Accepted in Int. Conference on Imaging for Crime Detection and Prevention (ICDP), 2013

An Individual-Specific Strategy for Management of Reference Data in Adaptive Ensembles for Person Re-Identification

Miguel De-la-Torre*†, Eric Granger*, Robert Sabourin*, Dmitry O. Gorodnichy‡

*École de technologie supérieure, Université du Québec, Montréal, Canada,
miguel@livia.etsmtl.ca; eric.granger@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

Keywords: Multi-Classifier Systems; Adaptive Biometrics; Face Recognition; Video Surveillance; Person Re-Identification

Abstract

In video surveillance, person re-identification refers to recognizing individuals of interest from faces captured across a network of video cameras. Face recognition in such applications is challenging because faces are captured with limited spatial and temporal constraints. In addition, facial models for recognition are commonly designed using limited reference samples from faces captured under specific conditions. Given new reference samples, updating facial models may allow maintaining a high level of performance over time. Although adaptive ensembles have been successfully applied to robust modeling of an individual's face, reference data samples must be stored for validation. In this paper, a memory management strategy based on Kullback-Leiber (KL) divergence is proposed to rank and select the most relevant validation samples over time in adaptive individual-specific ensembles. When new reference data becomes available for an individual, updates to the corresponding ensembles are validated using a mixture of new and previously-stored samples. Only the samples with the highest KL divergence are preserved in memory for future adaptations. The strategy is compared with reference classifiers using videos from the FIA data set. Simulation results show that the proposed strategy tends to select samples of statistically different subjects (so-called "wolfs") for validation, thereby reducing the number of samples per individual by up to 80%, yet maintaining a high level of performance.

1 Introduction

In many video surveillance applications, automated face recognition (FR) is increasingly employed to alert a human operator to the presence of individuals of interest appearing in either live (real-time monitoring) or archived (post-event analysis) videos. FR in video surveillance (FRiVS) is employed in a range of applications that involve still-to-video FR (e.g., watchlist screening) and video-to-video FR (e.g., person re-identification). This paper deals with the problem of re-identifying individuals in video streams coming from surveillance cameras, which can be used for search and retrieval, face tagging, video summarization and other security-related applications.

Given an individual of interest, the operator of a human-centric decision support system for person re-identification captures reference facial trajectories[1] corresponding to an individual appearing in video feeds, and designs a facial model (*e.g.* templates or statistical representation) to be stored in a gallery. Facial models are typically designed a priori using high quality captures (reference trajectories) obtained under controlled conditions. Then, during operations, facial trajectories captured in live or archived video streams are compared against facial models of individuals enrolled to the system.

Person re-identification in video surveillance is typically performed across a network of surveillance cameras. Accurate and timely responses are required for face trajectories captured in potentially complex semi-constrained (e.g., inspection lane, portal and checkpoint entry) and unconstrained (e.g., cluttered free-flow scene at an airport or casino) environments. Automated systems require robust operation under a wide variety of conditions, and must be fast and scalable to several enrolments and input videos from several IP cameras.

Moreover, the unobtrusive capture of video sequences with target individuals provides only a limited amount of high quality reference samples to design facial models. Abundant non-target facial trajectories are regrouped in the cohort model (CM, non-target individuals enrolled to the system) and universal model (UM, non-target individuals from operational trajectories). These models provide a great source of information for designing discriminant face models, leading the need to select the most relevant samples that avoid biasing matchers towards the negative class [11]. Finally, changes in the physiognomy of individuals lead to changes in the classification environment, and facial models may not be representative of all operational conditions, thus, exhibiting poor performance. Updating facial models has been shown to improve or maintain a high level of performance over various operating conditions [1, 3].

This paper is focused on adaptive video-to-video FR using multi-classifier systems (MCSs). It is assumed that faces cap-

[1] A facial trajectory is defined as a set of facial ROIs (produced by face segmentation) that correspond to the same high quality track of an individual across consecutive frames.

tured within trajectories (obtained from post-analysis of video feeds) are used for supervised incremental learning of facial models. Although adaptive ensembles have been applied to face modeling [1, 3, 18], they require the storage of reference validation samples in a long term memory (LTM) to preserve accuracy. One challenge for practical implementation is bounding the growing number of reference samples collected over several updates. Bounding the size of LTMs raises the issue of selecting the most relevant samples to be preserved in memory to maintain performance [5]. The selection of the most relevant validation samples, as well as the size of individual-specific LTMs also depends on the specific target individual.

In this paper, a strategy is proposed to select the most representative validation samples for an individual to be stored in a fixed size LTM. It is assumed that an ensemble of 2-class classifiers or detectors per target individual (EoD, target vs. non-target) is used for face matching. When a new reference trajectory becomes available, its ROI samples are combined with non-target samples from the CM and UM selected using one sided selection (OSS) [11]. The corresponding EoD is updated and validated using a mixture of new and pre-stored samples in LTM. The least relevant samples are discarded. Among different relevance measures inspired by techniques in active learning, analysis on synthetic data shows that the Kullback-Leibler (KL) divergence is able to accurately rank samples in the overlapping area between target and non-target populations.

