Adaptive Skew-Sensitive Fusion of Ensembles and their Application to Face Re-Identification

Miguel De-la-Torre*†, Eric Granger*, Robert Sabourin*
* 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

***Abstract*—Adaptive classifier ensembles have been shown to improve the accuracy and robustness of systems for face recognition (FR) in video surveillance. However, it is often assumed that the proportions of faces captured for target and non-target individuals are balanced, or they are known a priori, and constant over time. Some active approaches have been proposed to update the ensemble during operations according to class imbalance of the input data stream. Beyond the estimation operational class imbalance, these approaches commonly generate diverse pools of classifiers by selecting balanced training data, limiting the potential diversity provided by the abundant non-target data. In this paper, a skew-sensitive ensemble is proposed to adaptively combine classifiers trained with data selected to have varying levels of imbalance and complexity. Given a face re-identification application, faces captured for each person appearing in the scene are tracked and regrouped into trajectories. During enrollment, faces in a reference trajectory are combined with those of selected non-target trajectories to generate a pool of 2-class classifiers using data with various levels of imbalance and complexity. During operations, the level of imbalance is periodically estimated by comparing input trajectories and pre-computed histograms using Hellinger distance quantification. Ensemble fusion functions are then adapted based on the imbalance and complexity of operational data. Finally, ensemble scores are accumulated over trajectories for robust spatio-temporal FR. Results obtained in experiments with synthetic data and Face in Action videos reveal that the proposed approach can significantly improve performance across operational imbalances.**

I. Introduction

In video surveillance, face re-identification applications involve recognizing the face of individuals that have previously appeared over a network of video cameras. Systems for video-to-video FR are commonly employed in such applications to match facial trajectories[1] captured in either live (real-time monitoring) or archived (post-event analysis) videos against the facial models of all target individuals enrolled to the system. Given a video surveillance system, an analyst may capture a reference facial trajectory corresponding to a target individual appearing in video feeds, and then design or update a facial model (e.g., a set of templates or a statistical representation) to be stored in the gallery. In face re-identification scenarios, several persons may appear before a camera view point, and their appearance typically varies either abruptly or gradually due to changes in, e.g., illumination, blur, scale and pose. Changes in the capture conditions are associated with changes in the underlying class distribution of operational data in the face matching space. Uneven proportions of ROI patterns from target and non-target individuals are related to the prior probability of individuals, and are commonly referred to as class imbalance or skew in pattern recognition literature.

Some recent systems for video-to-video FR have been successfully designed using adaptive ensembles of 2-class (target vs. non-target) classifiers [1]. They allow to represent and update facial models based on new reference trajectories, yet avoid knowledge corruption. Other ensemble-based methods have also been proposed to address the class imbalance problem in video FR [2]. This paper focuses on the design of facial models using adaptive skew-sensitive ensembles of 2-class classifiers.

Several methods have been proposed to generate diversified pools of classifiers by varying the complexity (class overlap and dispersion) [3] and imbalance [4] of class distributions. Moreover, recent research suggests that specialized selection and fusion methods that consider both complexity and imbalance may lead to robust ensembles [5]. A representative techniques for active fusion under changing class imbalance is skew-sensitive Boolean combination (SSBC). It allows to estimate class proportions using the Hellinger distance between histogram representations of operational and validation samples [2]. However, the accuracy of estimates on operational imbalance is limited by the number of imbalance levels (histograms) generated using validation samples. Furthermore, using classifier generation methods that only rely on data complexity limits the diversity of ensembles.

In this paper, adaptive skew-sensitive classifier ensembles are proposed to address applications (like face re-identification) where the operational class imbalance changes over time. During enrollment of a target individual, facial captures from a reference trajectory are combined with selected non-target captures from the universal (UM) and cohort (CM) models[2] to generate a diverse pool of 2-class classifiers using data with various levels of imbalance and complexity. Training/validation sets with different imbalances and complexities are built through random under-sampling, and cover a range of imbalances from 1:1 to a maximum imbalanced 1:λ^{max}, where λ^{max} is estimated empirically. During operations, the level of imbalance is periodically estimated from the input data stream using a Hellinger distance (HD) quantification method. The operational level of imbalance is estimated employing

[1] A trajectory is defined as a set of facial regions of interest (ROIs) captured in video that correspond to a same high quality track of a person appearing across consecutive frames.

[2] In this paper, the UM is a database containing non-target trajectories from selected unknown people appearing in scene. The CM is database with trajectories belonging to other target individuals enrolled to the system.

