WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia IJCNN

Incremental Update of Biometric Models in Face-Based Video Surveillance

Miguel De-la-Torre*†, Eric Granger*, Paulo V. W. Radtke*, Robert Sabourin*, Dmitry O. Gorodnichy‡
* École de technologie supérieure, Montréal, Canada
{miguel@livia.etsmtl.ca, eric.granger@etsmtl.ca, radtke@livia.etsmtl.ca, Robert.Sabourin@etsmtl.ca}
†Centro Universitario de Los Valles, Universidad de Guadalajara, Ameca, México
‡Science and Engineering Directorate, Canada Border Services Agency, Ottawa, Canada
{dmitry.gorodnichy@cbsa-asfc.gc.ca}

***Abstract*—Video-based face recognition of individuals involves matching facial regions captured in video sequences against the model of individuals enrolled to a face recognition system. Due to a limited control over operational conditions, classification systems applied to face matching are confronted with complex pattern recognition environments that change over time. Therefore, the facial model of an individual tends to diverge from the underlying data distribution. Although a limited amount of reference data is often collected during initial enrollment, new samples often become available over time to update and refine models. In this paper, an adaptive ensemble of classifiers is proposed to update facial models in response to new reference samples. To avoid knowledge corruption linked to incremental learning of monolithic classifiers, and maintain a high level of performance, this ensemble exploits a learn-and-combine approach. In response to new reference samples, a new 2-class Probabilistic Fuzzy ARTMAP classifier is trained and combined to previously-trained classifiers in the ROC space. Iterative Boolean Combination is employed for fusion of 2-class classifiers of each individual in the decision space. Performance is assessed in terms of AUC accuracy and resource requirements under different incremental learning scenarios with new data extracted from the Faces in Action data set. Simulation results indicate that the proposed system significantly outperforms reference classifiers and ensembles for incremental learning.**

I. Introduction

The emergence video surveillance networks are comprised of a growing number of low cost digital IP video cameras, and video analytics systems are seen as solution to manage massive quantities of data obtained from the growing number of low cost digital IP video cameras. It has become increasingly difficult for human operators to analyse all video feeds. In this context, video-based face recognition is becoming an important function for decision support in enhanced surveillance and security systems. Biometric systems allow to perform automated recognition of individuals based on their physiological and/or behavioral traits [1].

In video surveillance applications, faces appearing in a video stream can be rapidly and covertly detected, tracked and recognized to determine whether these faces correspond to individuals of interest. Facial regions acquired from a scene are subject to considerable variations due to limited control over operational conditions from unconstrained scenes (e.g., illumination, pose, facial expression, orientation and occlusion). Moreover, the physiology of individuals may change over time, and new knowledge may suddenly emerge whereas previously acquired data may eventually become obsolete in dynamically changing classification environments.

Face recognition (FR) systems are comprised of a biometric database with a facial model for each individual enrolled to the system. Facial (biometric) models may consist of a set of one or more templates –genuine reference samples acquired during enrollment process. To improve robustness and reduce resources, biometric models may also consist of parameters of a statistical estimation obtained by training a classifier. Neural or statistical classifiers then implicitly define a model of an individual's facial traits by mapping the finite set of reference samples to a decision space.

The collection and analysis of labeled biometric data from individuals is often expensive and time consuming. In such cases, classifiers are designed during an a priori enrollment phase using sparse and unbalanced reference samples collected according to an unknown data distribution. Facial BMs are often poor representatives of faces to be recognized during operations [2]. The underlying data distribution corresponding to individuals enrolled to the system is complex mainly due to inter- and intra-class variability, to changes that occur during operations, to variations in the interaction between sensor and individual, to the large number of input features and individuals, and to limitations of cameras and signal processing techniques [3]. The performance of biometric systems may decline considerably because state-of-the-art neural and statistical classifiers employed for matching depend heavily on the availability of representative reference data and prior knowledge. During operations, the probability of seeing an individual of interest in scenes may be quite low, and BMs may incorporate a significant amount of uncertainty with respect to the unconstrained video scene. In addition, new information may suddenly emerge, and underlying data distributions may change gradually or abruptly in the classification environment. Performance may decline over time as BMs deviate from the actual data distribution [3], [2], [4].

Some adaptive biometric systems have been proposed in the literature to refine biometric models (BMs) according to the intra-class variations in input samples [5]. With self-adaptive or semi-supervised learning strategies, BMs are initially designed during enrollment using labeled training data, and then

978-1-4673-1490-9/12/$31.00 ©2012 IEEE

updated with highly confident unlabeled data obtained during operations [6], [2]. These strategies are however vulnerable to outliers, dispersion and overlap in class distributions. Stringent criteria are required for selection of highly confident data, to minimize the probability of introducing imposter data into updated biometric models.

In this article, supervised learning strategies are considered, and new data samples are assumed to be analyzed and labeled by an operator with expert knowledge of intra-class variations. If labeled data becomes available, for instance, over multiple enrollment sessions, or when operational videos are analyzed off-line, and can allow an operator to gradually build facial BMs over time. Adaptive biometric systems in literature have used newly-acquired reference samples to update the selection of a user's template from a gallery via clustering and editing techniques [4]. Others have performed on-line learning of genuine samples over time to update each user's single super template [7]. It is however difficult to represent intra-class variations with a single template [5].