The strategy proposed to manage a LTM is evaluated on face trajectories collected in semi-constrained environments from the CMU-FIA database. Three capture sessions with three months separation are considered for experiments. In this test case, the adaptive MCS is composed of an ensemble of 2-class Probabilistic Fuzzy ARTMAP (PFAM) classifiers for each enrolled subject. Average performance is presented and Doddington zoo [13] analysis is employed to compare individual-specific parameters for LTM management. Using the menagerie terminology introduced in [12], this analysis allows to categorize subjects into 4 groups: sheep, goat, wolf and lamb-like individuals according to their performance.

2 Adaptive Face Recognition in Video

Assume that video streams are captured from one or more video cameras. During operations, FRiVS involves several processing steps. First, segmentation isolates the facial ROIs corresponding to faces appearing in each frame using, e.g., the Viola-Jones algorithm. In order to build face trajectories, a tracker (e.g., CAMSHIFT) simultaneously follows the face of individuals in scene and assigns a same ID to facial ROIs from the same individual. Then, feature extraction extracts and selects discriminant features for classification from the extracted ROIs and arranged into feature vectors. Common feature extraction-selection techniques include the Local Binary Pattern (LBP) algorithm and Principal Component Analysis (PCA). Input feature vectors are compared with facial models, producing matching scores that are compared to individual specific thresholds. In video surveillance applications, the system detects all matching identities where matching scores surpass thresholds. Finally, a decision fusion allows to combine tracking IDs with the output classifier predictions and accumulate responses over a face trajectory. This process allows for reliable spatio-temporal detection of persons of interest [15].

In literature, matching for FRiVS has been addressed as an open-set problem, where the number of individuals of interest is greatly outnumbered by non-target individuals. Multi-class classifiers have been used in video surveillance with a rejection threshold for unknown individuals. A multi-class classifier designed to address the open set problem in video face recognition is the TCM-kNN [12]. This matcher takes advantage of transductive inference to generate a class prediction based on randomness deficiency. Modular architectures with a detector (1- or 2-class classifier) per individual have been proposed, allowing to set individual-independent parameters [8]. An individual-specific approach is based on the identification of the decision region(s) in the feature space of individual specific faces, and training a dedicated feed forward neural network for each individual of interest [10]. Another example is an SVM-based modular system that was applied to an access control scenario [4]. To improve accuracy and reliability ensembles of 2-class classifiers or detectors (EoD) have been proposed to implement individual-specific detectors. EoDs are co-jointly trained using a dynamic particle swarm optimization (DPSO) based training strategy, generating a diversified pool of ARTMAP neural networks. Trained detectors are selected and combined using boolean combination (BC) [17].

Adaptive systems for FR in video have also been proposed in literature to maintain a high level of performance. These allow to update facial models over time through supervised incremental learning of new data. An incremental learning strategy based on DPSO has been proposed for video-based access control. It allows to evolve an ensemble heterogeneous multi-class classifiers from new data, using a LTM to store validation samples for fitness estimation and to stop training epochs. This approach reduces the effect of knowledge corruption [1]. Another adaptive MCS for FRiVS is composed of an ensemble of binary 2-class classifiers per individual, a DPSO module and a LTM. ARTMAP neural networks are used as ensemble members, and the combination function is updated using BC [3]. Learn++ is another well-known ensemble-based technique for incremental learning that has been applied to FR. It employs Adaboost to generate a new set of weak classifiers every time new data becomes available, and combines old and new classifiers using weighted majority voting [18].

To assure a high level of accuracy, adaptive MCSs require the storage of reference validation samples in a LTM. However, memory limitations imposed by real-world systems prevent the indefinite growth of the amount of stored validation samples. In literature, editing algorithms like the condensed nearest neighbor have been used to manage a gallery of templates in template matching systems, and bound the amount of reference samples stored in memory [5]. In this paper adaptive MCSs are considered for FRiVS, where an ensemble of 2-class classifiers is used to estimate the facial model of individuals of interest [3]. An individual-specific strategy is proposed to manage (rank and select) the most informative validation samples over time for each adaptive ensemble.

3 Selection of Representative Samples

Some methods in literature allow to select a subset of representative samples for validation, and the criteria for representativeness is related to the level of information provided for the specific system. Fig. 1 presents the levels of selection that are relevant for ensembles of binary 1- or 2-class classifiers.

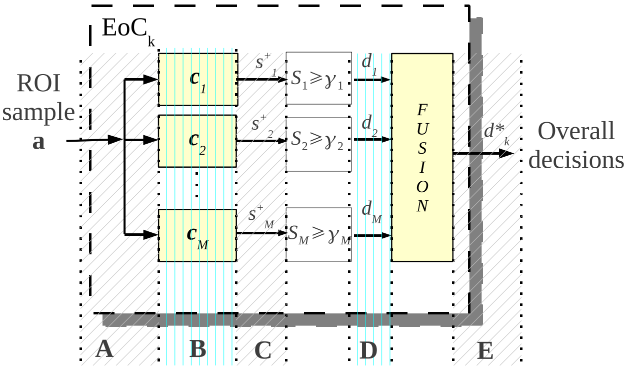


Figure 1. Levels of ranking that are relevant for an ensemble of detectors (1 or 2-class binary classifiers) for individual k.

