IET Computer Vision

Research Article

IET Journals
The Institution of Engineering and Technology

Individual-specific management of reference data in adaptive ensembles for face re-identification

ISSN 1751-9632
Received on 22nd October 2014
Revised on 23rd January 2015
Accepted on 25th February 2015
doi: 10.1049/iet-cvi.2014.0375
www.ietdl.org

Miguel De-la-Torre[1] ✉, *Eric Granger*[1], *Robert Sabourin*[1], *Dmitry O. Gorodnichy*[2]

[1]*École de Technologie Supérieure, Université du Québec, Montréal, Canada*
[2]*Science and Engineering Directorate, Canada Border Services Agency, Ottawa, Canada*
✉ *E-mail: mikefixer@hotmail.com*

Abstract: During video surveillance, face re-identification allows recognition and targeting of individuals of interest from faces captured across a network of video cameras. In such applications, face recognition is challenging because faces are captured under limited spatial and temporal constraints. In addition, facial models for recognition are commonly designed using a limited number of representative reference samples from faces captured under specific conditions, regrouped into facial trajectories. Given new reference samples (provided by an operator or through some self-updating process), updating facial models may allow maintenance of a high level of performance over time. Although adaptive ensembles have been successfully applied to robust modelling of an individual's facial appearance, reference data samples from a trajectory must be stored for validation. In this study, 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 samples become available for an individual, updates to the corresponding ensemble 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. This strategy is compared with reference classifiers using videos from the face in action data. Simulation results show that the proposed strategy tends to select discriminative samples from wolf-like individuals for validation. It allows maintenance of a high level of performance, while reducing the number of samples per individual by up to 80%.

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 focuses on the problem of re-identifying individuals from faces captured using video surveillance cameras, as found in search and retrieval, face tagging, video summarisation and other security-related applications.

Using a decision support system for person re-identification, the operator seeks to capture reference facial trajectories corresponding to a target individual of interest appearing in video feeds, and designs a facial model (e.g. templates or statistical representation) to be stored in a gallery. A facial trajectory is defined as a set of facial captures (regions of interest (ROIs) produced by face segmentation) that correspond to the same high quality track of a same individual across consecutive frames. 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 into the system.

Face re-identification in video surveillance is typically performed across a network of surveillance cameras. Accurate and timely responses are required for FR from 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.

The unobtrusive capture of video sequences with target individuals provides only a limited amount of high quality reference samples to design facial models. Indeed, faces captured in video surveillance incorporate variations because of pose, illumination, blur, restoration, expression and so on. Updating facial models with new reference target trajectories has been shown to improve or maintain a high level of performance over various capture conditions [1–3]. 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 source of information for designing discriminant face models, leading to the need to select the most relevant samples that avoid biasing matchers towards the negative class [4].

This paper is focused on adaptive video-to-video FR using multi-classifier systems (MCSs). It is assumed that faces captured within trajectories (obtained from post-analysis of video feeds) are used to update facial models. Although adaptive ensembles have previously been applied to face modelling [1, 2, 5], 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 [6]. 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 two-class

classifiers or detectors per target individual (EoD, target against non-target) is used for face matching. When a new reference trajectory becomes available, its target samples extracted from captured ROIs are combined with non-target samples from the CM and UM selected using one-sided selection (OSS) [4]. The corresponding EoD is updated and validated using a mixture of new and pre-stored samples in LTM. Among different relevance measures inspired by techniques in active learning, the Kullback–Leibler (KL) divergence is proposed to accurately rank samples in the overlapping area between target and non-target populations. The least relevant samples are discarded.

The strategy proposed to manage an LTM is evaluated on face trajectories collected in semi-constrained environments from the CMU-FIA database [7]. Three capture sessions with three months' separation are considered for experiments on a scenario with gradual changes, whereas a single capture session with frontal, right and left capture views are considered for a scenario with abrupt changes. For validation, the adaptive MCS is composed of an ensemble of 2-class ARTMAP classifiers for each enrolled individual. Average performance is presented and Doddington zoo [8] analysis is employed to compare individual-specific parameters for LTM management. Using the menagerie terminology introduced in [9], this analysis allows to be categorised the subjects into four groups of individuals (sheep, goat, wolf and lamb) according to their performance.

2 Adaptive FR 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, for example, the Viola–Jones algorithm [10]. 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 to 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 tracking IDs to be combined with the output classifier predictions and accumulate responses over a face trajectory. This process allows for reliable spatio-temporal detection of persons of interest [11].

In the 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 FR is the TCM-kNN [9]. 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 setting of individual-independent parameters [12]. 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 [13]. Another example is an SVM-based modular system that was applied to an access control scenario [14]. 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 optimisation (DPSO)-based training strategy, generating a diversified pool of ARTMAP neural networks. Trained detectors are selected and combined using Boolean combination (BC) [15].

Adaptive systems for FR in the video have also been proposed in the literature to maintain a high level of performance. This means that facial models can be updated over time through supervised incremental learning of new data. An incremental learning strategy based on DPSO has been proposed for video-based access control. An ensemble heterogeneous multi-class classifiers can evolve from new data, using an 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 an LTM. ARTMAP neural networks are used as ensemble members, and the combination function is updated using BC [2]. Learn++ is another well-known ensemble-based technique for incremental learning that has been applied to the 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 [5].