Neural and statistical classifiers that allow for supervised incremental learning of new data provide the means to efficiently maintain accurate and up-to-date face models of individuals. However, the decline of performance caused by knowledge corruption (i.e. BMs) remains a fundamental issue for monolithic classifiers [8]. As with most research in literature, these techniques are suitable for designing classification systems with an adequate number of samples acquired from ideal and static environments, where class distributions are balanced and remain unchanged over time. Ensemble-based techniques like Learn++ [9] and its variants, where classifier ensemble are trained independently on the newly-acquired training data and then combined with previously-trained classifiers may provide a more robust solution. Ensemble-based approaches may avoid knowledge corruption at the expense of growing complexity, and the need to store reference samples in a Long Term Memory (LTM) for validation.

In this paper, an efficient adaptive ensemble of classifiers is proposed to update facial models in response to new reference samples. To avoid knowledge corruption linked to incremental learning (IL) of classifiers, this ensemble exploits a learn and combine approach. When new reference samples become available, a new classifier is trained and combined to previously-trained classifiers in the ROC space. Iterative Boolean Combination [10] is employed to combine 2-class PFAM classifiers of an individual in the decision space.

The performance of the learn-and-combine classification is compared to other IL approaches for updating BMs in face-based video surveillance. The learning strategy is characterized in different scenarios, including learning ones using limited data, and using a LTM to store reference samples for validation and design of fusion functions. The Face in Action database [11] is employed for proof-of-concept simulations. Video streams are pre-processed with state-of-the-art segmentation, and feature extraction and selection algorithms on regions of interest (ROIs). The classification sub-system is maintained with the learn-and-combine strategy. Reference classification algorithms include k-NN (k- Nearest Neighbor) and PFAM that learn in batch mode (retraining with all accumulated data), and PFAM and Learn++ in incremental mode.

The rest of the paper is organized as follows. Section II provides survey on state of the art techniques used for video based FR and open set classification in biometrics. In section III different techniques for adaptive biometrics are discussed, focusing on classification approaches for IL. The learn-and-combine approach for IL is then described in Section IV. The experimental methodology used in comparison is depicted in Section V, and simulation results are presented and discussed in Section VI.

II. Face Recognition in Video-Surveillance

Assume that 2D images are captured in one or more video cameras. Face recognition in video involves several processing steps. First, the segmentation process isolates the facial ROIs corresponding to face appearing in subsequent input frames. Among a wide range of techniques appearing in literature, appearance-based methods for image segmentation like the Viola-Jones algorithm, have been shown to efficiently detect facial ROIs in video using a high level of accuracy.

Next, the tracking function follows the movement or expression of faces across video frames, while the classification function seeks to match input feature patterns to the face models of individuals enrolled to the system. The tracking features are typically the position in frames, speed, acceleration, and track number assigned to each ROI in the scene, while classifiers will require invariant and discriminant classification features extracted from ROIs. Feature extraction modules then extract specific characteristics for, e.g., particle filter tracking, on feature vector $\mathbf{b}$, and classification on feature vector $\mathbf{a}$ extracted and selected using, e.g., Local Binary Patterns or Principal Component Analysis. During classification, input $\mathbf{a}$ is compared to the facial model of individual i stored within a biometric database, producing a similarity score $S_i(\mathbf{a})$.

Biometric matching may be implemented by training a statistical or neural network classifier using reference data. With neural network classifiers, for instance, the biometric model (BM) of individuals is defined using the hyperparameters, synaptic weights, and architecture. Finally, for each video frame, the decision module may combine and accumulate the responses from the tracking and classification modules. Several powerful techniques have been proposed to recognize faces in static 2D images [12]. A common approach to recognize faces in video consists in exploiting only spatial information (*i.e.*, appearance), and applying extensions of static image-based techniques on high quality face images produced through segmentation. The predominant techniques are appearance-based methods like Eigenfaces, and feature-based methods like Elastic Bunch Graph Matching [12].

FR systems for video-surveillance may exploit spatio-temporal information on the appearance and motion of faces detected in a scene. The advantages of video FR include an increase in contextual knowledge and data in video [13]. For example, track-and-classify systems may combine spatial

information with information on motion and appearance of faces in a scene [14]. Given a video sequence, the ROIs corresponding to an individual may be tracked, and the responses may be accumulated over time for improved performance. Regardless, the performance of these techniques may degrade considerably when applied in real-world applications.

One of the main challenges is that BMs are not representative because they are designed using limited and incomplete data captured from uncontrolled environments. Facial captures are then subject to considerable variations due to limited control over operational conditions when acquiring images from unconstrained scenes (e.g., illumination, pose, facial expression, orientation and occlusion). Moreover, physiology of the individuals may change over time, either temporary (e.g., haircut, glasses, etc.) or permanently (e.g., scars, aging). New informations, such as input features and output classes, may suddenly emerge and previously acquired data may eventually become obsolete in dynamically changing classification environments.

Video Surveillance problems are addressed as an open-set or open-world problem, where the number of individuals of interest is greatly outnumbered by other individuals. Li and Wechsler [15] proposed a variant of TCM-kNN (Transduction Confidence Machine- k-NN) for surveillance applications which considers new input patterns in order to tune a specialized rejection threshold. Tax and Duin also propose a multi-class classifier of 1-class binary classifiers per class, in which posterior probabilities are normalized to apply a common rejection threshold across all classes, but adapted to each distribution [16].