At the *input data level* **(A)** the dataset itself is used to filter out redundant samples, information about data distributions of samples is not required. At the *classifier level* **(B)** the relevance measure of samples is retrieved from the internal response of the classifiers in the ensemble, to an input sample $\mathbf{a}$. At the *classifier score level* **(C)**, the output scores $S^+_m(\mathbf{a})$ of M classifiers in the ensemble may be combined to produce a measure of relevance. When probabilistic classifiers are used as base classifiers, the computation of relevance measures is based on the combined estimated posterior probability (classification scores S^+_m). At the *classifier decision level* **(D)**, the output predictions $d_m(\mathbf{a})$ of classifiers in the ensemble are combined. Voting strategies can be used to generate a relevance measure like vote entropy. Finally, at the *ensemble decision level* **(E)**, the global output of the ensemble can be used as a measure of the informativeness of the input sample.

Uninformed Selection. Unlike other levels, methods from level **A** do not require previously trained classifiers to provide information in the selection process. For instance, random under-sampling is the easiest non-heuristic method that randomly eliminates samples from the majority class. Other methods exploit the geometric relationship between samples in feature space, like the condensed nearest neighbor rule (CNN) and one sided selection (OSS) [7].

OSS is considered in this paper to select representative samples from the CM and UM. It aims to eliminate the samples from the majority (non-target) that are distant from the decision boundary in the original set D. It starts by building a training set D' with all target samples and one randomly selected non-target sample. Then, 1-NN is trained on D', and used to classify the remaining non-target samples. Misclassified non-target samples are incorporated to D', which at the end will constitute a consistent subset of D.

Informed Selection. Methods at levels **C** and **D** are independent of classification algorithm used in the ensemble as well as combination strategy, and allow to rank and select representative samples. The only constraint imposed by level **C** lies in the compatibility of scores produced by classifiers, a limitation that can be defeated by using normalization strategies.

A method that operates at level **C** is the *average margin sampling*. It is inspired on the *margin sampling* proposed by Scheffer *et al* in [19], and is defined as

$$AMS(\mathbf{a}) = \frac{1}{M}\sum_{m}^{M} MS_m(\mathbf{a}) \ , \quad (1)$$

where M is the number of ensemble members, and $MS_m(\mathbf{a})$ is the margin sampling estimated for each ensemble member c_m given the input sample $\mathbf{a}$. Margin sampling is computed by

$$MS(\mathbf{a}) = S(\omega_{max}, \mathbf{a}) - S(\omega_{2max}, \mathbf{a}) \ , \quad (2)$$

where $\omega_{max}, \omega_{2max}$ are the first and the second most probable class labels respectively, and $S(\omega)$ is the output score (*e.g.* posterior probability) of a given classifier for class ω. Margin sampling aims to incorporate the posterior probability of the second most likely class label to the relevance measurement.

The disagreement between base classifiers on a test sample $\mathbf{a}$ has also been used as a measure of relevance. For instance, the *Kullback-Leibler* (KL) divergence (or relative entropy), proposed by McCallum and Nigam, operates at level **C** [9]. The KL divergence is defined as

$$KL(\mathbf{a}) = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{i\in\Omega} S^i_m(\mathbf{a}) \log \frac{S^i_m(\mathbf{a})}{\hat{P}^i_{EoD_k}(\mathbf{a})}\right) \ , \quad (3)$$

where M is the number of classifiers in the ensemble, and $\hat{P}^i_{EoD_k}(\mathbf{a})$ given by Eqn. 4 is the consensus probability that the class $i \in \Omega$ is the correct label for sample $\mathbf{a}$, given the scores $S^i_n(\mathbf{a})$ produced by the base classifiers.

$$\hat{P}^i_{EoD_k}(\mathbf{a}) = \frac{1}{M}\sum_{n=1}^{M} S^i_n(\mathbf{a}) \ . \quad (4)$$

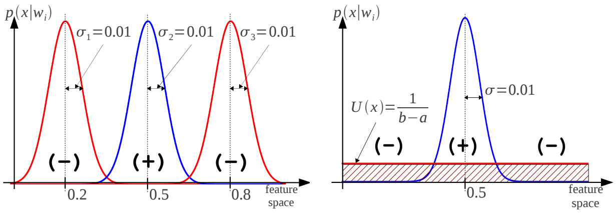
For KL divergence, the most informative samples are those with the largest average difference between the class distributions of any one of the committee members and the consensus.

An example of level **D** relevance measure is the *vote entropy* [2], defined as

$$VE(\mathbf{a}) = -\sum_{i\in\Omega} \frac{V(\omega_i, \mathbf{a})}{M} \log \frac{V(\omega_i, \mathbf{a})}{M} \ , \quad (5)$$

where $V(\omega_i, \mathbf{a})$ is the number of votes for the class $\omega_i \in \Omega$ provided by the ensemble. Similarly to KL divergence, VE increases with the disagreement in the ensemble members, but its resolution (*e.g.*, ranking levels) is bounded by the number of base classifiers in the ensemble.