To assure a high level of accuracy, adaptive MCSs require the storage of reference validation samples in an LTM. However, memory limitations imposed by real-world systems prevent the indefinite growth of the amount of stored validation samples. In the literature, editing algorithms like the condensed nearest neighbour have been used to manage a gallery of templates in template matching systems, and bound the amount of reference samples stored in memory [6]. 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 [2, 16]. 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 the literature allow selection of 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.

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 $\boldsymbol{a}$. At the 'classifier score level' (C), the output scores $S_m^+(\boldsymbol{a})$ 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

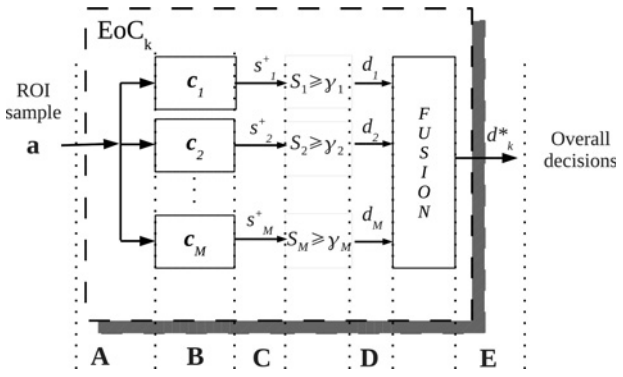

Fig. 1 *Levels of ranking that are relevant for an ensemble of detectors (1- or 2-class binary classifiers) for individual k*

A sample ROI $\boldsymbol{a}$ in the feature space feeds the EoC for a given individual k. Each classifier c_m produces a positive score s_m^+, that is compared to a decision threshold γ_m that produces a prediction d_m. All decisions are then combined with the fusion function to produce a single overall decision d_k^*

'classifier decision level' (D), the output predictions $d_m(\boldsymbol{a})$ 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.

3.1 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 neighbour (CNN) rule and OSS [17].

OSS is considered in this paper to select representative samples from the CM and the 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.

3.2 Informed selection

Methods at levels C and D are independent of classification algorithm used in the ensemble as well as combination strategy, and can 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 normalisation 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* [18], and is defined as

$$\mathrm{AMS}(\boldsymbol{a}) = \frac{1}{M}\sum_{m}^{M} \mathrm{MS}_m(\boldsymbol{a}) \tag{1}$$

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

$$\mathrm{MS}(\boldsymbol{a}) = S(\omega_{\max}, \boldsymbol{a}) - S(\omega_{2\max}, \boldsymbol{a}) \tag{2}$$

where $\omega_{\max}$, $\omega_{2\max}$ 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 $\boldsymbol{a}$ has also been used as a measure of relevance. For instance, the KL divergence (or relative entropy), proposed by McCallum and Nigam, operates at level C [19]. The KL divergence is defined as

$$\mathrm{KL}(\boldsymbol{a}) = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{i\in\Omega} S_m^i(\boldsymbol{a}) \log \frac{S_m^i(\boldsymbol{a})}{\hat{P}^i_{\mathrm{EoD}_k}(\boldsymbol{a})}\right) \tag{3}$$

where M is the number of classifiers in the ensemble, and $\hat{P}^i_{\mathrm{EoD}_k}(\boldsymbol{a})$ given by (4) is the consensus probability that the class $i \in \Omega$ is the correct label for sample $\boldsymbol{a}$, given the scores $S_n^i(\boldsymbol{a})$ produced by the base classifiers

$$\hat{P}^i_{\mathrm{EoD}_k}(\boldsymbol{a}) = \frac{1}{M}\sum_{n=1}^{M} S_n^i(\boldsymbol{a}) \tag{4}$$

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' [20], defined as

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

where $V(\omega_i, \boldsymbol{a})$ 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.

3.3 Synthetic analysis

For more insight on the selective capacity of the relevance measures, two synthetic 2-class problems were designed in the one-dimensional (1D) space. Fig. 2 shows the original probability distributions of data. Central Gaussian distribution in Figs. 2*a* and *b* have a centre of mass $\mu_2 = 0.5$. Centres of mass of the non-target distributions in Fig. 2*a* are $\mu_1 = 0.2$ and $\mu_3 = 0.8$, and in Fig. 2*b* the non-target samples are randomly drawn according to a uniform distribution. All Gaussians have a variance of $\sigma = 0.01$.

A pool of seven probabilistic fuzzy ARTMAP (PFAM) classifiers was trained for each problem using balanced data. The PFAM classifier combines the fuzzy ARTMAP learning to encode category prototypes and update centres of mass of estimated class distributions [21]. A DPSO learning strategy was used for base classifiers generation and hyperparameter optimisation [1].