In other applications like speaker recognition, the "Universal Background Model" (UBM) is widely used for better discrimination between target voice from all other sounds [17]. This UBM is built by selecting samples of the background sound that characterizes a recording environment, and is used to discriminate between the individual (speaker) of interest and other sounds. In the same manner, the cohort model is a set of selected samples from non-target samples from already known voices to discriminate known individuals from other known speakers. These cohort and UBM models constitute an important source of discriminative information for system design.

In this paper, a Universal Face Model formed with samples from individuals that do not appear in the watch list, and a Cohort Face Model comprised of samples from other individuals within the watch list. The rejection threshold is estimated per individual using a validation data set that allows the selection of an operating point in the ROC space.

III. Adaptive Biometrics

The statistical representation of individual traits using biometric models often diverges from the real biometric trait due to intra-class variability, aging, varying capture conditions, different capture conditions, and interaction with the capture systems. Adapting biometric models of interest to new data can be used to limit the impact of such variability.

The collection and analysis of biometric data from individuals is often expensive and time consuming. In such cases, classifiers are therefore designed using some prior knowledge of the underlying data distributions, and a limited amount of learning data. It is possible however to acquire new facial images at some point in time after a classifier has originally been trained and deployed for operations. These can be acquired during re-enrollment sessions, or during post-analysis of video streams [1].

For accurate and timely recognition of individuals, it is important to efficiently adapt facial models over time in response to new or changing input features, data samples, priors, classes and environments. Adaptive biometric systems in literature traditionally incorporate newly-acquired reference samples to update the selection of a user's template from a gallery via clustering and editing techniques [4], to improve representation of intra-class variations with a single template, and some adaptive biometric systems have been proposed to refine BMs according to the intra-class variations in input samples [5]. Others have performed on-line learning of genuine samples over time to update each user's single super template [7].

With self-adaptive or semi-supervised learning strategies, biometric models are initially designed during enrollment using labeled training data, and then updated with highly confident unlabeled data obtained during operations [6], [2]. These strategies are however vulnerable to outliers, dispersion and overlap in class distributions. Stringent criteria are required for selection of highly confident data, to minimize the probability of introducing imposter data into updated BMs.

In this paper, supervised learning strategies are considered, and new data samples are assumed to be analyzed and labeled by an operator with expert knowledge of intra-class variations. If labeled data becomes available, for instance, over multiple enrollment sessions, or when operational videos are analyzed off-line, and can allow an operator to gradually build facial BMs.

Neural and statistical classifiers that allow for supervised incremental learning of new data provide the means to efficiently maintain an accurate and up-to-date face model of individuals.

However, the decline of performance caused by corruption of knowledge (BMs) remains a fundamental issue for monolithic classifiers. These techniques are suitable for designing classification systems with an adequate number of samples acquired from ideal and static environments, where class distributions are balanced and remain unchanged over time. Some high-level architectures, based on well-known pattern classifiers, e.g., Ensembles of Classifiers, have also been proposed [9], where classifiers are trained independently on the newly-acquired training data, and then combined with previously-trained classifiers.

In literature, some promising pattern classification algorithms have been reported for supervised IL in environments where distributions are fixed. For example, the ARTMAP [18] and Growing Self-Organizing [19] families of neural network classifiers, have been designed with the inherent ability to

perform IL. In addition, some well-known pattern classifiers, such as the Support Vector Machine (SVM), and the Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) neural networks have been adapted to perform IL [20], [21], [22].

A. Probabilistic Fuzzy ARTMAP

A representative example of ARTMAP classifiers is Fuzzy ARTMAP, which integrate a Fuzzy ART model to process analog and binary valued inputs to the ARTMAP architecture. The probabilistic variant proposed by Lim and Harrison in [23], combines the Fuzzy ARTMAP learning to encode prototype vectors and update centers of mass of estimated distributions. Then, the probabilistic neural network is used for probability estimation. In this way, the output for an input pattern $\mathbf{a}$ for each category j is represented as a hyper-spherical gaussian probability density function

$$g_j(\mathbf{a}) = \frac{1}{(2\pi)^{M/2}\sigma_j^M} e^{\left(-\frac{(\mathbf{a}-\mathbf{w}_j^{a-c})^T(\mathbf{a}-\mathbf{w}_j^{a-c})}{2\sigma_j^2}\right)}, \qquad (1)$$

where the variance σ_j is the ratio of the sqared minimum Euclidean distance between $\mathbf{w}_j^{a-c}$ and any other center vector, to the value of an overlap parameter $r > 0$.

B. Learn++

A well-known approach for adapting ensembles incrementally is Learn++ proposed by Polikar et al. in [9]. This technique is inspired on the AdaBoost algorithm, and incorporates a new set of classifiers to the ensemble every time new data becomes available. In the original work the classical MLP was used as the base classifier, with its parameters adjusted to preserve resources and not necessarily produce a high accuracy. The generation of the pool of classifiers each time a new dataset D_t becomes available, is performed using a bagging strategy, by training distinct instances of MLPs on bootstrap replicates of the training set. Selection criteria integrate a *fixed size* set of classifiers, in which every classifier produces an average error lower than random selection ($\epsilon < 1/2$), and new added classifiers does not increase the overall classification error over random selection ($\epsilon_{global} < 1/2$).