Synthetic Analysis. For more insight on the selective capacity of the relevance measures, two synthetic 2-class problems were designed in the 1D space. Fig. 2 shows the original probability distributions of data. Central Gaussian distribution in Fig. 2a and 2b have a center of mass $\mu_2 = 0.5$. Centers of mass of the non-target distributions in Fig. 2a are $\mu_1 = 0.2$ and



(a) Problem 1 **(b)** Problem 2

Figure 2. Data distributions used to generate the training data for both problems. Central Gaussian distributions in both figures generate the positive (+) samples, and left and right distributions generate negative (-) samples.

$\mu_3 = 0.8$, and in Fig. 2b the non-target samples are randomly drawn according to a uniform distribution. All Gaussians have a variance of $\sigma = 0.01$.

An ensemble of 7 probabilistic Fuzzy ARTMAP (PFAM) classifiers was trained for both problems on balanced training sets. The PFAM classifier combines the Fuzzy ARTMAP learning to encode category prototypes and update centers of mass of estimated class distributions [14]. A DPSO learning strategy was used for base classifiers generation and hyperparameter optimization [1].

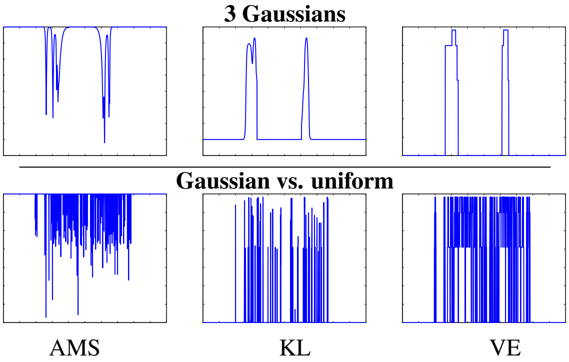


Figure 3. Value of relevance measures obtained over the feature space with an EoD (PFAM) for the 3 Gaussians (top) and Gaussian vs. uniform (bottom) problems.

The value of relevance measures produced by the ensembles are presented on Fig. 3. The three measures show a good characterization of the overlapping region between target and non-target populations, specially on the problem with three Gaussians. Vote Entropy shows a lower resolution than KL divergence and AMS, and the smoothness of the KL divergence curve shows a better representation of the overlapping area. In this paper, the KL divergence is employed to implement a strategy to assess the relevance of reference samples to manage a fixed size memory.

4 Individual-Specific Management of LTM

Fig. 4 presents the modular architecture for FRiVS that allows for supervised adaptation of facial models from new trajectories. During operations, the system will process ROIs segmented in each frame, and along input trajectories. ROI feature vectors are extracted and presented to each EoD$_k$. Using a face tracking algorithm, different faces in a video sequence are followed frame to frame and regrouped, and the successive predictions p_k from EoD$_k$ for each trajectory are accumulated over time for spatio-temporal recognition, in order to provide an overall prediction for each track ID. Finally, an individual specific threshold is applied to the accumulation curves of each EoD$_k$ in order to generate an overall decision d_k for each EoD$_k$. Note that there are several accumulation modules per track ID, to simultaneously recognize several people at a time in the scene.

During design/update, each EoD$_k$ performs independent supervised incremental learning. When a new trajectory T_k becomes available for a person k, OSS is used to form a consistent individual-specific training set D_k with all target samples and non-target samples selected from CM and UM. Then, a DPSO-based strategy is employed to generate a new pool of diversified binary classifiers that are combined with previously trained detectors corresponding to person k [3]. A fixed size LTM is maintained with validation samples that are representative of the overlapping zone between target and non-target distributions. The KL divergence measure (Eq. 3) is employed to rank reference samples and store the λ_k most representative in the LTM, where λ_k is the size of the LTM for person k enrolled to the system. At each adaptation step, new validation samples are combined with those stored in the LTM to accurately estimate a new fusion function and select an operations point.

Algorithm 1 shows the procedure followed by the management strategy to rank and select representative validation samples to be stored in the LTM$_k$. When a new validation set D with target and non-target samples becomes available for individual k, all samples are ranked according to the KL divergence. Then, the $\lambda_k/2$ highest ranked target samples, as well as the $\lambda_k/2$ highest ranked non-target samples are preserved, whereas the rest are discarded.

Algorithm 1: KL relevance subsampling for the EoD$_k$.

```
Input  : $D$, $S_k(a_i)$, $\lambda_k$    // Validation data, scores
                                         // and size of $LTM_k$
Output : $Dr$                            // Representative samples
for $a_i \in D$ do
    $r_i = KL(S_k(a_i))$                 // Rank with Eq. 3
$D \Leftarrow sort(D, r, d)$             // Sort $D$ according to $r_i$
$Dr^+ \Leftarrow first\_pos(D, \lceil \frac{\lambda_k}{2} \rceil)$
$Dr^- \Leftarrow first\_neg(D, \lceil \frac{\lambda_k}{2} \rceil)$
$Dr \Leftarrow Dr^+ \cup Dr^-$
```

The new set Dr is formed from old and new validation samples that are difficult to classify by old and new classifiers. Then, the selection is based on past and present information retrieved from the classifiers by choosing the samples in the overlapping area of the target and non-target distributions. Thus, the proposed selection strategy allows to store the samples that contain the most relevant information to define the decision frontier.

