Many features like LBP [9], [24], GMCL [8] are used in the literature for sclera feature representation. First multi-angled sclera recognition was proposed in [2], further in [7] multi-angled sclera recognition was proposed using multispectral imagery. Few advance work on sclera is found using features like in SIFT [14], OLBP [15], Dense -SIFT [16]. In [16] K-means followed by Spatial Pyramid matching [12] was used to enhance the feature set.

First multi-modal eye recognition techniques using sclera and iris was proposed in [3]. Here a score fusion based technique was adopted to combine the sclera and the iris feature. Further in [5] a quality fusion technique was used to combine sclera and iris feature and in [10] feature level combination was used to establish sclera and iris based multi-modal biometrics. A survey on sclera recognition is recorded in [17].

To this date, adaptive biometric is relatively less studied and little is known regarding its usefulness. So, the state of the art related to it is not matured yet. Therefore, in the next section we discuss various generic adaptive learning methods that can be used in adaptive biometrics domain. We also implement one of the adaptive/online learning methods to demonstrate the efficacy of the used online learning method in multi-modal ocular (sclera and iris) biometric domain.

1.3 Generic Adaptive Learning Methods

In this chapter we address the potential incremental learning techniques which can be applied in the biometric domain to create an adaptive learning system. The motivation to create an adaptive learning system for biometrics is explained below.

In real world scenarios, where we use machine learning algorithms, we often have to deal with cases where the input data changes its nature with time. In order to maintain the accuracy of the learning algorithm, we fre-

quently have to retrain our learning system, thereby making the system inconvenient and unreliable. This problem can be solved by using learning algorithms which can learn continuously with time (incremental/online learning). In contrast, offline learning works fine in an ideal scenario where there is no change in the underlying distribution of the input with time. However, for various reasons this does not often hold in real time problems that we intend to address (i.e. of robust biometric system) using machine learning.

In contrast to offline learning, ideally, incremental/online learning can be simultaneously trained and tested. Precisely, it need not stop performing its task (i.e. prediction or classification) if the learner has to update its learning parameters. Learning parameters can be updated as soon as the new training data is available. This leads to the creation of a never ending learning process which can adjust itself even if the environment changes and can perform learning while performing task.

The critical assumption on which most of the incremental learning algorithms are based upon is that previous data is completely or partially accessible. Based on this assumption, to handle streaming data they apply the time windowing technique of either fixed or variable size [48, 42 and 49]. Others have handled streaming data by weighting models in the ensemble [47, 50 and 51] or by weighting the data [52] or by retaining only the relevant subset of previous data [44 and 45]. We assume for our experimentation that we have no access to the previous data, thus making the algorithm capable to handle the scenario where old data is inaccessible.

In the next section we shall discuss some potential methods which can be used to make an adaptive biometric system along with some preliminary results that was produced using one of the biometric problems.

1.3.1 Ensemble of Classifiers

Since the inception of ensemble-based classification, it has been one of the most studied classification methods [54, 55]. Ensemble based classifiers have often been used in past for performing incremental/online learning [47, 50, 51, 53, 56]. The principle behind the ensemble decision is that the individual predictions combined appropriately, should have better overall accuracy, on average, than any individual ensemble member [56]. There are various reasons why an ensemble-based classification is chosen over a single classifier, a few are listed below.

1. No free lunch theorem states that in the absence of prior knowledge about a problem, no one classifier is universally better than any other classifier [57], this also includes random guessing.

2. In case of extremely high dimensional data, a single classifier's complexity may scale with the dimensionality of the data thus making the generation of a reliable single classifier infeasible. Instead of a single classifier, generate multiple classifiers on different subsets of features thus reducing the complexity of each classifier trained on the subset.

3. Single classifiers may not work well with data that are too little or too large in size. To work around this problem, ensembles can generate classifiers on multiple bootstrap datasets.

4. Reduces bias towards majority class (class that is well represented by training samples). And generating single strong classifier may be infeasible due to computational costs.

The success of ensemble learning algorithms is believed to depend both on the accuracy and on the diversity among the base learners [58] and some empirical studies revealed that there is a positive correlation between accuracy of the ensemble and diversity among its members [59, 60]. Breiman [69] also shows that random forests with lower generalization error have lower correlation among base learners and higher base learners' strength. Besides, he derives an upper bound for the generalization error of random forests which depends on both correlation and strength of the base learners.

Literature suggests that there is a trade-off between base learner's accuracy and diversity, meaning that lower accuracy may indicate higher diversity. However, study in [62] shows that relationship between accuracy and diversity is not straightforward and lower accuracy may not essentially mean higher diversity. A recent study in [77, 78] discuss that when, how and why ensembles of learning machines can help to handle concept drift in online learning, through a diversity study in the presence of concept drift. This paper presents an analysis of low and high diversity ensembles combined with different strategies to deal with concept drift and proposes

a new approach "Diversity for Dealing with Drifts" (DDD) to handle drifts. DDD maintains ensembles with different diversity levels, exploiting the advantages of diversity to handle drifts and using information from the old concept to aid the learning of the new concept. The authors claim that DDD is accurate both in the presence and in the absence of drifts.

In a recent study in [64] has reiterated the efficacy of ensemble-based learning to create an adaptive/online learning system for handwritten character recognition. We have used that method to make an adaptive biometric system which learns using one sample at a time. The system as presented in [64] is briefly described below.

The block diagram shown in Fig.1 shows the overall picture of the online learning method proposed in [64]. The method proposed in the paper is to conduct both online supervised as well as online semi-supervised learning. In general, to conduct semi-supervised learning abundant unlabeled data is required but unfortunately we have very limited number of samples/class in our biometric learning problem. Therefore we only conduct online supervised learning but with availability of more data in future we intend to apply semi-supervised online learning as well. Technical details are given below:

Let us introduce some notation to describe the data. Training dataset $\chi = \{(X^1, Y^1) \dots (Y^N, Y^N)\}$ where $X^i = x_1^i \dots x_D^i$, $\quad X^i \epsilon R^D$, are the samples in a D dimensional feature space and $Y \epsilon \{1, \dots K\}$ are the corresponding labels for a K-class classification problem.