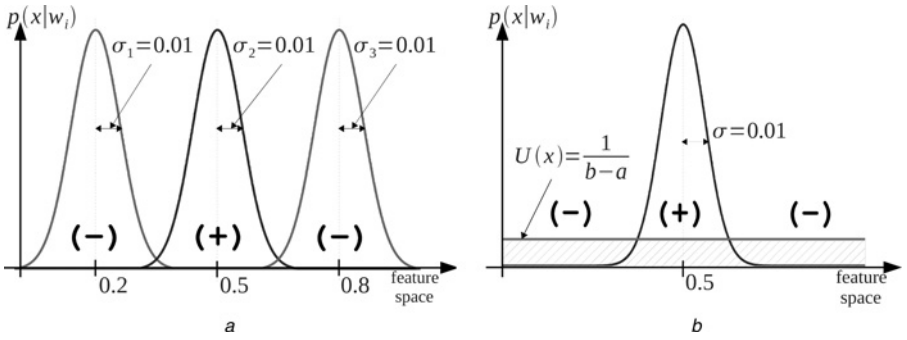

Fig. 2 *Data distributions used to generate the training data for both problems*

Central Gaussian distributions in both figures generate the positive (+) samples, and left and right distributions generate the negative (−) samples
a Problem 1
b Problem 2

IET Comput. Vis., 2015, Vol. 9, Iss. 5, pp. 732–740

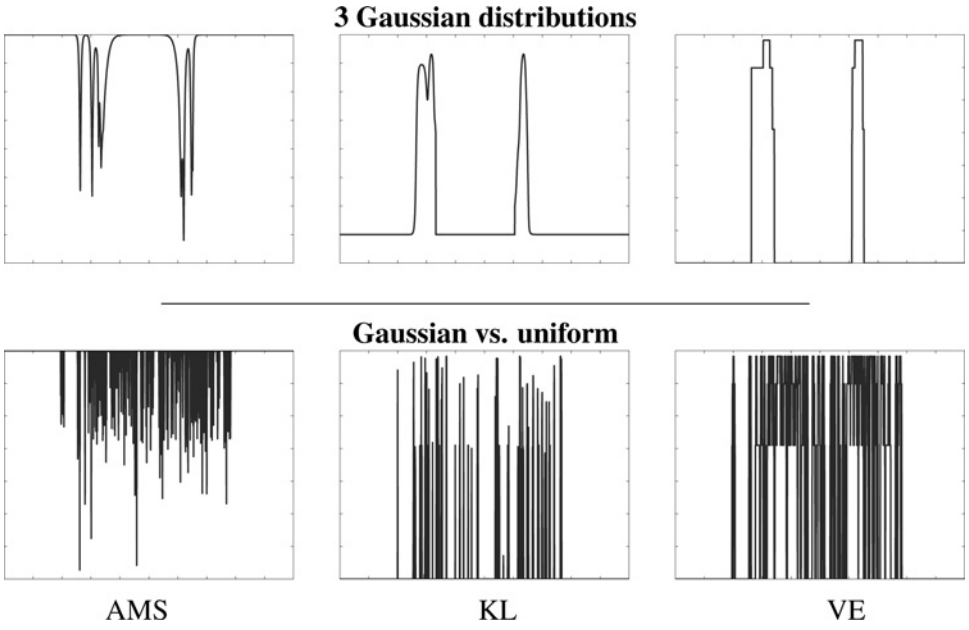

Fig. 3 *Value of relevance measures obtained over the feature space with an EoD (PFAM) for the three Gaussians (top) and Gaussian against uniform (bottom) problems*

The value of relevance measures produced by the ensembles is presented in Fig. 3. The three measures show a good characterisation of the overlapping region between target and non-target populations, especially 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 the ROI patterns extracted from 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 recognise 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 [2]. 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 (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 (see Fig. 5) 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.

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 can store the samples that contain the most relevant information to define the decision frontier.

5 Experimental methodology

The CMU face in action (FIA) database [7] is employed to characterise the proposed strategy in a person re-identification scenario that presents gradual and abrupt changes. The FIA database consists of 20 s videos of face data from 180 participants mimicking a passport checking scenario. An array of six cameras horizontally positioned at the face level capture the scene at 30 fps. Pairs of cameras were positioned at 0° (frontal) and ± 72° (left and right) angles 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 utilise the Sony ICX424 sensor, with a maximum resolution of 640×480 pixels and a 6 mm diagonal image size. Data has been captured on three sessions separated by a three months interval for each individual.

Facial trajectories were formed with facial regions segmented using the Viola–Jones algorithm [10] (see Fig. 6). An ideal face tracker is assumed, and all images were scaled to the resolution of the smallest face obtained after face detection (70×70 pixels). The multi-scale LBP [22] feature extractor has been used with three different block sizes (3×3, 5×5 and 9×9), along with pixel intensities features. Resulting features were combined into feature