IV. Learn and Combine Approach

Fig. 1 presents an adaptive multi-classifier system (MCS) for face recognition in video surveillance that allows for update facial models in response to new reference samples. It is composed of a LTM, an ensemble of binary 2-class classifiers or detectors (EoDs) $\mathbf{P}_i$ per individual, and a dynamic optimization module. To avoid knowledge corruption linked to IL of classifiers, this ensemble exploits a learn-and-combine approach. When new reference samples become available, a new classifier is trained and combined to previously-trained classifiers in the ROC space. Iterative Boolean Combination is employed to combine 2-class PFAM classifiers of an individual in the decision space.

Fig. 2 depicts the learn-and-combine strategy used to update the ensemble of classifiers in each module $\mathbf{P}_i, 1 \leq i \leq k$.

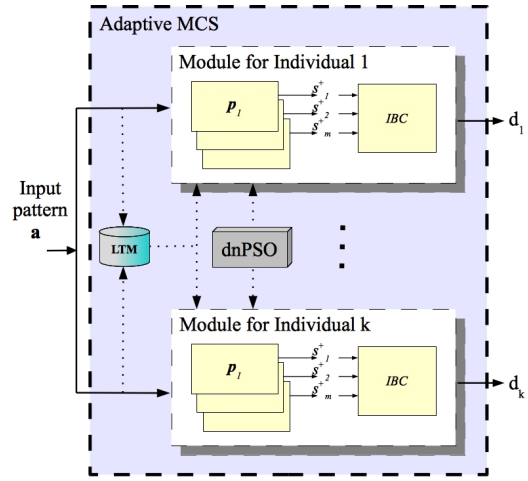


Fig. 1. Adaptive MCS for FR in video surveillance. Dotted arrows indicate pathways for enrollment/update with validation data.

When a new data block D_t is available, a subset is used for training D_t^t, and the rest is stored in the LTM. Three independent validation sets are maintained within the LTM, D_t^e to stop the training epochs of classifiers, D_t^f for fitness evaluation (PFAM parameter optimization) and D_t^c to estimate the fusion function and thresholds using iterative boolean combination (IBC). The learning strategy is based in the dnPSO optimization algorithm [24]. It generates a diversified pool of PFAM classifiers in the hyperparameter space, and the global best p_t is selected and added to the ensemble $\mathbf{P}_i$. The combination function for $\mathbf{P}_i$ is then updated using IBC and validated on D_t^c. Finally, an operating point is selected for prediction in the ROC space based on a false alarm rate pre-defined by an human operator.

Iterative boolean combination (IBC) [10] has been used to fuse multiple crisp or soft 1- or 2-class classifiers at decision level in the ROC space. IBC has been shown to outperform reference techniques and has a linear complexity with respect to the number of classifiers. Given an ensemble of classifiers $\mathbf{P}_i = \{p_1, \ldots, p_t\}$, IBC starts by combining all pairs of operating points (ROC space vertices) for two classifiers according to different Boolean functions. The convex hull of the newly generated operating points are successively combined with operating points of the remaining classifiers, one at a time, until all classifiers have been combined to provide an overall convex hull. The proposed learn-and-combine approach uses IBC to combine previously-trained classifiers with those trained on new data.

V. Experimental Methodology

The Face in Action (FIA) database [11] is used to observe the impact on performance when BMs are updated using the learn and combine strategy. The FIA database consists of 20 second videos of face data from 180 individuals mimicking a passport checking scenario. Grayscale frames are extracted from video sequences for training (session 1), and test (session 2 captured three months later). Only frontal indoor images

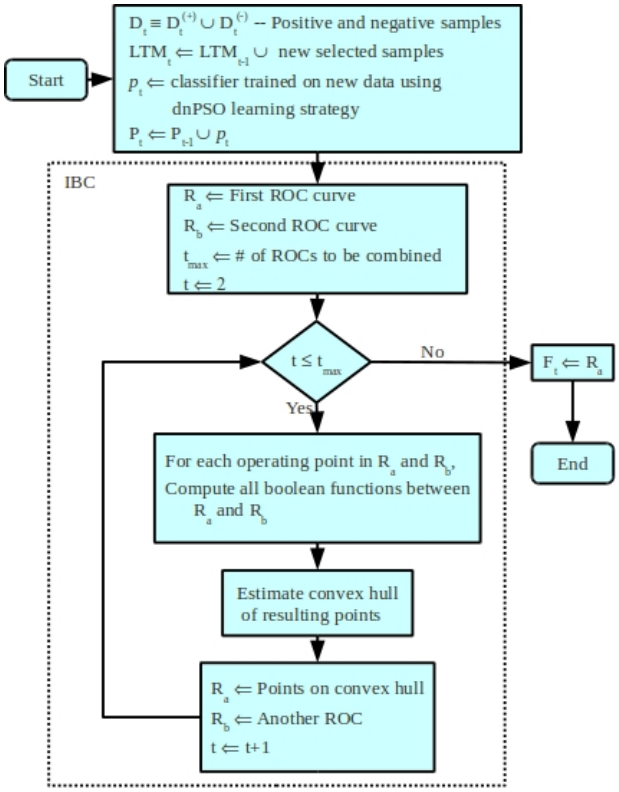


Fig. 2. Algorithm for learn-and-combine strategy employed for incremental learning of new reference samples with the system shown in Fig 1.

are used from both focal lengths (4-mm and 8-mm), and all images are resized to the highest possible resolution of the smallest face obtained after face detection with the well known Viola-Jones algorithm (70x70 pixels). MSLBP (Multi Scale Local Binary Patterns) [25] is used as a feature extractor with three different block sizes (3×3, 5×5 and 9×9), along with pixel intensities features. Resulting features are stacked in feature vectors, and PCA is applied selecting the 32 principal characteristics.