Using Bayes rule and conditional independence among the feature given the class label (assumption used to formulate Naïve Bayes classifier), we can write the posterior probability as:

$$P(Y = y_k | x_1 \dots x_D) = \frac{P(Y = y_k) \prod_i P(x_i | Y = y_k)}{\sum_j P(Y = y_j) P(x_1 \dots x_D | Y = y_j)}$$

So to train our classifier we fit a Gaussian $\mathcal{N}\left(x_i; \widehat{\mu_{ik}}, \widehat{\sigma_{ik}^2}\right)$ to each $P(x_i | Y = y_k)$, and we estimate mean and variance for the same using the training data. We perform Maximum Likelihood Estimation (MLE) to find the mean $\widehat{\mu_{ik}}$ and variance $\widehat{\sigma_{ik}^2}$ of $P(x_i | Y = y_k)$ for each feature x_i, which is just equal to sample mean and sample variance respectively. And the classification rule for a new sample $X^{new} = < x_1 \dots x_D >$ can be written as

$$Y^{new} = \arg max_{y_k} P(Y = y_k) \prod_i P(x_i|Y = y_k)$$

$$Y^{new} = \arg max_{y_k} \pi_k \prod_i \mathcal{N}\left(x_i^{new}; \widehat{\mu_{ik}}, \widehat{\sigma_{ik}^2}\right)$$

In order to make an ensemble of B classifiers, we repeat the following steps B times: Randomly select F features from the pool of D features. Estimate the learning parameters for the classifier.
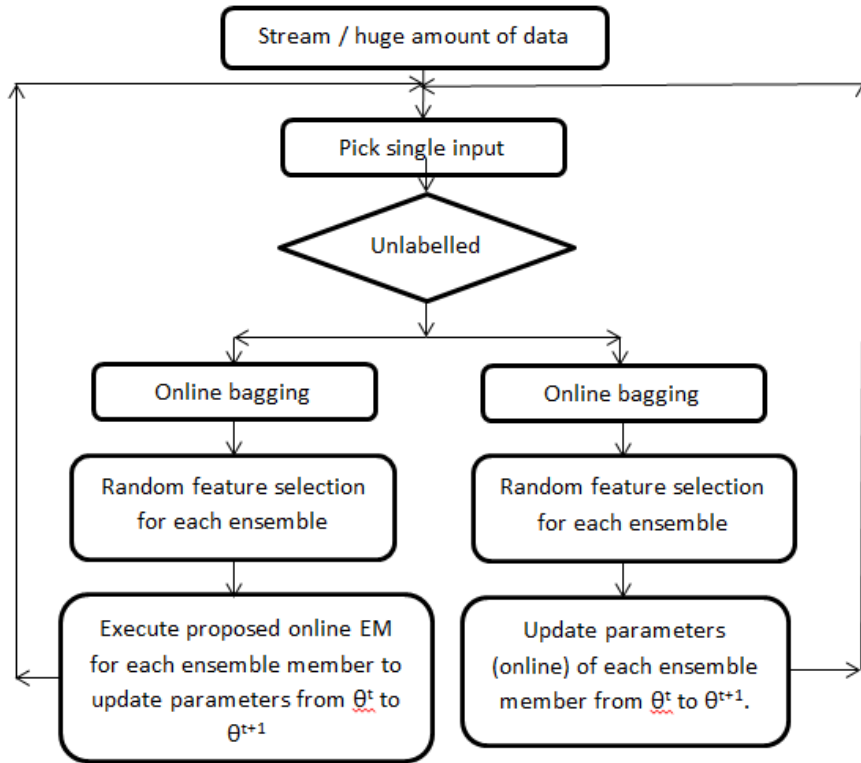

Fig. 1.1: Block diagram representing the batch learning (MLE) of Augmented Naive Bayes network.

A test sample will be classified by each classifier in the ensemble and the class which gets the majority vote by the ensemble will get assigned to the test sample. In the above explained way we train an ensemble of B

classifiers as an initialization step, by just 1 samples/class. Beyond this with time, as we get more samples of a class, we can update its respective learning parameter as shown below in an online (on the fly) manner using 1 sample at a time. This enables the system to adapt to the changes in underlying distribution of input samples. Initialization (by just one labeled training sample):

$$\widehat{\mu_{ik}} = x_{ik}^1 \; ; \; \widehat{\sigma_{ik}^2} = \sigma_0 \; ; \; \pi_k = \frac{1}{no.\,of\,classes} \; ; \; c_k = 1 \; ; \; \alpha = \alpha_0$$

Where $c_k = no.\,of\,samples\,used\,so\,far\,for\,training,$

  $\alpha = decides\,the\,length\,of\,memory\,of\,the\,classifier\,(\alpha < 1)$

Repeat steps below for all the incoming labeled training samples for any class k:

$$c_k = c_k + 1 \; ; \; \eta_k = \left(\frac{1-\alpha}{c_k} + \alpha\right)$$

  $where\,\eta_k\,is\,learning\,rate\,for\,class\,k$

$$\mu_{ik}(t) = (1 - \eta_k)\mu_{ik}(t-1) + \eta_k x_i^j \delta\left(Y^j = y_k\right)$$

$$\sigma_{ik}^2(t) = (1 - \eta_k)\sigma_{ik}^2(t-1) + \eta_k \left(x_i^j - \mu_{ik}(t)\right)^2 \delta\left(Y^j = y_k\right)$$

In the same paper [64], another method has been suggested to conduct online learning in semi-supervised manner. The only difference between supervised online learning and semi-supervised online learning lies in the definition of learning rate. Learning rate definition for semi-supervised learning is:

$$\eta_k = q_k \left(\frac{1-\alpha}{c_k} + \alpha\right)\lambda \; ; \; (\alpha < 1)$$

Where λ = weight factor applied to moderate the contribution of unlabelled data in the parameter estimation step.

If an incoming new sample is unlabeled then the trained classifier is used to produce the posterior $q_k = P(Y^j = y_k|X)$ corresponding to all k (class). This posterior will be used to calculate the learning rate corresponding to all the classes and subsequently this learning rate will be used

to update learning parameters of all the classes as is done in case of supervised online learning.

In [64] the authors have upgraded the Naive Bayes network structure and have suggested method to conduct an online learning for that upgraded network. The network structure was upgraded to get rid of the Naïve Bayes unrealistic assumption of conditional independence between different features given the class label. The structure was upgraded with a restriction that all the features will have at most two parents (earlier each had just one) Fig. 1. The improved structure was proved to be working much better in the concerned application. The technical detail is briefly explained below.
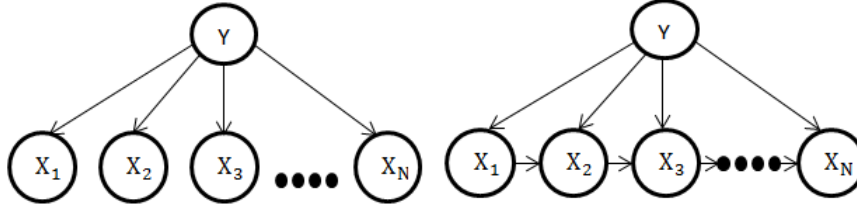

Fig. 1.2: Left: Bayes net for Naive Bayes. Right: Bayes net for Augmented Naive Bayes.