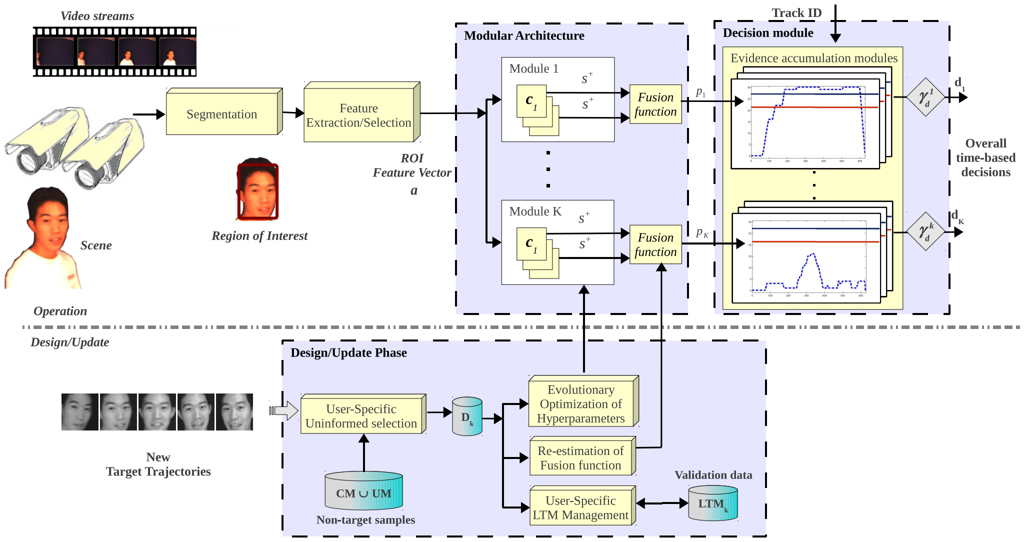


Figure 4. Adaptive MCS for FRiVS. In the design/update phase, when a new face trajectory T_k becomes available for a person k, a training set D_k is formed with all its target samples, and non-target samples selected from CM and UM using OSS. Then, an evolutionary optimization strategy is employed to generate a new pool of diversified classifiers with optimized hyper parameters, and the decision-level fusion function is updated based on new data and pre-stored reference samples (from the LTM). Finally the λ_k most relevant samples from previous and newly-learned trajectories are stored in LTM according to the KL divergence.

5 Experimental Methodology

The proposed LTM management strategy is characterized for person re-identification scenario, using the CMU Face in Action (FIA) database [6]. The FIA database consists of 20 second videos of face data from 180 participants mimicking a passport checking scenario. Faces are captured at 30 frames per second, with a resolution of 640 × 480 pixels. An array of 6 cameras horizontally positioned at the face level capture the scene. Pairs of cameras were positioned at 0^o (frontal) and $\pm 72^o$ (left and right) angle with respect to the individual. Three cameras were set to an 8-mm focal-length (zoomed), resulting in face areas around 300 × 300 pixels, and the other three to a 4-mm focal-length (unzoomed) resulting in face areas around 100 × 100 pixels. The cameras utilize the Sony ICX424 sensor, with a maximum resolution of 640x480 pixels and a 6mm diagonal image size. Data has been captured on three sessions separated by a three months interval for each individual.

Facial trajectories are formed with frontal facial regions segmented using the Viola-Jones algorithm [20], and an ideal face tracker is assumed. All images are scaled to the resolution of the smallest face obtained after face detection (70x70 pixels). The Multi Scale LBP [16] feature extractor has been used with three different block sizes (3 × 3, 5 × 5 and 9 × 9), along with pixel intensities features. Resulting features are combined into feature vectors, and PCA is applied to select the 32 most discriminant projected features.

Ten individuals were randomly selected for re-identification (with FIA ID 2, 58, 72, 92, 147, 151, 176, 188, 190 and 209), and one EoD$_k$ is designed for each. 88 of the remaining individuals are selected as part of the universal model (UM), and the rest are considered as never seen test individuals. It is important to highlight that individuals from the UM never appear in test. Face trajectories from individuals of interest contain between 80 and 239 facial regions, and non-target training and test samples differ in each dataset.

Prior to computer simulations, four data subsets have been prepared. Trajectories in the design dataset D are comprised of target ROIs from the the zoomed view of capture session 1. The test/adaptation datasets D_1 to D_3 have been constructed with ROIs from the unzoomed view of capture sessions 1 to 3 respectively. Non-target samples are independently selected for each of the training/validation sets picked from the cohort model (CM) and UM, using OSS [11]. The CM comprises trajectories from non-target individuals enrolled to the system.

The classifiers were initially trained using trajectories in the design set D, and tested on trajectories in D_1 for the first evaluation. For an evaluation in a gradually changing environment, after performance evaluation on D_1 the classifiers were updated with trajectories in D_1 and tested on D_2. The same process was repeated for update/test on D_2 and D_3 respectively. The proposed approach was updated with only the new labeled dataset. In contrast, TCM-kNN was trained on batch mode, learning from scratch the previous and new samples.