Ten individuals of interest are randomly selected to form the watch list (labeled as 176, 190, 209, 188, 147, 151, 58, 2, 23 and 106). The 2-class classification module $\mathbf{P}_i$ is trained using a balanced set of samples: 50% of positives samples from the individual of interest, and the remaining 50% of negative samples are drawn uniformly from the cohort model (other 9 individuals in the watch list) and universal model (samples from 88 random unknown individuals outside the watch list). Other individuals are considered unknown individuals (negative class), and their samples appear only in the test set. Training samples for each module are randomly distributed in five blocks, $D_t, 1 \leq t \leq 5$, each preserving the proportion of data. Each data block D_t has a fixed size, where $|D_t^t| = 32$ (16 samples per class) and validation data sets D_t^e, D_t^f and D_t^c all have 16 samples (8 samples per class). The test set D_{tst} contains a total 41,044 samples, from which $x, 309 \leq x \leq 469$ samples are from the positive class. Negative class samples in D_{tst} are as follows: $3736 - x$ samples from the cohort model, 9,270 samples from the universal facial model and 23,038 samples from individuals never seen by the system.

To asses the performance of the learn-and-combine approach, a comparison with reference classification algorithms is conducted. They are the k-NN and PFAM that learn in batch mode (retraining with all accumulated data from scratch), PFAM in incremental mode, and Learn++ as the reference ensemble based classifier for IL. Our experiments use PFAM as the base classifier for the Learn++ algorithm, instead of an MLP classifier. PFAM classifiers are trained using a dynamic niching particle swarm optimization (dnPSO) [24] based learning strategy to optimize the PFAM hyper-parameters. Validating the number of epochs up to convergence is performed on D_t^e, whereas particle fitness is evaluated on D_t^f. The dnPSO algorithm has an initial swarm of 60 particles, and a maximum of 5 particles within each of the 6 subswarms. The algorithm is set to run a maximum of 30 iterations, allowing 5 extra iterations to ensure convergence. Once the dnPSO global best particle is found, its classifier is combined to the ensemble.

The fusion function for both learn-and-combine (IBC) and Learn++ (weighted majority voting) are updated using D_t^c. The original Learn++ estimates the weights of the combination function using only the training data set D_t^t. Two different ways to estimate combination weights are also considered. One is to use the validation set D_t^c to estimate the combination weights and the other is to join the validation and training data in a single set ($D_t^t \cup D_t^c$) to estimate the weights. Table I shows the AUC accuracy of these three different ways of using the combination data with Learn++ on an IL scenario. Results indicate that the best option is using $D_t^t \cup D_t^c$ to estimate Learn++ majority vote weights.

TABLE I
LEARN++ ANALYSIS ON A SPECIFIC INDIVIDUAL.

Weights on	D_1	D_2	D_3	D_4	D_5
D_t^t	**0.601** ±0.089	0.618 ±0.074	**0.666** ±0.047	0.639 ±0.075	0.638 ±0.074
D_t^c	0.586 ±0.125	0.609 ±0.133	0.529 ±0.132	0.559 ±0.123	0.650 ±0.114
$D_t^c \cup D_t^t$	0.509 ±0.085	**0.692** ±0.104	0.625 ±0.096	**0.743** ±0.061	**0.670** ±0.074

The experimental protocol simulates a dynamic update scenario, where labeled data for all individuals become available in 5 different data blocks $D_t, 1 \leq t \leq 5$. The experiment is replicated ten times using 2×5-fold cross validation. Incoming data blocks are randomly divided in 5 folds, from which two are used for training (D_t^t), one to stop training epochs D_t^e, one for fitness estimation D_t^f and the last one for classifier fusion function evaluation D_t^c. Each fold is assigned to a different training/validation set at each replica of the experiment to generate a standard error measure on five different assignments. The experiment is repeated a second time with a different selection of samples in folds to complete the 2×5-fold cross validation. Algorithm 1 details the complete experimental procedure. To compare the performance of all classifiers, empirical ROC curves are estimated on the D_{tst}

test data set and compression (ratio between training samples and classification model size) is calculated for each D_t.

Algorithm 1 Experimental protocol for each module.

```
D^t ⇐ Training data (D_t(+)) and UM (D_t(−))
D_tst ⇐ Test data from all individuals
// 2x5 = 10 replicas of the experiment
for r = 1...10 do
  // Data block available at t_max times
  for t = 1...t_max do
    D_t^t ⇐ 2 folds // Training
    D_t^e, D_t^c, D_t^f ⇐ 1 fold
    Train k-NN on batch D_t^t
    Train PFAM_batch on batch D_t^t
    Update PFAM_inc on block D_t^t
    Update the Learn++ on block D_t^t
    Update the L&C on block D_t^t
    Estimate performance on D_tst
  end for
end for
```

To verify the impact of training with limited data, the same experiments is performed with smaller D_t^t training sets. Two different limited set sizes are used, $|D_t^t| = 16$ (8 samples per class) and $|D_t^t| = 4$ (2 samples per class), an extreme case that tries to replicate the limitations faced by actual watch list based screening applications. The use of the LTM is also evaluated, first using an unlimited size LTM, keeping all validation samples for combination with new validation data. A fixed size LTM with 8 samples per class for each validation dataset has also assessed, storing samples from each data block using a FIFO strategy. Every time a new data block is available, the LTM is combined with a random selection from the new validation data and old samples that does not fit in the LTM are discarded.