$P(X_i|X_{i-1}, Y)$ can be parameterized by the following Gaussian distribution: $P(X_i|X_{i-1}, Y) = N(\beta_0 + \beta_1 X_{i-1}, \sigma^2)$

The author in [64] performs MLE estimation to evaluate the learning parameters β_0, β_1 and σ^2 for supervised learning

$$\beta_1 = \frac{E_D[X_i X_{i-1}] - E_D[X_i]E_D[X_{i-1}]}{E_D[X_{i-1}X_{i-1}] - (E_D[X_{i-1}])^2} = \Sigma_{X_{i-1}X_{i-1}}^{-1}\Sigma_{X_iX_{i-1}}$$

$$\beta_0 = \mu_{X_i} - \Sigma_{X_{i-1}X_{i-1}}^{-1}\Sigma_{X_iX_{i-1}}\mu_{X_{i-1}}$$

$$\sigma^2 = \Sigma_{X_iX_i} - \Sigma_{X_iX_{i-1}}\Sigma_{X_{i-1}X_{i-1}}^{-1}\Sigma_{X_iX_{i-1}}$$

The   authors   further   propose   to   update $E_D[X_i]$, $E_D[X_{i-1}]$, $E_D[X_{i-1}X_{i-1}]$,

$E_D[X_i X_{i-1}]$ in an online manner in order to update $\beta_0, \beta_1 \ and \ \sigma^2$ because they are the building block which is obvious from their definition. Hence, it can be written as:

$$E_D[X_i](t) = (1 - \eta_k)E_D[X_i](t - 1) + \eta_k X_i$$

$$E_D[X_{i-1}](t) = (1 - \eta_k)E_D[X_{i-1}](t - 1) + \eta_k X_{i-1}$$

$$\sigma^2_{X_{i-1}}(t) = (1 - \eta_k)\sigma^2_{X_{i-1}}(t - 1) + \eta_k\big(X_{i-1} - E_D[X_{i-1}](t)\big)^2$$

Covariance between two RVs, A and B is given by:

$$\sigma_{A,B} = E[AB] - E[A]\,E[B]$$

Therefore: $E_D[X_{i-1}X_{i-1}](t) = \sigma^2_{X_{i-1}}(t) + \{E_D[X_{i-1}](t)\}^2$

Similarly: $\sigma_{X_i X_{i-1}}(t) = (1 - \eta_k)\sigma_{X_i X_{i-1}}(t - 1) + \eta_k\big(X_i - E_D[X_i](t)\big)\big(X_{i-1} - E_D[X_{i-1}](t)\big)$

Using the above equations:

$$E_D[X_i X_{i-1}](t) = \sigma_{X_i X_{i-1}}(t) + E_D[X_i](t)E_D[X_{i-1}](t)$$

In the equations above, η_k refers to learning rate and its definition is same as in case of Naïve Bayes online learning case. And along similar lines, online semi-supervised learning was proposed with a changed definition of learning rate. It has been shown that the upgraded network performs much more accurately as it captures the relationship between different features and accordingly learns. Details can be found in [64].

1.3.2 Incremental/Adaptive Support Vector Machines (SVMs)

SVM is based on a kernel method; however unlike suboptimal kernel methods, as in the case of a kernel method based on clustering, kernel methods for SVMs are optimal, with the optimality being rooted in convex optimization. Realizing the theoretical strength of SVMs, researchers have developed incremental versions of them. And considering the fact that incremental SVMs have never been explored in the biometric domain, it becomes imperative to discuss SVMs and their incremental versions in the context of adaptive biometrics.

Classification and regression methods based on SVMs [23, 66] are very powerful, which generalize well even in case of very sparse and high dimensional data. SVM is based on Vapnik's structural risk minimization induction principle which carries out searching over hypothesis classes of varying capacity with best generalization performance.

A two class classifier based on SVM can be represented as: $f(X) = w.\varphi(X) + b$ are learned from the data $\chi = \{(X^1, Y^1) \dots (Y^N, Y^N)\}$ where $X^i = x_1^i \dots x_D^i, X^i \epsilon R^D$ and $Y^i \epsilon \{-1,1\}$ by minimizing

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \xi_i^p$$

For $p \in \{1,2\}$ subject to the constraints (soft margin)

$$Y^i\big(w.\varphi(X^i) + b\big) \geq 1 - \xi_i^p, \; \xi_i^p > 0 \; \forall \; i \; \in \{1, \dots \dots, N\}$$

A set of slack variables are introduced for the system to allow few samples to be on the wrong side of the margin (to handle overlapping class distributions) but impose penalty of $\xi_i^p = |Y^i - f(X^i)|^p$ over the objective cost for the samples to be on the wrong side of the margin boundary. Value of $\xi_i^p = 0$ for being on the correct side of the margin boundary. p = 1 is what generally preferred in practice because of the robustness to outliers that hinge loss offers as compared to the quadratic loss which corresponds to p = 2. The goal is to minimize above the objective function while softly penalizing the points that lie on the wrong side of the margin boundary. Parameter C > 0 controls the tradeoff between the slack variable penalty and the margin. Above minimization can be done using quadratic programming but to simplify and take advantage of the kernel trick the above minimization problem is expressed in its dual form

$$\min_{0\leq\alpha^i\leq C} W = \frac{1}{2}\sum_{i,j=1}^{N} \alpha^i Q^{ij}\alpha^j - \sum_{i=1}^{N}\alpha^i + b \sum_{i=1}^{N} Y^i\alpha^i$$

With   the   Lagrange   multiplier   (and   offset)   b   and $Q^{ij} = Y^iY^j\varphi(X^i).\varphi(X^j)$. The above dual form of the original minimization problem must satisfy the famous Karush-Kuhn-Tucker (KKT) condi-

tion. The KKT conditions generally involves the Primal constrains, dual constrains and complementary slackness. Therefore the above dual form along with the KKT condition gives rise to a linearly constrained quadratic programming problem. And there are standard solvers available to solve them. The resulting dual form of the SVM is then $f(X) = \sum_{i=1}^{N} y^i \alpha^i \; \varphi(X^i) \varphi(X) + b$. Point to be noted here is that the transformed sample now only appear in dot product. Therefore one can employ a positive definite kernel function to implicitly project the input samples into some high dimensional (which can be infinite) space and calculate the dot product to perform classification or regression in that space.