The MCS used for LTM analysis was composed of an ensemble of 2-class Probabilistic Fuzzy ARTMAP (PFAM) classifiers per individual, EoD$_k$ (PFAM). The DPSO learning strategy was used for classifiers generation and hyperparameters optimization, and BC was applied for decision level fusion of classifiers on the ROC space [3]. The LTM was managed according to the KL divergence with six individual-specific values of λ_k were explored: 0, 25, 50, 75, 100 and ∞.

Evaluation was performed following 2 × 5-fold cross-validation for 10 independent trials. Target samples from the learning set were randomly split according to a uniform dis-

fpr (%) ↓	tpr (%) ↑	F_1 ↑	pAUC (5%) ↑
TCM-kNN (batch learning)			
20.13 ±0.419 → 22.81 ±0.414 → 18.32 ±0.187	**90.65 ±1.425** → **54.26 ±3.220** → **87.91 ±1.666**	0.0935 ±0.00339 → 0.0580 ±0.00391 → 0.1747 ±0.00442	88.71 ±1.47 → 48.54 ±3.34 → 83.16 ±2.29
PFAM			
0.95 ±0.184 → 1.20 ±0.122 → 1.91 ±0.235	80.84 ±2.048 → 54.06 ±3.465 → 84.52 ±2.315	0.6648 ±0.01930 → **0.4377 ±0.02880** → 0.6656 ±0.02432	90.40 ±1.21 → 69.18 ±2.86 → 87.75 ±1.66
Learn++			
0.60 ±0.068 → **0.57 ±0.038** → 1.19 ±0.108	16.90 ±2.365 → 11.87 ±1.804 → 20.57 ±2.780	0.1613 ±0.01669 → 0.1278 ±0.01368 → 0.1917 ±0.01953	47.87 ±2.71 → 36.81 ±2.45 → 34.19 ±2.64
EoD$_k$ (PFAM) $LTM_{KL,\lambda_k=\infty}$			
0.62 ±0.09 → 0.67 ±0.05 → **0.84 ±0.07**	77.02 ±2.10 → 45.51 ±3.63 → 76.70 ±2.71	**0.6789 ±0.0177** → 0.4041 ±0.0308 → **0.6909 ±0.0231**	**92.88 ±0.81** → **72.03 ±2.76** → **93.64 ±0.84**

Table 1. Average performance of the system on 10 individuals and 10 trials, for $D_1 \to D_2 \to D_3$. Operations point at $fpr = 1\%$.

tribution, in 5 folds of the same size. The folds were first distributed in three different design sets, including two folds for training (D_t^t), $1\frac{1}{2}$ folds to stop training epochs (D_t^e), and $1\frac{1}{2}$ folds for fitness evaluation (D_t^f). Once the classifiers were trained, D_t^e and D_t^f are combined, randomized and divided in two equally distributed subsets to produce a validation data for threshold/fusion function estimation (D_t^c), and to select the operations point (D_t^s). Each fold was assigned to a different training/validation set at each replica of the experiment. At replication 5, the five folds were regenerated after a randomization of the sample order for each class, and the process was repeated to generate a standard error on ten different assignments.

Reference approaches in comparison include TCM-kNN, single PFAM in incremental learning mode and Learn++ with 7 PFAM base classifiers. TCM-kNN was trained with a fixed $k = 1$ on a batch learning scheme. PFAM classifiers used in all other approaches, were trained using DPSO based learning strategy to optimize hyperparameters. Validating the number of training epochs for classifier convergence was performed on D_t^e, whereas particle fitness was evaluated on D_t^f. The DPSO algorithm was initialized with a swarm of 60 particles, and a maximum of 5 particles within each of the 6 subswarms. The algorithm was set to run a maximum of 30 iterations, allowing 5 extra iterations to ensure convergence. Once the global best particle is found, its classifier and the 6 local bests from each subswarm were added to the EoD.

6 Simulation Results

Table 1 presents the average performance obtained after incremental learning of data blocks D, D_1 and D_2 (test on D_1, D_2 and D_3 respectively) for the 10 individuals of interest. Results are compared in the ROC space with the partial AUC for $0 \leq fpr \leq 0.05$ (pAUC (5%)), and at a specific operations point selected on the validation ROC curve for an $fpr = 1\%$. In that table, the EoDs (PFAM) with a LTM managed with KL divergence (LTM$_{kl,\lambda_k}$) generally provides a higher level of performance in terms of pAUC (5%) w.r.t. reference systems.

By analyzing the performance at ($fpr = 1\%$), it can be noticed that EoDs allow for a lower fpr versus incremental PFAM. Conversely, TCM-kNN yields the highest fpr due to the difficulty faced by multi-class classifiers in finding multiple boundaries during the same optimization process: within the cohort, and between individuals of interest. In contrast, TCM-kNN shows the highest tpr, followed by the PFAM, even though the operating point is selected using the same validation data for all approaches. The PFAM can provide good generalization for the target class, but face difficulties establishing the limit to non-target samples. Ensemble based classifiers provide accurate rejection of non-target samples.