Given the responses of a 2-class PFAM classifier for a set of test samples, the true positive rate (tpr) is the proportion of positives correctly classified over the total number of positive samples. The false positive rate (fpr) is the proportion of negatives incorrectly classified (as positives) over the total number of negative samples. A ROC curve is a parametric curve in which the tpr is plotted against the fpr. The area under the ROC curve (AUC) or the partial AUC is known as a robust scalar summary of classifiers performance. The AUC assesses ranking in terms of class separation –the fraction of positive-negative pairs that are ranked correctly. For instance, with an AUC = 1, all positives are ranked higher than negatives indicating a perfect discrimination between classes. A random classifier has an AUC = 0.5, and both classes are ranked at random. The average performance of the classifiers are assessed in terms of partial Area Under the ROC Curve (pAUC) for the first 10% of false positive rate. An operating point then can be selected by fixing a maximum allowance of fpr, and selecting its corresponding threshold.

Resources requirement of different classification systems is measured in terms of compression –the average number of training patterns per category prototype. For k-NN the number of prototypes saved comprise the whole training data set, showing a constant $compression = 1$. PFAM based approaches require the space to store a number of neurons in the F2 layer, that in general is lower to the number of training samples. In PFAM based approaches compression is estimated as

$$compression = \frac{|D_t^t|}{\sum_i |F2_i|},$$

where $|D_t^t|$ is the number of training samples and $|F2_i|$ is the number of neurons in each PFAM classifier p_i.

VI. Results

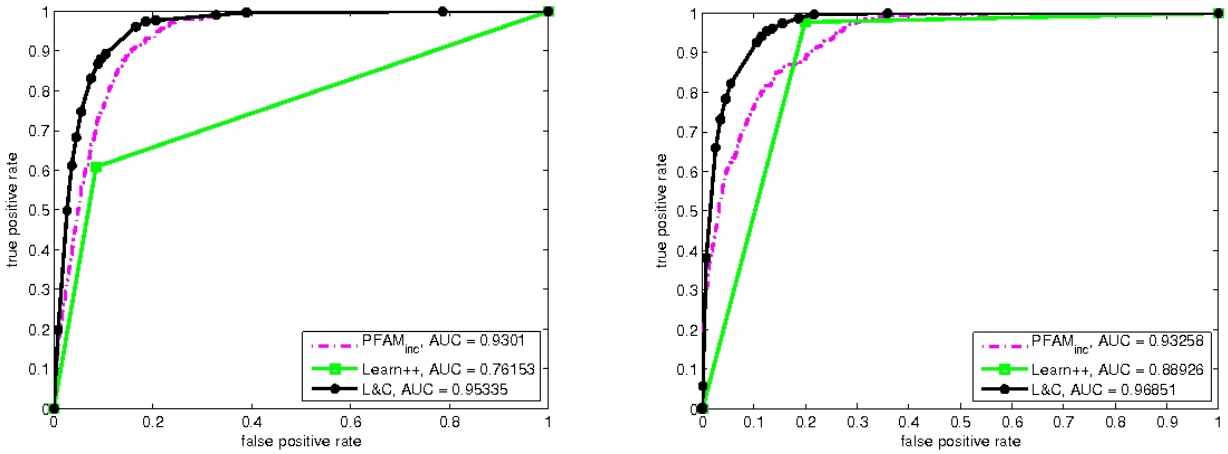
A. Baseline experiment

The learn-and-combine strategy performance is at least comparable to other tested approaches, if not better, as indicated by Table II. Average learn-and-combine AUC values are equivalent to those provided by the PFAM and k-NN classifiers on batch mode. The best compression level is attained by the PFAM incremental classifier, which changes its current classification model when presented to new training data. However, this classifier feature introduces knowledge corruption, indicated by a steady decrease in the average AUC value. Whereas the PFAM batch classifier provides good performance with higher compression levels, the required learning time when presenting new data at an specific time $t, t \geq 2$ is much smaller with learn-and-combine. Instead of learning all data as the PFAM batch classifier, the learn-and-combine approach trains one new PFAM classifier only with new data. Thus, it is preferable to use the learn-and-combine strategy for a better overall performance.

It is interesting to note that whereas AUC values for learn-and-combine are similar to the other compared approaches, the pAUC at 10% fpr (false positive rate) in Table II indicate that learn-and-combine outperforms all other tested approaches at times $t, t \geq 2$ (after at least on incremental update). Fig. 3 details the ROC curves of the three tested incremental approaches at two different times, $t = 2$ and $t = 5$ to better illustrate the evolution of AUC values.

B. Limited Data

Table III presents the performance of the system with limited training data, using both 8 and 2 samples per class in the training set (16 samples per class were used in the baseline experiment). Average AUC values follow the same trends observed in the baseline experiment (Table II), and the same outperformance conclusions regarding the learn-and-combine approach are still valid. Comparing average AUC values between Tables III and II, it can be observed that attained AUC is lower with limited data, but still acceptable given the restricted training data. One possible explanation for those results is that the intra-class variation is smaller than in other classification problems (such as handwritten recognition), and few training samples are required to train a discriminant classifier for a watch list based application. This observation is important for actual watch list based screening



(a) **(b)**

Fig. 3. Sample ROC curves for incremental approaches, module of the individual labeled 190, at replica 3 of the experiment, after update at time $t = 2$ **(a)** and $t = 5$ **(b)**.