In the mid-90's Support Vector Machines (SVMs) emerged and subsequently researcher's interest in its online version arose. Early work on this subject by [67] suggests that for each new batch of data, a Support Vector Machine is trained on the new data and the Support Vectors from the previous learning step. And the logic behind this approach is that the decision function of an SVM depends only on its Support Vectors, i.e. training an SVM on the Support Vectors alone results in the same decision function as training on the whole data set. Because of this, one can expect to get an incremental result that is equal to the non-incremental result, if the last training set contains all examples that are Support Vectors in the non-incremental case. However the shortcoming of this approach is that as there are typically only very few Support Vectors, their influence on the decision function in the next incremental learning step may be very small if the new data is distributed differently. Note: Support Vectors are a sufficient description of the decision boundary between the examples, but not of the examples themselves.

The above problem was addressed in [68] by making a clever change in the objective function to be optimized, and i.e. by making the error on old Support Vectors (which represent the old learning set) more costly than an error on a new example. Details can be found on the concerned paper. At the same time [69] exploits the locality of the RBF kernel to build online SVM. The authors do not use all the previous support vectors (as done in [67, 68]), instead it only uses the support candidates in the neighborhood of the new incoming sample. Though deciding the neighborhood is critical, the method would be fast. However above three approaches and methods proposed by [70]; provide only approximate solution and may require many passes through the dataset to reach a reasonable level of convergence.

An exact solution to the problem of online SVM learning has been found by [71]. Their incremental algorithm updates an optimal solution of an SVM training problem after one training example is added (or removed). In this the authors construct the solution recursively one point at a time such that the KKT condition is satisfied over all the data already seen as well as the new incoming samples.

The first order condition on W reduces to KKT condition:

$$g^i = \frac{\partial W}{\partial \alpha^i} = \sum_{j=1}^{N} Q^{ij}\alpha^j + Y^i b - 1 = Y^i f(X^i) - 1$$

$$g^i = \begin{cases} \geq 0 & \alpha^i = 0 \\ = 0 & 0 < \alpha^i < C \\ \leq 0 & \alpha^i = C \end{cases}$$

$$\frac{\partial W}{\partial b} = \sum_{j=1}^{N} y^j \alpha^j = 0$$

This partitions the data into three categories:

A) $x^i \in S \subset \chi$ where S is a set of margin support vectors, strictly on the margin (i.e. $Y^i f(X^i) = 1$).

B) $\in O \subset \chi$, where O is the set of other vectors for which $Y^i f(X^i) > 1$ i.e. the sample is on the correct side of the margin boundary (correctly classified).

C) $x^i \in E \subset \chi$ where E is a set of error vectors $Y^i f(X^i) < 1$, sample is on the wrong side of the margin boundary but not necessarily misclassified.

The set R $= \{O \cup E\}$ is a set of reserve vectors. Lower case letters s, e, o and r will be used to refer to such kind of partitions.

By writing the KKT conditions before and after an update $\Delta\alpha$ we obtain the following conditions that must be satisfied after an update [71]:

$$\begin{bmatrix} \Delta g^c \\ \Delta g^s \\ \Delta g^r \\ 0 \end{bmatrix} = \begin{bmatrix} Y^c & Q^{cs} \\ Y^s & Q^{ss} \\ Y^r & Q^{rs} \\ 0 & Y^{sT} \end{bmatrix} \begin{bmatrix} \Delta b \\ \Delta\alpha^s \end{bmatrix} + \Delta\alpha^c \begin{bmatrix} Q^{ccT} \\ Q^{csT} \\ Q^{crT} \\ Y^c \end{bmatrix}$$

It is easy to see that $\Delta\alpha^c$ is in equilibrium with $\Delta\alpha^s$ and Δb in order for the above condition to hold. Considering the fact that $\Delta g^s = 0$, from line 2 and 4 of the above equation we can write:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & Y^{sT} \\ Y^s & Q^{ss} \end{bmatrix} \Delta s + \begin{bmatrix} Y^c \\ Q^{csT} \end{bmatrix} ; where \ \Delta s = \begin{bmatrix} \Delta b \\ \Delta\alpha^s \end{bmatrix}$$

Above linear equation can be solved for Δs

$$\Delta s = \beta\Delta\alpha^c$$

Where

$$\beta = - \begin{bmatrix} 0 & Y^{sT} \\ Y^s & Q^{ss} \end{bmatrix}^{-1} \begin{bmatrix} Y^c \\ Q^{csT} \end{bmatrix}$$

is the gradient of the manifold of optimal solutions parameterized by α^c Similarly from line 1 and 3

$$\begin{bmatrix} \Delta g^c \\ \Delta g^r \end{bmatrix} = \gamma\Delta\alpha^c$$

Where

$$\gamma = \begin{bmatrix} Y^c & Q^{cs} \\ Y^r & Q^{rs} \end{bmatrix} \beta + \begin{bmatrix} Q^{ccT} \\ Q^{crT} \end{bmatrix}$$

is the gradient of the manifold of gradient g^r at an optimal solution parameterized by α^c. These refined set of equations shows that that update process is controlled by very simple sensitivity relation: $\Delta s = \beta\Delta\alpha^c$ and $[\Delta g^c \quad \Delta g^r]^T = \gamma\Delta\alpha^c$, where β is the sensitivity of Δs with respect to $\Delta\alpha^c$ and γ is the sensitivity of $\Delta g^c \ and \ \Delta g^r$ with respect to $\Delta\alpha^c$.

At this stage to carry out the parameter update process, the key is to find out the largest possible increment of $\Delta\alpha^c$ and subsequently Δs and Δg is updated. Authors of [71] have very exhaustively addressed all the cases by

which one can determine the largest value of $\Delta\alpha^c$. And once the step is determined, one can follow the steps of algorithm 1 given in [72] to carry out online learning.

Finding an absence of a well-accepted implementation of the work by [71]; [72] proposed a new design of storage and numerical operation which speeds up the incremental SVM training by a factor of 5 to 20. On the similar line, [73] have applied the accumulated knowledge of optimization to the computational problem presented by the SVM to propose a very efficient way of training SVM in incremental fashion.

1.3.3 Incremental/Adaptive Neural Network

Neural Networks being one of the oldest methods of machine learning, it is very obvious that umpteen amount of work have been done on that domain and many of them deals with incremental and adaptive learning. Here we list few briefly which can be useful to create an adaptive biometric system.

Fuzzy ARTMAP: This is a neural network based structure and it is one of the earliest methods used in incremental learning. The fuzzy ARTMAP has two fuzzy ART modules that are linked via an inter-art module known as "map field". The map field is used to form predictive categories for learning class association. Fuzzy ARTMAP will generate new decision clusters in response to new input patterns that are sufficiently different from previously seen instances. The 'sufficiently different' patterns are controlled using a free parameter of ARTMAP known as the vigilance parameter. ARTMAP is sensitive to the vigilance parameter especially in presence of significant noise in the training data. Using stability and match tracking, fuzzy ARTMAP automatically constructs as many categories as are needed to learn any static training set to 100%. Thus, fuzzy ARTMAP may over-fit, leading to poor generalization.