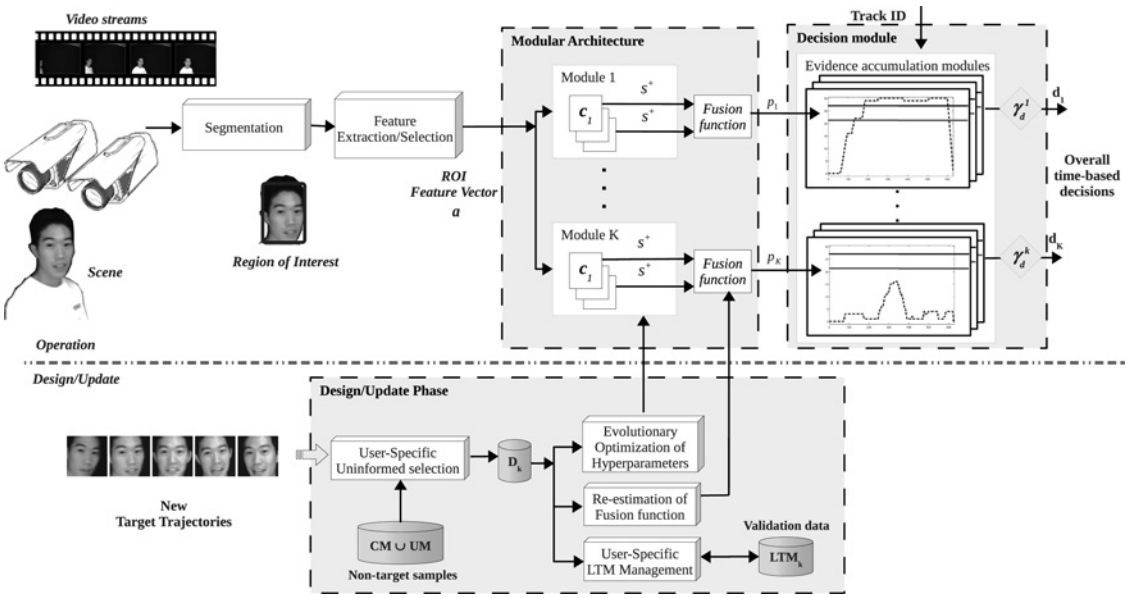

Fig. 4 *Adaptive MCS for FRiVS*

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 optimisation strategy is employed to generate a new pool of diversified classifiers with optimised 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

Algorithm 1:

Input : $D, S_k(a_i), \lambda_k$ // Validation data, scores and size of LTM_k

Output : Dr // Representative samples

1 **for** $a_i \in D$ **do**

2 $\quad r_i = KL(S_k(a_i))$ // Rank with Eq. 3

3 $D \Leftarrow sort(D, r, d)$ // Sort D according to r_i

4 $Dr^+ \Leftarrow first\_pos(D, \lceil \frac{\lambda_k}{2} \rceil)$

5 $Dr^- \Leftarrow first\_neg(D, \lceil \frac{\lambda_k}{2} \rceil)$

6 $Dr \Leftarrow Dr^+ \cup Dr^-$

Fig. 5 *Algorithm 1 KL relevance subsampling for the EoD$_k$*

vectors, and PCA was applied to select the 32 most discriminant projected features.

Ten individuals were randomly selected for re-identification, and one EoD$_k$ was designed for each. 88 of the remaining individuals are selected as part of the UM, and the rest are considered as never seen test individuals. The CM comprises trajectories from non-target individuals enrolled to the system. 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, five data subsets have been prepared. Trajectories in the design dataset D are comprised of target ROI patterns from the zoomed view of capture session 1. In order to build a scenario with gradual changes (age), the test/adaptation datasets D_1–D_3 have been constructed with ROI patterns from the unzoomed view of capture sessions 1–3, respectively. On the other hand, for the scenario with abrupt changes (pose), the test/adaptation datasets D_F, D_R and D_L have been constructed with ROI patterns from the unzoomed view of capture session 1, with the frontal, right and left cameras, respectively. Non-target samples are independently selected for each of the training/validation sets picked from the CM and UM, using OSS [4].

The classifiers were initially trained using trajectories in the design set D, and tested on trajectories in D_1 (or equivalently D_F for the scenario with abrupt changes), obtaining the performance for the first evaluation. After performance evaluation on D_1 (D_F) the classifiers were updated with trajectories in D_1 (D_F) and tested on D_2 (D_R). The same process was repeated for update/test on D_2 (D_R) and D_3 (D_L), respectively, in both scenarios with gradual and abrupt changes.

The approaches capable of incremental learning [PFAM, Learn++ (PFAM) and EoD$_k$ (PFAM)] were updated with only the new labelled 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 optimisation, and BC was applied for decision level fusion of classifiers on the ROC space [2]. 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 ten independent trials. Target samples from the learning set were randomly split according to a uniform distribution, in five-folds of the same size. The folds were first distributed in three different

IET Comput. Vis., 2015, Vol. 9, Iss. 5, pp. 732–740

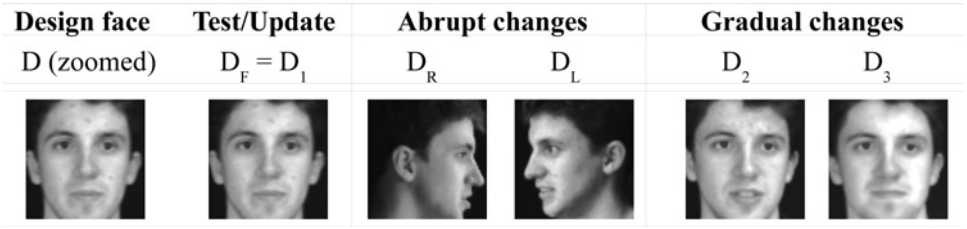

Fig. 6 *Samples of design/update facial regions from one of the individuals enrolled to the system (ID 188)*

Faces were detected in video sequences from the FIA database using the Viola–Jones face detector trained with frontal faces for gradual changes, and frontal, right and left poses for abrupt changes

design sets, including two-folds for training (D_t^t), 1(1/2) folds to stop training epochs (D_t^e) and 1(1/2) folds for fitness evaluation (D_t^f). Once the classifiers were trained, D_t^e and D_t^f are combined, randomised and divided into 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 sets at each replica of the experiment. At replication 5, the five-folds were regenerated after a randomisation 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 seven 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 optimise 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 initialised 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 five extra iterations to ensure convergence. Once the global best particle is found, its classifier and the six local bests from each subswarm were added to the EoD.