Table 2 presents the performance of the ensemble during incremental learning for two individuals, using $\lambda_k = 25, 75$ and 100. EoD$_{58}$ (for individual with ID 58) was selected because of its good initial performance (pAUC (5%) $\geq$ 95%). This individual is easy to detect by the system ($tpr > 80\%$), and easy to differentiate against non-target individuals ($fpr < 1\%$) – it is a *sheep*-like subject in the Doddington zoo taxonomy [12]. Conversely, EoD$_{188}$ was selected because of its low initial performance (pAUC (5%) < 95%). It corresponds to an individual that even though is easy to detect by the system ($tpr > 80\%$), it is also easy to impersonate ($fpr > 1\%$). For this EoD$_{188}$, the test on D_1 throws 32 non-target individuals that are wrongly detected more than 1% of the time (*wolves*). Given the number of wolves, the EoD$_{188}$ corresponds to a *lamb*-like individual.

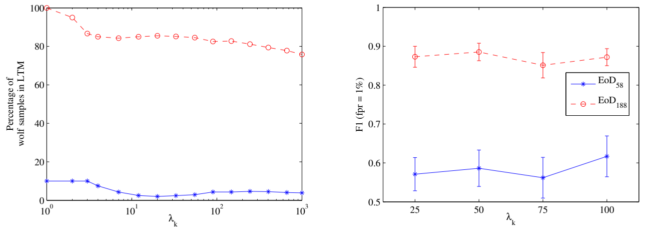
	EoD$_{58}$	**EoD**$_{188}$
LTM$_{KL,\lambda_k=25}$		
fpr	0.23 ±0.09 → 0.87 ±0.07 → 3.92 ±0.71	2.54 ±0.57 → 1.01 ±0.10 → **0.84 ±0.24**
tpr	84.43 ±3.33 → 39.49 ±7.01 → 90.93 ±3.02	89.58 ±4.26 → 84.88 ±5.36 → 97.29 ±0.82
F_1	84.92 ±2.29 → 40.29 ±6.07 → 57.10 ±4.27	47.20 ±5.37 → 65.94 ±3.81 → **87.30 ±2.70**
pAUC (5%)	98.45 ±0.23 → 72.46 ±3.74 → 97.18 ±1.09	91.12 ±2.41 → 96.43 ±0.80 → 99.64 ±0.07
LTM$_{KL,\lambda_k=75}$		
fpr	0.23 ±0.09 → 0.84 ±0.10 → 4.29 ±0.62	2.54 ±0.57 → 1.02 ±0.10 → 1.07 ±0.31
tpr	84.43 ±3.33 → 41.49 ±7.76 → 94.65 ±3.25	89.58 ±4.26 → 89.53 ±3.21 → **97.60 ±0.64**
F_1	84.92 ±2.29 → 41.71 ±6.41 → 56.19 ±5.23	47.20 ±5.37 → 68.38 ±2.57 → 85.11 ±3.27
pAUC (5%)	98.45 ±0.23 → 71.92 ±3.50 → **98.60 ±0.77**	91.12 ±2.41 → 96.21 ±0.67 → 99.63 ±0.09
LTM$_{KL,\lambda_k=100}$		
fpr	0.23 ±0.09 → 0.84 ±0.08 → **3.64 ±0.73**	2.54 ±0.57 → 1.09 ±0.14 → **0.84 ±0.19**
tpr	84.43 ±3.33 → 38.28 ±8.46 → **95.81 ±1.63**	89.58 ±4.26 → 88.08 ±3.06 → **97.60 ±0.52**
F_1	84.92 ±2.29 → 38.08 ±7.05 → **61.68 ±5.25**	47.20 ±5.37 → 66.69 ±3.20 → 87.20 ±2.19
pAUC (5%)	98.45 ±0.23 → 71.91 ±3.56 → 98.36 ±0.79	91.12 ±2.41 → 96.25 ±0.55 → **99.67 ±0.09**

Table 2. Average performance of the EoD$_{58}$ and EoD$_{188}$ after tests on $D_1 \to D_2 \to D_3$.

Regarding the F_1 measure for EoD$_{58}$ after test on D_2, results show a performance that declines more importantly for EoD$_{58}$ with $\lambda_{58} = 100$. In this case, with $\lambda_{58} = 75$ it allows for a better performance. On the other hand, after update on D_2 (test on D_3) the appearance of new representative samples in the LTM leads to a recovery in the performance. Performance shown by EoD$_{58}$ after testing on D_3 suggests that sheep-like individuals benefit from high λ_k values.

A different trend is shown by EoD$_{188}$, which in general increases in performance every time it is updated. A comparison between λ_{188} values shows that there is no significant difference between using a large or small LTM, indicating that the performance of the EoD$_{188}$ for this lamb-like individual is maintained using this KL-based selection, even with small λ_{188} values (e.g. $\lambda_{188} = 25$). Note that the average number of samples selected by OSS for validation in experiments is 139.1 ±5.07 (global average for the 10 individuals over the 10 trials), and $\lambda_{188} = 25$ samples constitutes the 17.97% of the data.

Samples from wolf-like individuals degrade the fpr of EoDs for lamb-like individuals, and are useful for system's validation, allowing for better discrimination. Fig. 5 shows the average percentage of samples from wolf-like individuals selected by the KL algorithm for the EoD$_{58}$ and EoD$_{188}$, using a λ_k that grows up to 1000 samples from D_1. It can be seen that the percentage of samples from wolf-like individuals remains close to 80% for EoD$_{188}$. These proportions indicate effectiveness of the KL selection to retain wolf-like samples in the LTM of lamb-like individuals. Conversely, it is less than 10% for the EoD$_{58}$, suggesting that the most informative non-target samples are not associated with wolf-like individuals.