Learn ++: This is one of the most notable families of incremental learning algorithm, which was first introduced by Polikar et al. in [46] and later upgraded by few other authors for e.g. [47]. It creates multiple classifiers to each data chunk presented to the system. Inspired by AdaBoost [77, 78] for each chunk, the training set for each base learner is created by sampling examples according to a distribution of probability. Like AdaBoost, Learn++ maintains a distribution of instance weights; however Learn++ does not update the weights in the same manner as performed with Ada-

Boost. In AdaBoost, the distribution of probability is built to give higher priority to instances mis-classified by the last previously created classifier whereas Learn++ uses the ensemble decision, rather than the decision of the latest classifier. When a new dataset arrives, the distribution is re-initialized by evaluating the entire ensemble and initializing the distribution. Pros: Learn++ does not have to access the previous data chunks and it demonstrate considerable improvement at generalization when compared with fuzzy ARTMAP on common databases. *Cons:* Problem is that a new set of classifiers is created for each new data chunk. So, the ensemble size can become extremely large considering lifelong learning.

Self-Organized Incremental Neural Network (SOINN): This is an unsupervised incremental learning method which was proposed in [73] for topology learning and classification to handle noisy unlabeled data. This method is essentially a combination of self-organizing map [75] and competitive Hebbian learning [76] which can be used to learn the topology of the input data stream. The proposed algorithm makes a two layered neural network Fig 3. The first layer which represents a reasonable topological structure of unlabeled data gives a reasonable number of clusters and gives a typical prototype pattern of every cluster. Prior knowledge regarding the number of classes or codebook is not required.

The first layer learns the density distribution of the input pattern. Subsequently output of the first layer serves as the input of the second layer, where the different clusters are separated by detecting the low density overlap area. The method uses similarity threshold and a locally accumulated error-based insertion criterion to grow the system incrementally and accommodate the input patterns of online non stationary data distribution. It also uses and online criterion to delete nodes from the low probability regions and this enables the system to separate the cluster and simultaneously eliminate the noise noisy samples from the input data. Authors use "error radius" as the utility parameters to control the growth of number of nodes in the network and check successful node insertion. Though the method has been successfully applied on some real-time problems, it has several limitations for e.g. a) in case of high density overlap it is difficult for the method to separate the clusters b) several important parameters value need to be decided by the users which increase the chances that the system getting used sub optimally. On the similar lines authors have further modified the SOINN to make a semi-supervised incremental active learning system [74] which is very promising and claims have been made that this method can be very useful to create a never ending learning system.

Fig. 1.3: Flowchart of SOINN, shows the basic overflow of the proposed algorithm. This figure has been taken from [75].

1.4 Experimental details and results

In order to evaluate the effectiveness of the discussed incremental learning algorithms, we have used one of the discussed methods to develop an adaptive sclera biometric system. Fuzzy C-means clustering was used to segment the sclera [18-20].

Since the vessel patterns in the sclera are not prominent, the image was enhanced to make it visible. A new preprocessing approach for vein highlighting is proposed using adaptive histogram equalization [21]. And his-

togram equalization was performed with a window size of 42 x 42 (the window value was selected by analysis, window value that produces the best result was used for experimentation). Since the sclera vessel patterns are most prominent in the green channel, it was used for preprocessing to make the vessel structure more prominent

Furthermore, a bank of filter based on Discrete Meyer wavelet [22] was used to enhance the vessel patterns. Low pass reconstruction of the above mentioned filter was used to enhance the image.

We use sclera feature extraction based on the Dense Local Directional Pattern (D-LDP) [80] which is an extended version of LDP proposed in [11]. For extracting the patch/dense descriptors, each image was divided into a regular dense grid of three different special pyramidal planes. A higher order LDP with 4 orientations was used here for the feature extraction.

A histogram of bin size 512, for of each of these patches are calculated in multi-scales of 1, 2 and 3. Finally the histogram of each patch and each scale are calculated and concatenated column wise to get the final descriptor.

The irises were segmented along the radius by calculating the center and the radius by the integro-differential [81] operator and further enhanced using an adaptive histogram equalization technique. The red channel of the color image was used for iris image enhancement. Image level fusion was performed using iris and sclera, and subsequently the patterns were classified using the developed features.

In order to perform the experiment, UBIRIS version 1 [13] dataset was used in the experiment. This database consists of 1877 RGB images taken in two different sessions (1205 images in session 1 and 672 images in session 2), from 241 identities and images are represented in RGB color space. The database contains blurred images and images with blinking eyes. Both high resolution images (800×600) and low resolution images (200×150) are present in the database. For each individual 10 eye images were present. All the images are in JPEG format.

The dataset consists of different quality of images with respect to sclera region visibility. Some of the images are not occluded and have sclera regions visible which are of good quality; some of them are of medium qual-

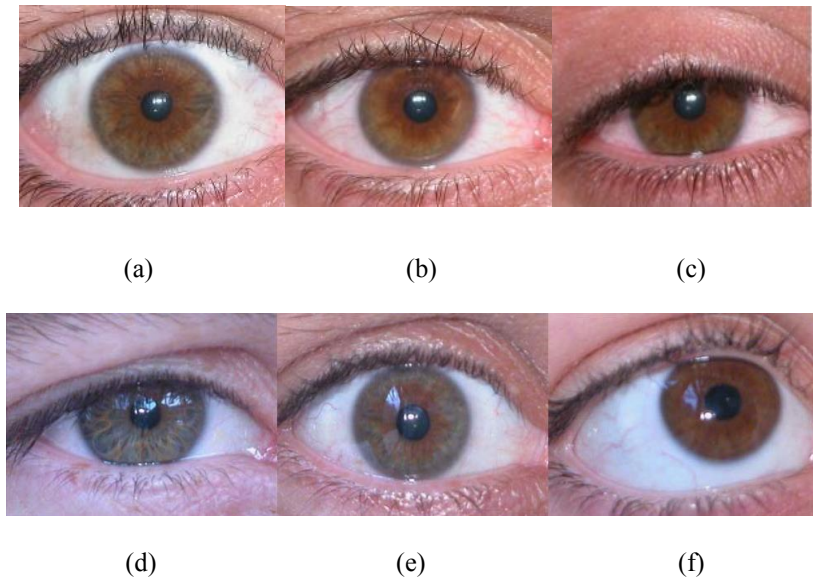ity and the third type was of poor quality with respect to sclera region visibility as shown in figure 1.5.