The evaluation measures used in comparison employ measures from the ROC (receiver operating characteristics) and PROC (precision–recall ROC) spaces. The true positive rate (tpr or recall) and false positive rate (fpr) are defined according to the number of true positives (TP), false positive (FP), true negative (TN) and

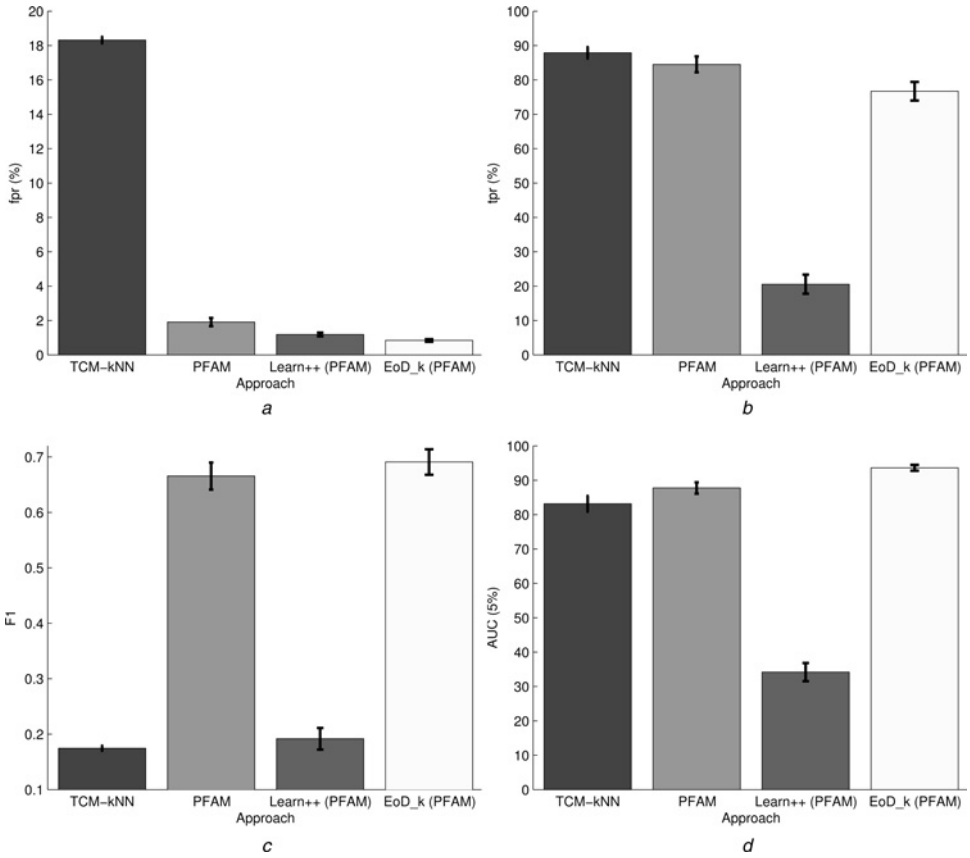

Fig. 7 *Average transaction-based performance of the different classifiers after two updates ($D_1 \rightarrow D_2 \rightarrow D_3$)*

More details on this comparison can be found in [16]
The tpr, fpr and F_1 measure are estimated at the operations point selected for a fixed fpr = 1%
a fpr
b tpr
c F_1
d pAUC (5%)

Table 1 Average performance of the system on ten individuals and ten trials, for the scenarios with gradual (top) and abrupt (bottom) changes

fpr (%) ↓	tpr (%) ↑	F_1 ↑	pAUC (5%) ↑
Gradual changes ($D_1 \rightarrow D_2 \rightarrow D_3$)			
EoD$_k$ (PFAM) LTM$_{KL,\lambda_k=\infty}$			
0.62 ±	77.02 ±	0.6789 ±	92.88 ±
0.09→0.67 ±	2.10→45.51 ±	0.0177→0.4041 ±	0.81→72.03 ±
0.05→0.84 ±	3.63→76.70 ±	0.0308→0.6909 ±	2.76→93.64 ±
0.07	2.71	0.0231	0.84
Abrupt changes ($D_F \rightarrow D_R \rightarrow D_L$)			
EoD$_k$ (PFAM) LTM$_{KL,\lambda_k=\infty}$			
0.62 ±	77.02 ±	0.6789 ±	92.88 ±
0.09→5.38 ±	2.10→13.48 ±	0.0177→0.0571 ±	0.81→22.0747 ±
1.13→2.73 ±	2.444→11.68 ±	0.0121→0.0605 ±	2.598→19.68 ±
0.34	2.42	0.0147	2.5450

Operations point selected at fpr = 1%.

false negative (FN) predictions obtained on an evaluation set. The tpr and fpr are given by

$$\text{tpr} = \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{fpr} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{7}$$

On the other hand, the F_1 measure is defined in the POC space, and is given by

$$F_1 = \frac{2}{1/\text{precision} + 1/\text{recall}} \tag{8}$$

where the precision is defined by

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

6 Simulation results

Fig. 7 presents the average performance of the system for the ten individuals of interest, after incremental learning. The ROC and PROC performance spaces are used for comparison, with partial area under the ROC curve for $0 \leq \text{fpr} \leq 0.05$ [pAUC (5%)], and empiric estimation of tpr, fpr and F_1 measure.