978-1-4799-1959-8/15/$31.00

the Hellinger distance between validation and operational histogram representations of class distributions in the feature space, HDx [6]. Pre-computed histograms and ensemble fusion functions are then adapted to the imbalance and complexity of operational data. Finally, a decision threshold is applied to the accumulation over time of positive ensemble scores for robust spatio-temporal recognition.

The proposed approach had been validated on synthetic and video surveillance data sets, and compared against reference approaches. The synthetic problem was designed using samples generated from Gaussian distributions in a two-dimensional feature space, where the overlap for target and non-target classes is controlled. The Carnegie Mellon University Face In Action (FIA) video database was used to emulate face re-identification applications.

II. Skew-Sensitive Ensembles

Ensembles methods combine classifiers with diversity of opinions to increase classification robustness and accuracy. The design process can be divided into three steps – generation of a diversified pool of base classifiers, selection and fusion of base classifiers to improve performance [3]. Representative examples of ensemble methods are bagging, boosting, random subspaces, which employ different sets of data or features from the training set to build distinct base classifiers [3], [7]. Connolly et al. [8], authors take advantage of diversity in the hyperparameter space of classifiers to produce diversity of opinions. Well-known selection strategies include greedy search, clustering-based methods and ranking-based methods, and fusion strategies are often divided in feature-based (concatenation), score-based (average, meta kNN) and decision-based (majority vote) [9].

Algorithms designed for dynamically changing environments in data distributions, and particularly in the class priors, can be categorized according to the use of a change detection mechanism [10]. Active approaches seek explicitly to determine whether and when a change has occurred in the prior probability before taking a corrective action [2], [10]. Conversely, passive approaches assume that a change may occur at any time, or is continuously occurring, and hence the classifiers are updated every time new data becomes available [11], [10].

Specialized architectures with ensembles of detectors (2-class classifiers) per individual enrolled to the system have been employed for FR in video surveillance [12]. In this case, the generation of classifiers is based on different decision bounds produced by varying classifier hyper-parameters and presentation order of training data [8]. Boolean combination was employed to select and combine the base classifiers in the ROC space [13]. Several active approaches in literature employ ensembles for classification in imbalanced environments [2], [14]. Changing imbalance can be estimated by adding a mechanism to detect changes in prior probabilities. Examples of such mechanisms are based in Hellinger distance [2], Kullback Leibler divergence [15], or accounting for class-specific performance (e.g., *recall*) [14].

This paper proposes an active approache for skew-sensitive ensembles designed for face re-identification. Skew-sensitive Boolean combination (SSBC) is considered as a representative example, which estimates class imbalance using the Hellinger distance between the distributions of validation data and the most recent unlabeled operational samples [2]. During training, SSBC assumes that a diversified pool of binary classifiers $\mathcal{P} = \{p_1, ..., p_n\}$ has been generated, and operates at the ensemble selection and fusion levels to take advantage of the diversity of opinions. Validation data with different levels of imbalance is used to estimate the operations points of the BC function (covering the whole ROC space). Two validation sets with the same imbalances, the first (`OPT`) are employed to estimate the operational imbalance, and the other (`VAL`) is employed to select the operation point with the proper estimated imbalance.

During operations, the histogram `opd` corresponding to the most recent operational samples is accumulated over time, and the closest level of class imbalance $\lambda^* \in \Lambda$ is estimated by comparing `opd` to the data sets in `OPT` using the Hellinger distance. Then, λ^* is used to select the BC that corresponds to that imbalance. In the case λ^* is not available on Λ_{BC}, the BCs for the two closest imbalances are merged, and the convex hull is estimated.

The known levels of class imbalance used by the approach form the set $\Lambda = \{\lambda^{bal} = 1 : 1, ..., \lambda^{max}\}$. A subset of class imbalances $\Lambda_{BC} \subset \Lambda$ is selected from Λ to optimize a subset of BCs E. The subset of imbalances Λ_{BC} should contain evenly distributed intermediate class imbalance levels between the minimum λ^{bal} and the maximum level of imbalance λ^{max} inclusively. The sets `OPT` and `VAL` are generated from imbalanced reference data that follows λ^{max}. Different data sets with the levels of class imbalance defined in Λ, where the amount of target samples remains fixed, while the amount of non-target samples are added to the set through random under sampling.