(a)   (b)   (c)

(d)   (e)   (f)

Fig 1.4: Different quality of eye images used in experiment (a) is the type of best quality images of Session 1, (b) is the type of medium quality of Session 1 (c) is the type of Poor quality of Session 1, (d) is the type of below average quality image of Session 2, (e) is the type of average quality of Session 2 (f) is the type of best quality image in Session 2

First session images have minimum noise in the form of reflections, luminosity and contrast, as it uses a standard image capturing framework inside a closed room with controled artificial lights. In the second session pictures were captured in more natural conditions, for example it uses both natural and artificial light, there by bringing in the changing image capturing settings into play. This difference in the capturing enviroment produces heterogeneous images with respect to reflections, contrast, luminosity and focus. Images collected at second session were captured by a vision system with or without minimal active participation from the subjects, adding several noise problems. There was a gap of two weeks between the two session.

These changes in environmental condition in the form of different gaze angle of the eye, data accruing techniques and time span gap were utilized to investigate the adaptability of the sclera texture for biometric identification.

It was observed that the number of participants in the two sessions were not the same which produced a very uneven number of samples corresponding to different individuals. Even the number of the population was differnt for the two sessions. The first session consisted of 241 users and in the second sesion there were 135 users.

We have used 50 out of 241 total classes present because many classes did not have data from both sessions which makes the number of samples too few to apply the learning algorithm over them. We did not use a few classes because the iris and sclera region of the participants were too occluded to be used for learning. Few examples of such images are given below in figure 1.6.
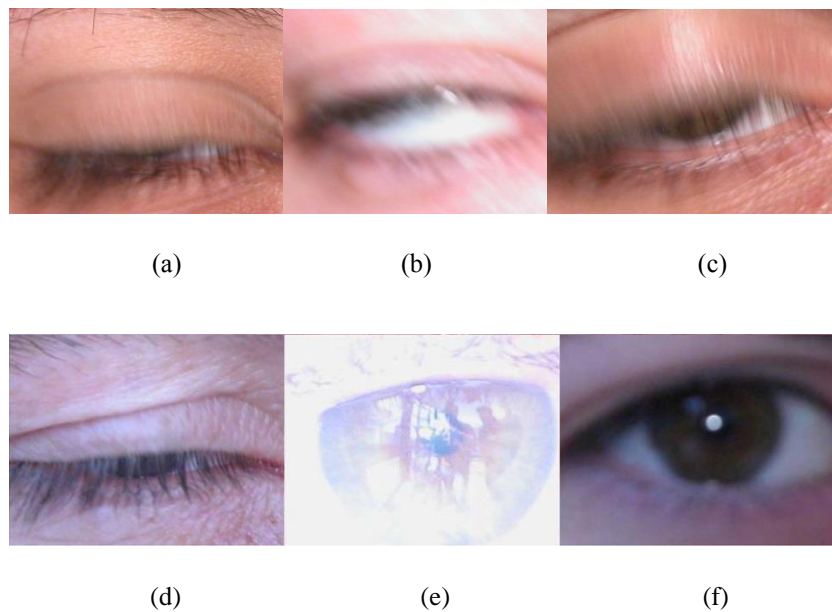

(a)   (b)   (c)

(d)   (e)   (f)

Fig 1.5: Example of closed and blurred eyes. (a), (b) and (c) are of session 1 and (d), (e) and (f) are of session 2.

For our experiment as a learner we used incremental Naive Bayes classifier as proposed in [64]. The number of features used for each class was 30480. The total number of classes used was 50. For each feature we fitted a single Gaussian of the form $P(x_i|Y = y_k)$ corresponding to each class. Where i and k refers to ith feature and kth class respectively. The classifier is initially trained offline with few numbers of samples using maximum likelihood estimation. Subsequently we updated the learning parameters of each Gaussian corresponding to each class in the online fashion as discussed in section 1.3.1 and [64]. The value of constant α was empirically decided to be 0.55 for the experiments. We did not create ensemble of classifiers and semi-supervised learning as we did not have enough data per class to create different classifiers. But creating an ensemble using a randomization technique is a powerful technique to create a boost in accuracy of the classifier. We shall try this once we have more data.

Few parameters used in the learning method employed are as follows:

A) c_k which keeps track of the number of samples that have been used for each class for online training.

B) α it is a constant which decides the length of the memory of the classifier, it influences the value of η_k which decides how much weight must be given (in general and when the convergence (when $c_k \rightarrow \infty$) is achieved) to the new incoming samples in the learning parameter estimation step.

C) η_k is the learning rate parameter whose value depends on the value of count of samples and the value of α and the value of c_k. The detailed discussion over the role of these parameters in learning process is given in [64].

The results table shows the accuracy of the applied method under different settings of the experiments. Considering the fact that Naive Bayes is a classifier which is based upon a very strong assumption of conditional independence among the features given the class, it is performing reasonably well. Therefore it can be assumed that the stronger version of incremental

learning algorithms could perform even better. This experimental process shows that larger accuracies can be achieved using incremental learning in the biometric domain by more experimentation with other adaptive learning techniques available in the machine learning literature.

Table 1.1. Results show that adaptive/online Naïve Bayes classifier is more accurate. Adaptive classifiers are initially trained with a few samples in batch/offline mode and later learn in an online/adaptive manner i.e. using one sample at a time. For example: row 2:- classifier was initially trained with 5 samples per class, subsequently it was adaptively trained with 2 samples per class and later tested with 3 samples per class.

Classifier Name	Training/Testing	Accuracy (%)
Naive Bayes (NB)	7 offline/3	60
Online/adaptive NB	5 offline + 2 online/3	72
Online/adaptive NB	5 offline + 3 online/2	86
Online/adaptive NB	5 offline + 4 online/1	88

1.5 Conclusions

In this chapter we discussed many potential existing adaptive learning methods, which can be applied in the biometric domain to create a robust adaptive biometric system. To demonstrate this we applied one existing learning method to create an adaptive multi-modal ocular approach using iris and sclera. It is evident from the experiments that the adaptive/incremental system applied outperforms the base classifier performance. It is also evident from the results that when the numbers of samples are increased, the adaptability also improves. Due to lack of availability of large number of samples, other theories like online semi-supervised learning methods were not tested.

In future we plan to explore all the promising adaptive techniques in the biometric domain and build a reliable never ending learning biometric system. In that regard we are in the process of collecting more data so that more sophisticated learning methods based on incremental semi-supervised learning can be applied to gain the leverage from the availability of large data.

References
1. R. Derakhshani, A. Ross, and S. Crihalmeanu. A new biometric modality based on conjunctival vasculature. Proceedings of Artificial Neural Networks in Engineering: 1-8, 2006.

2. Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp. Multi angled sclera recognition. IEEE Workshop on Computational Intelligence in Biometrics and Identity Management: 103 – 108, 2011.
3. Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp. Multimodal eye recognition. Proceedings of the International Society for Optical Engineering, 7708(770806):1-10, 2010.
4. Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp. A new biometric sclera recognition. IEEE transaction on System, Man AndCybernatics –PART A: System And Human, 42(3): 571-583, 2012.
5. Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp, Quality Fusion Based Multimodal Eye Recognition, IEEE International Conference on Systems, Man, and Cybernetics: 1297-1302, 2012.
6. M. H. Khosravi and R. Safabakhsh, Human eye sclera detection and tracking using a modified time-adaptive self-organizing map, Pattern Recognition, 41: 2571 – 2593, 2008.
7. S.Crihalmeanu and A. Ross, Multispectral sclera patterns for ocular biometric recognition, Pattern Recognition Letters, 33: 1860–1869, 2012.
8. S. P. Tankasala, P. Doynov, R. R. Derakhshani, A. Ross and S. Crihalmeanu, Biometric Recognition of Conjunctival Vasculature using GLCM Features, International Conference on Image Information Processing: 1-6, 2011.
9. K. Oh and K. Toh, Extracting Sclera Features for Cancelable Identity Verification, 5th IAPR International Conference on Biometric: 245-250, 2012.
10. V. Gottemukkula, S. K. Saripalle, S. P. Tankasala, R. Derakhshani, R. Pasula and A. Ross, Fusing Iris and Conjunctival Vasculature: Ocular Biometrics in the Visible Spectrum, IEEE Conference on Technologies for Homeland Security: 150-155, 2012.
11. Md. H. Kabir, T. Jabid, O. Chae, A Local Directional Pattern Variance (LDPv) based Face Descriptor for Human Facial Expression Recognition, In Proceedings of the IEEE Advanced Video and Signal Based Surveillance (AVSS): 526–532, August 2010
12. S. Lazebnik, C. Schmid, and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, In *Proc.* Computer Vision and Pattern Recognition (CVPR), 2: 2169–2178, 2006.
13. H. Proença and   L. A. Alexandre, UBIRIS: A noisy iris image database,Proceed. of ICIAP 2005 - Intern. Confer. on Image Analysis and Processing, 1: 970—977, 2005.
14. A. Das, U. Pal, M. F. A. Ballester and M. Blumenstein, A New Method for Sclera Vessel Recognition using OLBP, Chinese Conference on Biometric Recognition ,LNCS 8232: 370–377,2013.
15. M. A. Ferrer, A. Morales, A. Das, M. Blumenstein and U. Pal, Model based Sclera vessels segmentation with SIFT Recognition and its combination with Iris, Spanish biometric consodium, VII Jornadas de Reconocimiento Biometrico de Personas: 68-76, 2013

16. A. Das, U. Pal, M. F. Ballester and M. Blumenstein, Sclera Recognition Using D-SIFT, In 13th International Conference on Intelligent Systems Design and Applications: 74-79, 2013.
17. A. Das, U. Pal, M. Blumenstein and M. F. Ballester, Sclera Recognition - A Survey, Advancement in Computer Vision and Pattern Recognition: 917 -921, 2013.
18. J. C. Dunn, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics 3: 32-57, 1973.
19. J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
20. B. N. Li, C. K. Chui, S. Chang and S. H. Ong, Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. Computers in Biology and Medicine 41(1): 1-10, 2011.
21. S. M. Pizer, E. P. Amburn and J. D. Austin, Adaptive Histogram Equalization and Its Variations, Computer Vision, Graphics, and Image Processing 39: 355-368, 1987.
22. I. Daubechies, Ten lectures on wavelets, CBMS-NSF conference series in applied mathematics, SIAM Ed: 117–119, 1992.
23. H. William, A. Saul, T. William, B. P. Flannery, Support Vector Machines. Numerical Recipes: The Art of Scientific Computing (3rd ed.). New York: Cambridge University Press. ISBN 978-0-521-88068-8, 2007.
24. A. Das, U. Pal, M. F. Ballester and M. Blumenstein, Fuzzy Logic Based Sclera Recognition, In FUZZ-IEEE: 561-568, 2014.
25. A. Rattani, B. Freni, G. L. Marcialis, F. Roli, Template update methods in adaptive biometric systems: A critical review'. Proc. Third Int. Conf. Biometrics, Sardinia, Alghero: 847–856, 2009.
26. A. Das and R. Parekh, Iris Recognition using a Scalar based Template in Eigen-space, International Journal of Computer Science and Telecommunication, 3(5): 74-79, 2012.
27. A. Das and R. Parekh, Iris Recognition in 2D Eigen-space, International Journal of Computer Applications, 52(19): 1-6, 2012.
28. A. Das, Face Recognition in reduced Eigen plane, International Conference on Communications, Devices and Intelligent Systems: 620- 623, 2012.
30. N. Poh, A. Rattani and F. Roli, Critical Analysis of Adaptive Biometric Systems, IET Biometrics, 1(4): 179 - 187, 2012.
31. A. Rattani, G. L. Marcialis, E. Granger, F. Roli, A Dual-staged Classification-Selection Approach for Automated Update of Biometric Templates, Intl Conf. on Pattern Recognition (ICPR), Japan: 2972 - 2975 , 2012.
32. A. Rattani,: Adaptive biometric system based on template update procedures. PhD thesis, University of Cagliari, Italy, 2010
33. A. Rattani, B. Freni, G. L. Marcialis, F. Roli: Template update methods in adaptive biometric systems: A critical review. Proc. Third Int. Conf. Biometrics, Sardinia, Alghero, 2009, pp. 847–856

34. N. Poh, J. Kittler, S. Marcel, D. Matrouf, J. F. Bonastre.: Model and score adaptation for biometric systems: coping with device interoperability and changing acquisition conditions. Proc. Int. Conf.on Pattern Recognition, Istambul, Turkey, 1229–1232, 2010.

35. U. Uludag, A. Ross, A. Jain: Biometric template selection and update: A case study in fingerprints, Pattern Recognition., 37, (7): 1533–1542, 2004.

36. X. Jiang, W. Ser: 'Online fingerprint template improvement', IEEE Trans. Pattern Anal. Mach. Intell., 24, (8), .1121–1126, 2008.

37. C. Ryu, K. Hakil, A. K. Jain: 'Template adaptation based finger print verification'. Proc. 18th Int. Conf. on Pattern Recognition, HongKong, 582–585, 2006.

38. X. Liu, T. Chen, S. M. Thornton: Eigenspace updating for nonstationaryprocess and its application to face recognition', PatternRecognit., 36, (9), 1945–1959, 2003.