Fig. 7*a* shows that TCM-kNN yields the highest fpr, which is related to the difficulty faced by multi-class classifiers in finding multiple boundaries during the same optimisation process. In contrast, Learn++ (PFAM) and EoD$_k$ (PFAM) LTM$_{kl,\lambda_k}$ present the lowest fpr, proving the enhanced capacity of ensemble-based classifiers to discard non-target samples. Besides, Fig. 7*b* shows that TCM-kNN presents the highest tpr, followed by the PFAM and EoD$_k$ (PFAM) in the third place. In general, Figs. 7*c* and *d* show that the EoD$_k$ (PFAM) with an LTM managed with KL divergence presents the highest overall performance with a lower standard error.

Table 1 presents the average performance obtained after incremental learning in the scenarios with gradual and abrupt changes. Regarding the pAUC (5%), the tendency shown by the system in a scenario with gradual changes is characterised by an increase in the performance after two adaptations. An opposite tendency is shown on a scenario with abrupt changes, where the performance is constantly decreasing. This tendency is natural since facial models are designed with frontal faces, and it is required to recognise the individuals on right or left poses (see Fig. 6). However, the system behaves differently for each

Table 2 Average performance of the EoD$_{58}$ and EoD$_{188}$ after tests on scenarios of gradual ($D_1 \rightarrow D_2 \rightarrow D_3$) and abrupt ($D_F \rightarrow D_R \rightarrow D_L$) changes

	Gradual changes		Abrupt changes	
	EoD$_{58}$	EoD$_{188}$	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**	0.23 ± 0.09→29.51 ± 1.83→3.71 ± 0.407	2.54 ± 0.57→1.952 ± 0.17→3.17 ± 0.64
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	84.43 ± 3.33→43.33 ± 3.35→0.62 ± 0.15	89.58 ± 4.26→28.33 ± 2.05→**6.15 ± 0.87**
F_1 ↑	0.8492 ± 0.023→0.4029 ± 0.061→0.5710 ± 0.043	0.4720 ± 0.054→0.6594 ± 0.038→**0.8730 ± 0.027**	0.8492 ± 0.023→0.0134 ± 0.001→0.0016 ± 0.001	0.4720 ± 0.054→0.3119 ± 0.021→0.0370 ± 0.005
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	98.45 ± 0.23→8.15 ± 0.57→8.93 ± 0.4281	91.12 ± 2.41→38.71 ± 1.73→14.51 ± 0.97
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	0.23 ± 0.09→33.23 ± 1.71→**2.98 ± 0.13**	2.54 ± 0.57→2.62 ± 0.16→1.83 ± 0.29
tpr↑	84.43 ± 3.33→41.49 ± 7.76→94.65 ± 3.25	89.58 ± 4.26→**97.60 ± 0.64**	84.43 ± 3.33→48.33 ± 3.96→0.16 ± 0.049	89.58 ± 4.26→26.51 ± 1.86→5.38 ± 1.11
F_1 ↑	0.8492 ± 0.023→0.4171 ± 0.064→0.5619 ± 0.053	0.4720 ± 0.054→0.6838 ± 0.026→0.8511 ± 0.033	0.8492 ± 0.023→0.0122 ± 0.001→0.0007 ± 0.001	0.4720 ± 0.054→0.2743 ± 0.017→**0.0385 ± 0.007**
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	98.45 ± 0.23→8.44 ± 0.60→**9.78 ± 0.45**	91.12 ± 2.41→38.19 ± 1.22→17.94 ± 1.16
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**	0.23 ± 0.09→30.42 ± 1.56→4.14 ± 0.23	2.54 ± 0.57→2.59 ± 0.22→**1.74 ± 0.26**
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**	84.43 ± 3.33→45.00 ± 3.52→**4.06 ± 0.88**	89.58 ± 4.26→31.52 ± 2.08→5.38 ± 1.01
F_1 ↑	0.8492 ± 0.023→0.3808 ± 0.071→**0.6168 ± 0.053**	0.4720 ± 0.054→0.6669 ± 0.032→0.8720 ± 0.022	0.8492 ± 0.023→0.0125 ± 0.001→**0.0151 ± 0.004**	0.4720 ± 0.054→0.3231 ± 0.020→0.0379 ± 0.007
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**	98.45 ± 0.23→8.44 ± 0.61→9.33 ± 0.47	91.12 ± 2.41→41.25 ± 1.20→**19.42 ± 1.13**

Bold numbers symbolise significant difference with respect to the different sizes of LTM