TABLE II
AVERAGE PERFORMANCE (BASELINE EXPERIMENT).

Classifier	Compression	AUC	pAUC-10
		D_1	
k-NN	1.00±0.00	0.9127±0.020	0.4798±0.069
PFAM$_{batch}$	8.63±1.13	0.9499±0.022	0.7223±0.068
PFAM$_{inc}$	8.63±1.13	0.9499±0.022	0.7223±0.068
Learn++	**10.43±1.25**	0.8352±0.079	0.5477±0.098
L&C	9.00±1.19	**0.9523±0.024**	**0.7496±0.067**
		D_2	
k-NN	1.00±0.00	0.9345±0.016	0.5735±0.069
PFAM$_{batch}$	**10.38±1.74**	0.9522±0.020	0.7252±0.067
PFAM$_{inc}$	9.97±1.46	0.9440±2.09	0.6817±0.068
Learn++	9.58±0.95	0.7915±9.23	0.5046±0.104
L&C	8.56±1.14	**0.9622±0.019**	**0.7859±0.064**
		D_3	
k-NN	1.00±0.00	0.9398±0.016	0.5933±0.068
PFAM$_{batch}$	10.52±1.52	0.9512±0.018	0.7125±0.074
PFAM$_{inc}$	**11.82±1.83**	0.9382±0.022	0.6502±0.074
Learn++	9.46±0.91	0.8422±0.068	0.5080±0.097
L&C	8.22±1.10	**0.9649±0.018**	**0.7957±0.062**
		D_4	
k-NN	1.00±0.00	0.9486±0.014	0.6409±0.062
PFAM$_{batch}$	11.29±1.77	0.9551±0.020	0.7437±0.071
PFAM$_{inc}$	**13.96±2.39**	0.9343±0.022	0.6254±0.076
Learn++	9.50±0.87	0.8301±0.073	0.4788±0.092
L&C	7.79±0.89	**0.9702±0.015**	**0.8125±0.061**
		D_5	
k-NN	1.00±0.00	0.9496±0.013	0.6442±0.060
PFAM$_{batch}$	11.76±2.24	0.9589±0.014	0.7403±0.065
PFAM$_{inc}$	**16.07±3.00**	0.9164±0.032	0.5965±0.085
Learn++	9.34±0.89	0.8174±0.069	0.4280±0.089
L&C	7.32±0.79	**0.9732±0.013**	**0.8240±0.058**

applications, which relies on very limited data to train and update classifiers.

C. Long Term Memory

Using a LTM to store validation data produces the results in Table IV, detailing values for both the batch (unlimited) and FIFO LTM strategies. Considering the batch LTM, results are comparable to the baseline experiment in Table II and differences (if any) are not significant. One disadvantage of using such a batch LTM is that memory usage linearly increases over time, which is not the case for the fixed size FIFO LTM, which keeps a constant size. However, results for the FIFO LTM are still comparable to the baseline experiment, with no significant AUC value changes. These results are likely related to the static testing environment used and further experiments on a test set that changes over time, as the training data does, should provide more conclusive results.

VII. CONCLUSION

In this paper an adaptive ensemble of classifiers is proposed to update and refine BMs in response to new reference samples. This update strategy is analyzed with video surveillance application in mind. When new data becomes available, the *learn-and-combine* approach generates a new 2-class PFAM classifier, which is combined to previously trained classifiers in the ROC space. In this binary classification problem, positive samples are pooled with randomly selected negative samples from the universal facial model and the cohort model, for design and update of biometric face models. Results indicate that the learn-and-combine approach outperforms other tested approaches in terms of both AUC accuracy and compression. The approaches were also tested with limited training data, a relevant issue in surveillance applications, to find out that performance is acceptable even when only two samples per class are used to create/update classification models.

Current results on LTM usage are not conclusive, and further experiments are required to verify its impact on a changing test environment that reflects the changes in the training and validation data sets. Also, the characterization of the system in changing scenarios is still an interesting area to address, including handling changes in class priors and densities in the feature space.

TABLE III
AVERAGE AUC WITH LIMITED TRAINING DATA.

Samples per class	Classifier	D_1	D_2	D_3	D_4	D_5
8	k-NN	0.8758±0.030	0.9163±0.023	0.9287±0.022	0.9342±0.019	0.9407±0.017
	PFAM$_{batch}$	**0.9376±0.024**	0.9448±0.024	0.9476±0.024	0.9543±0.019	0.9547±0.017
	PFAM$_{inc}$	**0.9376±0.024**	0.9356±0.024	0.9293±0.024	0.9285±0.023	0.9145±0.031
	Learn++	0.7533±0.094	0.6867±0.104	0.6444±0.104	0.6484±0.096	0.6361±0.098
	L&C	0.9362±0.029	**0.9566±0.020**	**0.9618±0.018**	**0.9692±0.015**	**0.9709±0.014**
2	k-NN	0.7302±0.039	0.8121±0.036	0.8479±0.033	0.8724±0.032	0.8834±0.030
	PFAM$_{batch}$	**0.8743±0.037**	0.9083±0.030	0.9245±0.025	0.9335±0.024	0.9383±0.020
	PFAM$_{inc}$	**0.8743±0.037**	0.8837±0.038	0.8968±0.033	0.8952±0.033	0.9006±0.032
	Learn++	0.5698±0.112	0.5336±0.118	0.5826±0.111	0.5805±0.105	0.5419±0.108
	L&C	0.8660±0.039	**0.9253±0.029**	**0.9312±0.027**	**0.9485±0.023**	**0.9525±0.022**

TABLE IV
AVERAGE AUC OF IL CLASSIFIERS USING DIFFERENT MODELS OF LTM.