39. F, Roli, L. Didaci, G. L. Marcialis:Template co-update in multimodalbiometric systems. Proc. IEEE/IAPR Int. Conf. on Biometrics, Seoul,Korea, , 1194–1202, 2007.

40. F. Roli, G. L. Marcialis: Semi-supervised PCA-based face recognition using self-training. Proc. Joint IAPR Int. Workshop on S+SSPR06,HongKong, China, 560–568, 2006.

41. A. Rattani, G. L. Marcialis, F. Roli: Biometric template update usingthe graph mincut: a case study in face verification. Proc. Sixth IEEEBiometric Symp., Tampa, USA, 23–28, 2008.

42. M. Scholz, R Klinkenberg, Boosting classifiers for drifting concepts, IDA – Special Issue on Knowledge Discovery from Data Streams 11 (1): 3–28, 2007.

43. J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM, 51, 107–113, 2008.

44. M. A. Maloof, R. S. Michalski, Incremental learning with partial instance memory," Artificial Intelligence, 154(1-2), 95–126, 2004.

45. A. Shilton, M. Palaniswami, D. Ralph, A. C. Tsoi, Incremental training of support vector machines, IEEE Transactions on Neural Networks, 16, pp. 114– 131, 2005.

46. M. Muhlbaier, A. Topalis, R. Polikar, Learn++.NC: Combining Ensemble of Classifiers Combined with Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes," IEEE Transactions on Neural Networks, 20(1), 152–168, 2009.

47. G. Widmer, M. Kubat, Learning in the Presence of Concept Drift and Hidden Contexts, Machine Learning, 23, 69–101, 1996.

48. P. P. Rodrigues, J. A. Gama, J. A. Araújo, L. Lopes, L2GClust: local-to-global clustering of stream sources, in Proceedings of the 2011 ACM Symposium on Applied Computing, 1006–1011, ACM, 2011.

49. J. Z. Kolter, M. A. Maloof, Dynamic Weighted Majority: An Ensemble Method for Drifting Concepts, J. Mach. Learn. Res., 8, 2755–2790, December 2007.

50. N. W. Street, Y. Kim, A streaming ensemble algorithm (SEA) for large-scale classification, in KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 377–382, ACM, 2001.
51. R. Klinkenberg, Learning drifting concepts: Example selection vs. example weighting, Intelligent Data Analysis, 8, 281–300, 2004.
52. A. Fern, R Givan, Online ensemble learning: An empirical study, Machine Learning 53: 71–109, 2003.
53. R. Polikar, Bootstrap inspired techniques in computational intelligence, IEEE Signal Processing Magazine, 24, 57–72, 2007.
54. L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons, Inc., 2004.
55. G. Brown, Ensemble learning, Encyclopedia of Machine Learning, 1-9, 2010.
56. D. H. Wolpert, W. G. Macready, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation, 1(1), 67–82, 1997.
57. T. G. Dietterich, Machine learning research: Four current directions, Artificial Intelligence 18(4): 97–136, 1997.
58. T. G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization, Machine Learning 40(2): 139–157, 2000.
59. L. I. Kuncheva, C. J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Machine Learning 51: 181–207, 2003.
60. L. Breiman, Random forests, Machine Learning 45: 5–32, 2001.
61. E. KTang.,P. N. Sunganthan, X. Yao, An analysis of diversity measures, Machine Learning 65: 247–271, 2006.
62. L. L.Minku, A. White, X. Yao, The impact of diversity on on-line ensemble learning in the presence of concept drift, IEEE Transactions on Knowledge and Data Engineering (TKDE) 22: 730–742, 2010.
63. L. L. Minku, X. Yao, DDD: A new ensemble approach for dealing with concept drift, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2010.
64. R. Kunwar, U. Pal and M. Blumenstein. Semi-Supervised Online Learning of Handwritten Characters Using a Bayesian Classifier. Second IAPR Asian Conference on Pattern Recognition, Okinawa, Japan, 2013.
65. R. Kunwar, U. Pal and M. Blumenstein. Semi-Supervised Online Bayesian Network Learner for Handwritten Characters Recognition. Twenty second International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, 2014.
66. V. Vladimir, V.Vapnik. Statistical Learning Thoery. Springer-Verlag New York, 1998.
67. N. A. Syed, H. Liu, and K. K. Sung. Incremental learning with support vector machines. In SVM workshop, IJCAI, 1999.
68. S. Ruping. Incremental learning with support vector machines. Technical Report TR-18, Universitat Dortmund, SFB475, 2002.

69. L. Ralaivola, F. d'Alche Buc. Incremental support vector machine learning: A local approach. Lecture Notes in Computer Science, 2130:322–329, 2001.

70. J. Kivinen, A. J. Smola, R. C. Williamson. Online learning with kernels. In T. G. Diettrich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Inf. Proc. Systems (NIPS 01), 785–792, 2001.

71. G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, Advances in Neural Information Processing Systems,13, 409–415. MIT Press, 2001.

72. P. Laskov, C. Gehl, S. Kruger, K.R. Muller. Incremental Support Vector Learning: Analysis, Implementation and Applications. Journal of Machine Learning Research, 1909–1936, 2006.

73. A. Shilton, M. Palaniswami, D. Ralph, A. C. Tsol, Incremental Training of Support Vector Machines, IEEE transactions on Neural Networks, 16(1), 114-131, 2005.

74. F. Shen and O. Hasegawa, An Incremental Network for On-line Unsupervised Classification and Topology Learning, Neural Networks, 19(1), 90-106, 2006.

75. F. Shen, K. Sakurai, Y. Kamiya, O. Hasegawa: An Online Semi-supervised Active Learning Algorithm with Self-organizing Incremental Neural Network, International Joint Conference on Neural Network (IJCNN), 2007.

76. T. Kohonen, Self-organized Formation of Topologically Correct Feature Maps, Biol. Cybern, 43(1), 59-69, 1982.

77. T.M. Martinetz, S.G. Berkovich, K. J. Schulten, Neural-Gas Network for Vector Quantization and its Application to Time-Series Rrediction, IEEE Trans. Neural Networks, 4(4), 558-569, 1993.

78. Y. Freund, R. E. Schapire, Experiments with a new boosting algorithm, Proceedings of the Thirteenth International Conference on Machine Learning (ICML'96), Morgan Kaufmann, Bari, Italy, 148–156, 1996.

79. Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences 55(1): 119–139, 1997.

80. B. Zhang, Y. Gao, S. Zhao, J. Liu, Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor, IEEE TRANSACTIONS ON IMAGE PROCESSING,19(2), 2010

81. J. G. Daugman: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15(11), .1148–1161, 1993.