IET Comput. Vis., 2015, Vol. 9, Iss. 5, pp. 732–740

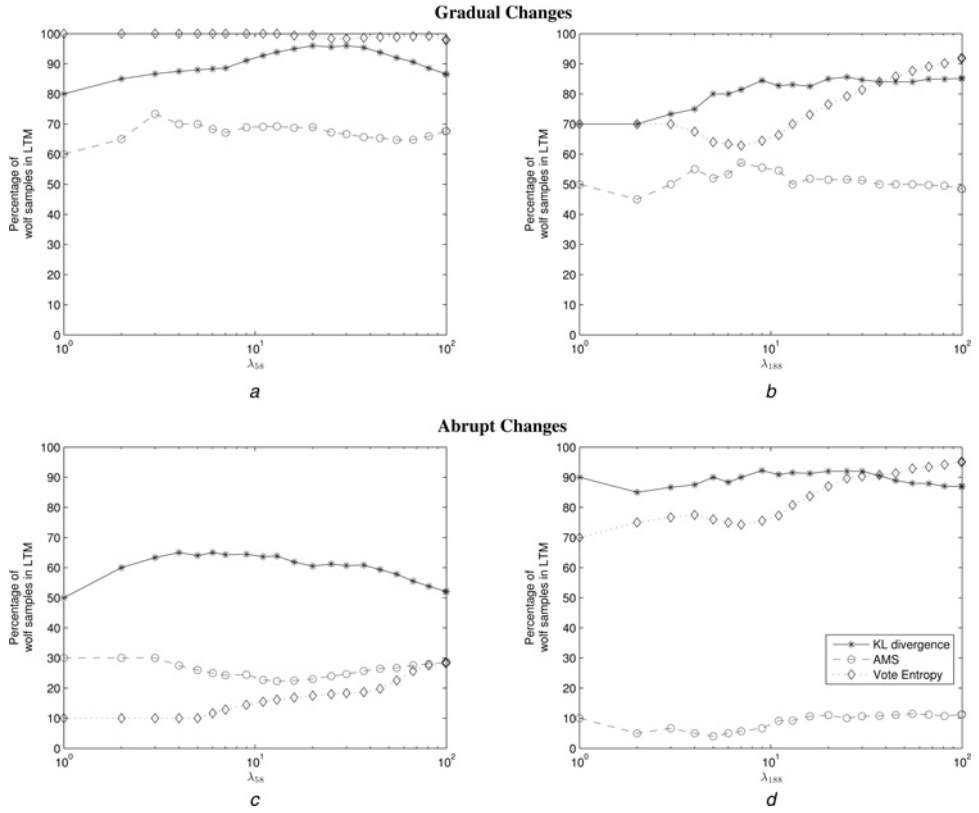

Fig. 8 *Average percentage of samples from wolf-like individuals for the EoD_{58} (a and c) and EoD_{188} (b and d), in the scenarios with gradual (upper graphs) and abrupt (lower graphs) changes*

a LTM_{58}, gradual changes
b LTM_{188}, gradual changes
c LTM_{58}, abrupt changes
d LTM_{188}, abrupt changes

individual in each scenario, and the impact of using an LTM is also different in each case.

Table 2 presents the performance of the ensemble during incremental learning for two individuals, using $\lambda_k = 25$, 75 and 100. EoD_{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 [9]. Conversely, EoD_{188} was selected because of its low initial performance (pAUC (5%) < 95%). It corresponds to a 'lamb'-like individual that even though is easy to detect by the system (tpr > 80%), it is also easy to impersonate (fpr > 1%). For individual 188, the test on D_1 throws 32 non-target individuals that are wrongly detected more than 1% of the time ('wolves').

Regarding the scenario with gradual changes, the F_1 measure for EoD_{58} after test on D_2, results show a performance that declines more importantly for EoD_{58} with $\lambda_{58} = 100$, and using a $\lambda_{58} = 75$ shows the best performance. However, after test on D_3, the appearance of new representative samples in the LTM leads to a recovery in the performance. A similar but smaller recovery is presented by EoD_{58} in the scenario with abrupt changes, suggesting that sheep-like individuals benefit from high λ_k values either in scenarios with gradual or abrupt changes.

A different trend is shown by EoD_{188} in the scenario with gradual changes, which in general presents a performance increases every time it is updated, regardless the value of λ_{188}. 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 $\pm$ 5.07 (global average for the ten individuals over the ten trials), and $\lambda_{188} = 25$ samples constitutes the 17.97% of the data.

Regarding the scenario with abrupt changes, the EoD_{188} shows a performance decrease as expected by pose changes. However, regarding its final performance, the use of large λ_{188} values significantly benefits its final performance. This suggests that lamb-like individuals are benefited by large λ values in scenarios with abrupt changes.

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. 8 shows the 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 100 samples characterising the scenarios with gradual and abrupt changes. The three selection strategies presented in Section 3 are compared. Regarding the scenario with gradual changes, it can be seen that LTM management strategies based on KL divergence and VE are successful in storing samples from wolf-like individuals, and the KL divergence retrieves the highest percentage for the lamb-like individual 188 (Fig. 8*b*). Results for the scenario with abrupt changes reveal that the KL divergence overcomes the other strategies at retrieving a greater proportion of samples from wolf-like individuals, either for lamb- or sheep-like target individuals. This becomes more evident for small values of λ.

Finally, when a new trajectory for an individual of interest becomes available, it takes around 150 min to update its facial model, and the modular architecture allows for parallel update of multiple facial models. The algorithm was implemented in Matlab® R2010B, running on Linux Gentoo, on a 2.53 GHz Intel® Xeon® processor. This makes the system appropriate for off-line update from, for example, daily police reports.