(a) Wolf samples in LTM (b) F$_1$ after test on D_3

Figure 5. Percentage of samples from wolf-like individuals and F_1 performance after test on D_3 for EoD$_{188}$ and EoD$_{58}$.

Finally, when a new trajectory for an individual of interest becomes available, it takes around 150 min. to update its facial model[2], and the modular architecture allows for parallel update of multiple facial models. This makes the system appropriate for off-line update from, e.g., daily police reports.

7 Conclusion

In this paper, an individual-specific strategy was proposed for management of reference samples used for validation of adaptive ensembles applied to person re-identification. When new reference samples becomes available for an individual enrolled to the system, its facial regions are combined with non-target samples from the universal and cohort models selected with OSS. Old and new validation samples are combined and ranked using Kullback-Leibler divergence, and the highest ranked are stored in a LTM for future validations. Its theoretical foundation lies on the relative entropy, for which the disagreement between ensemble members is an indicator of the informativeness of reference samples.

This strategy was tested on real-world CMU-FIA video data, and simulation results indicate that using the proposed strategy allows ensembles to maintain a level of performance comparable to that achieved by an ensemble where all validation samples are preserved, yet storing less than 20% of this samples. Comparing different LTM sizes (λ_k) for individual-specific ensembles suggests that sheep-like individuals benefit from high λ_k values, whereas low λ_k values may be selected for lamb-like individuals. This is related to the capacity of the management strategy to select samples from wolf-like individuals. Future research includes investigating strategies to find the optimal amount of samples required for each EoD, affecting a trade-off between performance and resources.

[2] Algorithm implemented in Matlab® R2010B, running on Linux Gentoo, on a 2.53GHz Intel® Xeon® processor.

References

[1] J.F. Connolly, E. Granger, and R. Sabourin. Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition. *PR*, 45:2460–2477, 2012.

[2] I. Dagan and S.P. Engelson. Committee-based sampling for training probabilistic classifiers. *ICML*, 150–7, San Francisco, USA, July 1995.

[3] M. De-la Torre, E. Granger, P. V. W. Radtke, R. Sabourin, and D. O. Gorodnichy. Incremental update of biometric models in face-based video surveillance. *IJCNN*, 1–8, Brisbane, Australia, June 2012.

[4] H. K. Ekenel, L. Szasz-Toth, and R. Stiefelhagen. Open-set face recognition-based visitor interface system. *CVS*, 5815, 43–52, Liege, Belgium, October 2009.

[5] B. Freni, G. Marcialis, and F. Roli. Template selection by editing algorithms: A case study in face recognition. *IAPR*, 5342, 745–754, Orlando, USA, December 2008.

[6] R. Goh, L. Liu, X. Liu, and T. Chen. The CMU Face In Action Database. *AMFG*, 255–263. CMU, 2005.

[7] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou. On the class imbalance problem. *ICNC*, 4, 192–201, Piscaatway, USA, August 2008.

[8] A.K. Jain and A. Ross. Learning user-specific parameters in a multibiometric system. *ICIP*, 57–60, Rochester, USA, September 2002.

[9] A. Kachites McCallum and K. Nigam. Employing EM and pool-based active learning for text classification. *ICML*, 350–8, San Francisco, USA, July 1998.

[10] B. Kamgar-Parsi, W. Lawson, and B. Kamgar-Parsi. Toward development of a face recognition system for watchlist surveillance. *Trans. on PAMI*, 33(10):1925–37, 2011.

[11] M. Kubat and S. Matwin. Addressing the curse of imbalanced training sets: One-sided selection. *ICML*, 179–186. Nashville, USA, July 1997.

[12] F. Li and H. Wechsler. Open set face recognition using transduction. *Trans. on PAMI*, 27(11):1686–97, 2005.

[13] G. Doddington, W. Liggett, A. Martin, M. Przybocki and D. Reynolds. Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance. *ICSLP*, 1351-1354. Sidney, Australia, December 1998.

[14] C. P. Lim and R. F. Harrison. An incremental adaptive network for on-line supervised learning and probability estimation. *Neural Networks*, 925–939. Houston, USA, June 1997.

[15] F. Matta and J.-L. Dugelay. Person recognition using facial video information: a state of the art. *JVLC*, 20(3):180–7, 2009.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Trans. on PAMI*, 24(7):971–87, 2002.

[17] C. Pagano, E. Granger, R. Sabourin, and D.O. Gorodnichy. Detector ensembles for face recognition in video surveillance. *IJCNN*, 1–8, Brisbane, Australia, June 2012.

[18] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar. Learn++: An Incremental Learning Algorithm for MLP Networks. *SMC*, 31(4):497–508, 2001.

[19] T. Scheffer, C. Decomain, and S. Wrobel. Active Hidden Markov models for information extraction. *AIDA*, 2189, 309–18, Berlin, Germany, June 2001.

[20] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 2(57):137–154, 2004.