The strength of the SSBC algorithm lies in the adaptive selection of suitable fusion functions (ROC operations points) according to the estimated operational imbalance. However, this technique assumes that the pool of classifiers is generated using balanced training data that provides enough diversity of opinions to classify when input operational data is imbalanced. Another issue with SSBC is the precision of class imbalance estimation, that is limited by the sampling strategy used to create the set of imbalances Λ. Specialized methods to quantify the class priors of unlabeled (operational) data have been proposed in literature [6], and two of them are summarized in the next section.

III. Estimation of Class Imbalance

Quantification (estimation of the class distribution in Bayesian terms) allows to approximate the number of samples belonging to each class in an unlabeled set [16], [17]. In the literature, different quantification methods appear and are based either on the classifier confusion matrix [18], the posterior probability estimates provided by a classifier [16], or the comparison of class conditional probability densities of data sets with known and unknown proportions [2], [6], [17]. Regarding the estimation task from the point of view of a classifier, two levels can be identified to estimate the class imbalance of a distribution represented by a set of unlabeled (operational) samples. Data-level estimation operates

in the feature space, employing the probability distribution of samples for each feature [2], [6]. On the other hand, score-level allows to employ the probability distribution of the scores generated by a probabilistic classifier.

Two representative quantification methods where recently proposed to use the Hellinger distance to estimate the prior probability of unlabeled data, either using the feature (HDx) or score (HDy) levels [6]. Given an unlabeled dataset $U = \{(\mathbf{a}^n) : n = 1, ..., N\}$ and a labeled validation dataset $V = \{(\mathbf{a}^m, l^m) : m = 1, ..., M\}$, the Hellinger distance between these two sets can be computed according to

$$HD(V,U) = \frac{1}{n_f}\sum_{f=1}^{n_f} HD_f(V,U), \tag{1}$$

where the feature-specific Hellinger distance is given by

$$HD_f(V,U) = \sqrt{\sum_{i=1}^{b}\left(\sqrt{\frac{|V_{f,i}|}{|V|}} - \sqrt{\frac{|U_{f,i}|}{|U|}}\right)^2}, \tag{2}$$

and where n_f is the number of features, b is the number of bins used to construct the feature-specific histogram representation of the probability density functions of the datasets. $|U|$ is the number of samples in U and $|U_{f,i}|$ is the number of samples whose feature f belongs to the bin i, similarly with $|V|$ and $|V_{f,i}|$ for the validation set V. The Hellinger distance between the probability densities of the unlabeled and validation sets can be computed by making the assumption

$$\frac{|V_{f,i}|}{|V|} = \frac{|S^-_{f,i}|}{|S^-|}P_v(-) + \frac{|S^+_{f,i}|}{|S^+|}P_v(+), \tag{3}$$

where $|S^-|$ is the number of non-target training samples and $|S^-_{f,i}|$ is the number of non-target samples whose feature f belongs to bin i in the histogram representation of the probability distribution of the training data S. Similarly, $|S^+|$ and $|S^+_{f,i}|$ are equivalent measures for the target class. The prior probability $P_v(+)$ (and similarly $P_v(-)$) can be manually assigned by the HDx quantification method.

For HDy, the Hellinger distance between the distributions of classifier outputs is estimated as

$$HD(V,U) = \sqrt{\sum_{i=1}^{b}\left(\sqrt{\frac{|V_{y,i}|}{|V|}} - \sqrt{\frac{|U_{y,i}|}{|U|}}\right)^2} \tag{4}$$

where $|U_{y,i}|$ and $|V_{y,i}|$ are the number of unlabeled and validation samples whose output y belongs to the bin $i = 1...b$. Similarly to the HDx method, the substitution to avoid sub-sampling and/or oversampling is given by

$$\frac{|V_{y,i}|}{|V|} = \frac{|S^-_{y,i}|}{|S^-|}P_v(-) + \frac{|S^+_{y,i}|}{|S^+|}P_v(+), \tag{5}$$

where $|S^+_{y,i}|$ and $|S^-_{y,i}|$ represent the number of non-target samples whose output y belongs to bin i in the histogram representation of the probability distribution of the scores.

The level of class imbalance in the proportions of a set of samples is related to the prior probability of target (and equivalently non-target) samples. Given an imbalanced validation set V with $|V|$ samples, this relationship follows the definition of prior probability given by

$$P(+) = 1 - P(-) = \frac{|V^+|}{|V|} = \frac{|V^+|}{|V^+| + |V^-|}, \tag{6}$$

where $|V^+|$ and $|V^-|$ correspond to the number of target and non-target samples in V respectively. In the notation followed in this paper, the level of imbalance is represented as

$$\frac{|V^+|}{|V^+|} : \frac{|V^-|}{|V^+|}, \tag{7}$$

and the number of target samples $|V^+|$ is given by the context. By simple algebraic substitution it is easy to see that both are representations of the same quantity. Hence, the HDx and HDy quantification methods provide an estimate of $P(+)$, and equivalently, an estimate of the class imbalance.