7 Conclusion

In this paper, an individual-specific strategy was proposed for the management of reference samples used for validation of adaptive ensembles applied to face re-identification. When new reference samples become available for an individual enrolled to the system, its facial regions are combined with non-target samples from the UM and CM selected with OSS. Old and new validation samples are combined and ranked using KL divergence, and the highest ranked are stored in an LTM for future validations. The theoretical foundation of this relevance measure lies on the relative entropy, where 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 emulating scenarios with gradual (aging) and abrupt (pose) changes in the classification environment. Simulation results indicate that using the proposed strategy allows individual-specific 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 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 KL divergence to rank and select samples from wolf-like individuals, compared to vote entropy and average margin sampling. Future research includes investigating strategies to find the optimal amount of samples required for each EoD, affecting a trade-off between performance and resources.

8 Acknowledgments

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. This work was also supported by the Program for the Improvement of the Professoriate of the Secretariat of Public Education, Mexico, and the Mexican National Council for Science and Technology.

9 References

1 Connolly, J.F., Granger, E., Sabourin, R.: 'Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition', *Pattern Recogn.*, 2012, **45**, pp. 2460–2477
2 De-la Torre, M., Granger, E., Radtke, P.V.W., Sabourin, R., Gorodnichy, D.O.: 'Incremental update of biometric models in face-based video surveillance'. Int. Joint Conf. Neural Networks, Brisbane, Australia, June 2012, pp. 1–8
3 De la Torre, M., Granger, E., Radtke, P.V.W., Sabourin, R., Gorodnichy, D.O.: 'Self-updating with facial trajectories for video-to-video face recognition'. Int. Conf. on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014, pp. 1–8
4 Kubat, M., Matwin, S.: 'Addressing the curse of imbalanced training sets: one-sided selection'. Int. Conf. Machine Learning, Nashville, USA, July 1997, pp. 179–186
5 Polikar, R., Udpa, L., Udpa, S.S., Honavar, V.: 'Learn++: an incremental learning algorithm for MLP networks', *Int. Conf. Syst. Man Cybern.*, 2001, **31**, (4), pp. 497–508
6 Freni, B., Marcialis, G., Roli, F.: 'Template selection by editing algorithms: a case study in face recognition'. Int. Association of Pattern Recognition, Orlando, USA, December 2008, vol. 5342, pp. 745–754
7 Goh, R., Liu, L., Liu, X., Chen, T.: 'The CMU face in action database'. Analysis and Modeling of Faces and Gestures, Beiging, China, October 2005, pp. 255–263
8 Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: 'Sheep, goats, lambs and wolves: a statistical analysis of speaker performance'. Int. Conf. Spoken Language Processing, Sidney, Australia, December 1998, pp. 1351–1354
9 Li, F., Wechsler, H.: 'Open set face recognition using transduction', *Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (11), pp. 1686–1697
10 Viola, P., Jones, M.: 'Robust real-time face detection', *Int. J. Comput. Vis.*, 2004, **2**, (57), pp. 137–154
11 Matta, F., Dugelay, J.-L.: 'Person recognition using facial video information: a state of the art', *J. Vis. Lang. Comput.*, 2009, **20**, (3), pp. 180–187
12 Jain, A.K., Ross, A.: 'Learning user-specific parameters in a multibiometric system'. Int. Conf. Image Processing, Rochester, USA, September 2002, pp. 57–60
13 Kamgar-Parsi, B., Lawson, W., Kamgar-Parsi, B.: 'Toward development of a face recognition system for watchlist surveillance', *Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (10), pp. 1925–1937
14 Ekenel, H.K., Szasz-Toth, L., Stiefelhagen, R.: 'Open-set face recognition-based visitor interface system'. Computer Vision Systems, Liege, Belgium, October 2009, vol. 5815, pp. 43–52
15 Pagano, C., Granger, E., Sabourin, R., Gorodnichy, D.O.: 'Detector ensembles for face recognition in video surveillance'. Int. Joint Conf. on Neural Networks, Brisbane, Australia, June 2012, pp. 1–8
16 De-la Torre, M., Granger, E., Sabourin, R., Gorodnichy, D.O.: 'An individual-specific strategy for management of reference data in adaptive ensembles for person re-identification'. Int. Conf. on Imaging for Crime Detection and Prevention, London, UK, December 2013, pp. 1–7
17 Guo, X., Yin, Y., Dong, C., Yang, G., Zhou, G.: 'On the class imbalance problem'. Int. Conf. on Computing, Networking and Communications, Piscataway, USA, August 2008, vol. 4, pp. 192–201
18 Scheffer, T., Decomain, C., Wrobel, S.: 'Active hidden Markov models for information extraction'. Int. Symp. on Intelligent Data Analysis, Berlin, Germany, June 2001, vol. 2189, pp. 309–318
19 Kachites McCallum, A., Nigam, K.: 'Employing EM and pool-based active learning for text classification'. Int. Conf. Machine Learning, San Francisco, USA, July 1998, pp. 350–358
20 Dagan, I., Engelson, S.P.: 'Committee-based sampling for training probabilistic classifiers'. Int. Conf. Machine Learning, San Francisco, USA, July 1995, pp. 150–157
21 Lim, C.P., Harrison, R.F.: 'An incremental adaptive network for on-line supervised learning and probability estimation'. Neural Networks, Houston, USA, June 1997, pp. 925–939
22 Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 971–987