LTM model	Classifier	D_1	D_2	D_3	D_4	D_5
Accumulative	PFAM$_{inc}$	0.9499±0.022	0.945±0.020	0.9447±0.020	0.9460±0.022	0.9474±0.022
	Learn++	0.8111±0.091	0.7772±0.095	0.7788±0.091	0.7666±0.092	0.7829±0.088
	L&C	**0.9523±0.024**	**0.9615±0.021**	**0.9682±0.016**	**0.9723±0.014**	**0.9732±0.014**
FIFO	PFAM$_{inc}$	0.9499±0.022	0.9452±0.020	0.9425±0.021	0.9403±0.024	0.9355±0.023
	Learn++	0.7887±0.095	0.7674±0.098	0.7686±0.095	0.7659±0.094	0.7645±0.091
	L&C	**0.9523±0.024**	**0.9615±0.021**	**0.9657±0.019**	**0.9711±0.014**	**0.9724±0.014**

ACKNOWLEDGMENT

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada, and the Defence Research and Development Canada Centre for Security Science Public Security Technical Program (PTSP 03-0401BIOM). This work was also supported by the Programa de Mejoramiento del Profesorado of the Secretaría de Educación Pública in México (PROMEP/103.5/094294).

REFERENCES

[1] A. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, June 2006.

[2] A. Rattani, "Adaptive biometric system based on template update procedures," Ph.D. dissertation, University of Cagliari, 2010.

[3] J. N. Pato and L. I. Millett, *Biometric Recognition: Challenges and Opportunities*, N. R. C. o. t. N. Whither Biometrics Committee, Ed., 2010.

[4] U. Uludag, A. Ross, and A. Jain, "Biometric template selection and update: a case study in fingerprints," *Pattern Recognition*, no. 37, pp. 1533–1542, 2004.

[5] F. Roli, L. Didaci, and G. L. Marcialis, *Adaptive biometric systems that can improve with use*. Springer, 2008, pp. 447–471.

[6] N. Poh, R. Wong, J. Kittler, and F. Roli, "Challenges and research directions for adaptive biometric recognition systems," vol. 5558 LNCS, Alghero, Italy, 2009, pp. 753–764.

[7] A. K. Jain and A. Ross, "Learning user-specific parameters in a multibietric system," *International Conference on Image Processing*, September 2002.

[8] E. Granger, J. Connolly, and R. Sabourin, "A comparison of fuzzy artmap and gaussian artmap neural networks for incremental learning," *Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA)*, 2008.

[9] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: An incremental learning algorithm for mlp networks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 31, no. 4, pp. 497–508, 2001.

[10] W. Khreich, E. Granger, A. Miri, and R. Sabourin, "Iterative boolean combination of classifiers in the roc space: An application to anomaly detection with hmms," *Pattern Recognition*, vol. 43, no. 8, pp. 2732 – 52, 2010.

[11] R. Goh, L. Liu, X. Liu, and T. Chen, "The cmu face in action (fia) database," Berlin, Germany, 2005, pp. 255 – 63.

[12] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, December 2003.

[13] F. Matta and J.-L. Dugelay, "Person recognition using facial video information: a state of the art," *Journal of Visual Languages and Computing*, vol. 20, no. 3, pp. 180–7, 2009.

[14] J. Connolly, E. Granger, and R. Sabourin, "An adaptive ensemble of fuzzy artmap neural networks for video-based face classification," *IEEE World Congress on Computational Intelligence*, 2010.

[15] F. Li and H. Wechsler, "Open set face recognition using transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1686 – 97, 2005.

[16] D. Tax and R. Duin, "Growing a multi-class classifier with a reject option," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1565 – 70, 2008.

[17] A. Brew and P. Cunningham, "Combining cohort and ubm models in open set speaker detection," vol. 48, no. 1, Van Godewijckstraat 30, Dordrecht, 3311 GZ, Netherlands, 2010, pp. 141 – 159.

[18] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "Artmap. supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, vol. 4, no. 5, pp. 565 – 588, 1991.

[19] B. Fritzke, "Growing self-organizing networks-why?" Brussels, Belgium, 1996, pp. 61 – 72.

[20] M. A. Maloof, "Incremental rule learning with partial instance memory for changing concepts," vol. 4, Portland, OR, United states, 2003, pp. 2764 – 2769.

[21] K. Okamoto, S. Ozawa, and S. Abe, "A fast incremental learning algorithm of rbf networks with long-term memory," vol. vol.1, Piscataway, NJ, USA, 2003, pp. 102 – 7.

[22] S. Ruping, "Incremental learning with support vector machines," *IEEE International Conference on Data Mining*, pp. 641–642, 2001.

[23] C. P. Lim and R. F. Harrison, "Probabilistic fuzzy artmap: An autonomous neural network architecture for bayesian probability estimation," *Artificial Neural Networks*, pp. 148–153, June 1995.

[24] A. Nickabadi, M. M. Ebadzadeh, and R. Safabakhsh, "Evaluating the performance of dnpso in dynamic environments," Singapore, Singapore, 2008, pp. 2640 – 2645.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971 – 87, 2002.