In the HDx and HDy quantification methods, the prior probabilities $P_v(+)$ and $P_v(-)$ are explicitly defined by a step size that divides the full range of values $0 \leq P_v(+) \leq 1$. The optimal size of each "small step" can be easily deducted by considering the maximum expected imbalance λ^{max}, which can be used to estimate the optimal size for these steps. In practice, the step size can be defined using the available validation set V, and is given by

$$STEPSIZE = P_{min}(+) = \frac{V^+}{V^+ + V^-}. \tag{8}$$

IV. Adaptive Skew-Sensitive Ensembles for Video-to-Video Face Recognition

Figure 1 depicts the proposed architecture for an adaptive skew-sensitive multi-classifier system (MCS) for video-to-video FR. It consists of a tracker, a skew-sensitive classification system with individual-specific ensembles, a spatio-temporal fusion module, a sample selection and a classifier design/update systems. It is an extension of the framework proposed in [1], and incorporates the functionality to adapt the individual-specific ensembles to the most recent operational imbalance.

A. Design/update phase

This phase is triggered when a new reference trajectory becomes available. Target samples are extracted and combined with non-target samples from UM and CM to form a learning data set D_k (for training and validation). D_k follows the maximum predefined imbalance λ^{max}, which is set *a priori* in accordance with the experience in the field. An individual-specific selection strategy is employed to choose non-target samples that achieves the maximum expected imbalance λ^{max}. The data set D_k is evenly divided for imbalanced generation (D_k^{GEN}) and fusion function computation (D_k^{val}). This allows to generate a pool $\mathcal{P}'_k$ of classifiers, which are incorporated to the previous pool following a *learn-and-combine* strategy (see Section IV-D). A long term memory (LTM) is employed to store individual-specific reference samples and avoid knowledge corruption [19]. Then, the validation samples used for combination are stored in the datasets opt^{max} for operational imbalance estimation (see Section IV-C) and the approximation of imbalanced BC. Finally, the skew-sensitive combination module allows to select the operations point with validation data according to the approximated imbalance λ^*.

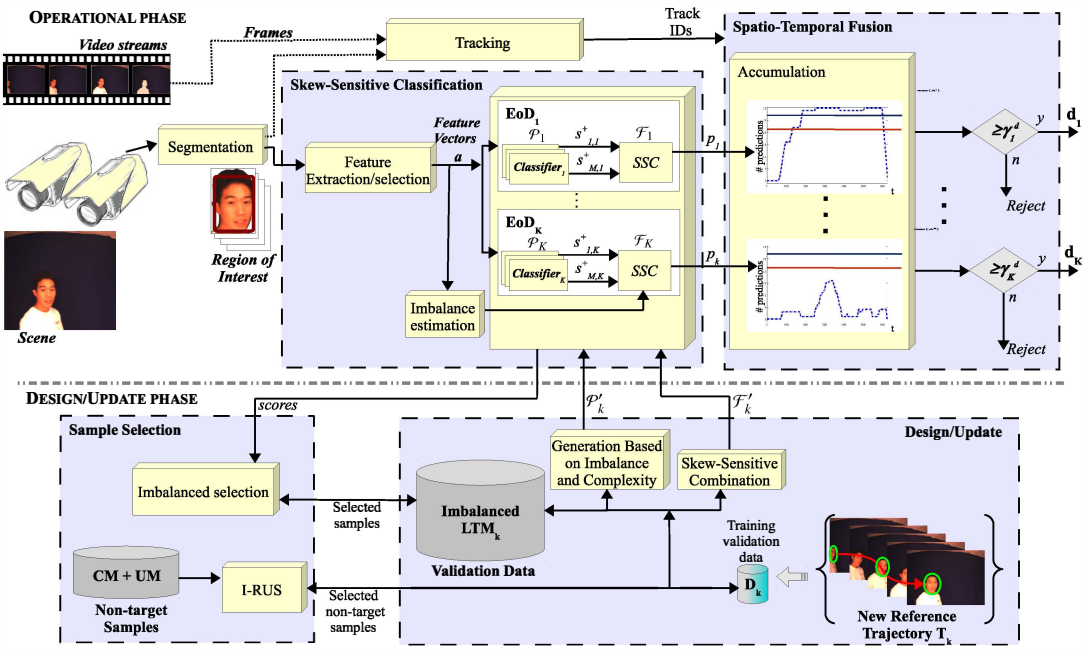


Fig. 1. Adaptive skew-sensitive MCS for video-to-video FR. It works in two different phases: operations and design/update of facial models (see [1]).

B. Operational Phase

The tracker follows the position of the segmented faces in video, building a face trajectory composed of consecutive ROIs from a same person in the scene. Simultaneously, features for classification are extracted and selected from segmented ROIs to form feature vectors ($\mathbf{a}$), which are processed by all the individual-specific ensembles of classifiers. The skew-sensitive ensemble for each enrolled individual k produces a sequence of predictions for input ROI patterns belonging to a face trajectory. In order to adapt the fusion function to the most recent operational imbalance, the feature specific histogram representation of the distribution of the operational data ($\mathbf{opd}$) from facial captures of the last predefined time period (e.g. 15 minutes) is computed. The most recent distribution stored in $\mathbf{opd}$ is employed to estimate the operational imbalance λ^* (see Section IV-C). Then, the combination function corresponding to the estimated operational imbalance λ^* is approximated, and the operations point ($\mathbf{op}$) in each individual-specific ensemble is selected. Finally, the spatio-temporal fusion module accumulates ensemble predictions over a fixed-size window of time. When the accumulation of predictions from an ensemble k surpasses a pre-defined detection threshold γ_k^d, the individual of interest k is detected in scene. If self-update is required and thee accumulation surpasses a second update threshold γ_k^u, the adaptation process starts employing all the ROIs belonging to the trajectory [1].

C. Approximation of Operational Imbalance

Initially, the classification system starts its operation considering a balanced operational environment. Feature vectors extracted from input facial regions are used to populate a data set with the most recent operational samples $\mathtt{ops}$. The $\mathtt{ops}$ set is renewed with new input samples over an user-defined period of time. The operational feature histogram is estimated based on the $\mathtt{ops}$ set. Then, the prior probability of the most recent target class distribution $P^*(+)$ of operational samples is estimated using HDx quantization, based on the feature histograms from unlabeled operational ($\mathtt{ops}$) and reference validation ($\mathtt{opt}^{\mathtt{max}}$) samples.

Let $|V^+|$ be the number of target samples in a validation data set V (e.g., $\mathtt{opt}^{\mathtt{max}}$). The number of non-target samples required to accomplish with the estimated class distribution $P^*(+)$ is given by

$$|V^-| = |V^+| \left(\frac{1}{P^*(+)} - 1 \right), \tag{9}$$

and the estimated class imbalance λ^* can be represented assuming $|V^+| = 1$ and substituting in Eqn. 7.

The HDx quantification method require a single validation set ($\mathtt{opt}^{max}$), which stores data from the abundant non-target samples that provide information from both imbalance and complexity in the feature space. The procedure for imbalance estimation is summarized in Algorithm 1.

Algorithm 1: Estimation of the level of imbalance λ^* from reference data $\mathtt{opt}^{max}$ and operational data $\mathtt{ops}$

Input : Data set $\mathtt{opt}^{max}$, Operational samples $\mathtt{ops}$, number of bins b
Output : Imbalance estimation λ^*
Estimate prior probability using HDx
Assume $|V^+| = 1$
Compute $|V^-|$ using Eqn. 9
Compute imbalance λ^* using Eqn. 7

The adaptation of the combination function based on the newly-approximated class imbalance λ^* is performed in accordance to the skew-sensitive algorithm, either by updating the combination weights (weighted voting or meta-classification combiners) or by selecting the imbalance-specific operations point (SSBC). The advantage of using an estimation of the prior probability as given by HDx provides a good estimation of the class imbalance, and the selection of the correct

imbalance in validation set VAL reduces the error propagation induced by some algorithms for imbalance estimation.

D. Design and Adaptation of Ensembles

The imbalance-based generation strategy proposed in this subsection allows to generate diversity of opinions, which can be successfully exploited with other skew-sensitive combination strategies. The operational imbalance in a real scenario suffers from constant changes, and should not assume a constant level of imbalance. Active skew-sensitive ensembles allow to estimate the operational imbalance, and select and combine the classifiers from a pool. Robustness of the ensembles may be enhanced with base classifiers trained on different levels of imbalance and complexity.

Limitations in resources make impractical to train a dedicated classifier for every possible level of imbalance, and a number of training imbalances should be fixed before training. The combination function is responsible for selecting of the classifiers with the proper imbalance and complexity, according to the operational data. In this way, given predefined minimum and maximum imbalances denoted by λ^{min} and λ^{max} respectively, a fixed number of imbalances is chosen between them.

Algorithm 2: Generation of diversified classifiers based on different levels of imbalance and complexity

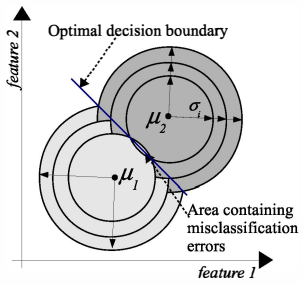
Input : Training data D_t, maximum imbalance λ^{max}_{GEN}, levels of imbalance $|\Lambda_{GEN}|$, size of subpools sp.
Output : $\mathcal{P}$ Pool of $|\Lambda_{GEN}| \times sp$ diversified classifiers.
Generate Λ_{GEN} by sampling the levels of imbalance with a log scale (e.g. 1:10^0, 1:$10^{\frac{log_{10}(\lambda^{max})\times i}{maxClsf-1}}$, ... 1:100)
Generate the imbalanced training sets D_i^{Imb} according to the imbalances in Λ_{GEN}
for i = *1*...$|\Lambda_{GEN}|$ **do**
  Train a new subpool with sp classifier $\mathcal{P}_i$ using D_i^{Imb} and a source of diversity
  $\mathcal{P} \Leftarrow \mathcal{P} \cup \mathcal{P}_i$

The procedure proposed for imbalance-based generation of diversified classifiers is shown in the Algorithm 2. In order to generate more diversity, the sub-pools of classifiers for each specific imbalance can be generated employing typical sources of diversity like different subsets of data, presentation orders, distinct hyperparameters, or other techniques.

V. Experiments with Synthetic Data

The objectives of these experiments include observing the impact on performance of designing adaptive ensembles trained on different levels of imbalance. The effectiveness of the skew-sensitive ensembles is compared to other ensemble techniques, and the generation of more than one classifier per imbalance level is evaluated, with diversity enhanced by different complexities.

To emulate a face re-identification scenario, a Gaussian distribution was employed to generate samples for the minority target class (individual of interest), and a second Gaussian distribution to draw samples for majority class (non-target individuals). The two overlapping multivariate Gaussian distributions with simple linear decision boundaries are shown in Figure 2a. These distributions are maintain a fixed center of mass $\mu_1 = [0, 0]$, $\mu_2 = [3.29, 3.29]$, and the degree of overlap was controlled by adjusting the covariance matrix σ of both distributions. Ten different imbalance levels were used to train 2-class probabilistic Fuzzy ARTMAP (PFAM) classifiers [20], corresponding to a logarithmic sampling between balanced and the maximum level of imbalance $\lambda^{max} = 1 : 1000$.



(a) Synthetic overlapping

	Total Probability of Error	Covariance Matrix
1.	1%	$\mathbf{I}$
2.	5%	$2 \times \mathbf{I}$
3.	10%	$3.304 \times \mathbf{I}$
4.	15%	$5.038 \times \mathbf{I}$
5.	20%	$7.672 \times \mathbf{I}$
6.	25%	$11.90 \times \mathbf{I}$

(b) Covariances matrices

Fig. 2. (a) Representation of the synthetic overlapping data set used for simulations and (b) covariance matrices used to control the degree of overlap between distributions ($\mathbf{I}$ is the 2×2 identity matrix). The covariance matrix allows to change the degree of overlap, and thus the total probability of error between classes. These parameters were extracted from [21].

Figure 3 illustrates a logarithmic scheme and the optimal decision boundaries for the imbalances selected between λ^1_{GEN} and λ^{max}_{GEN}. It produces evenly distributed decision boundaries and generates diversity evidenced in feature space.

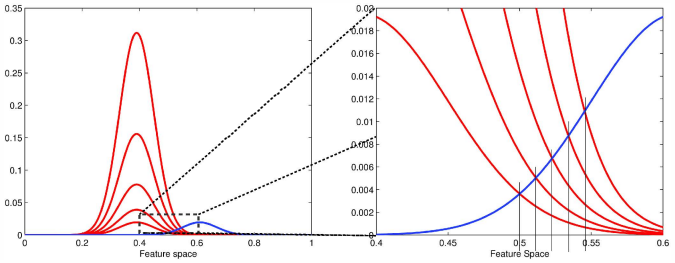
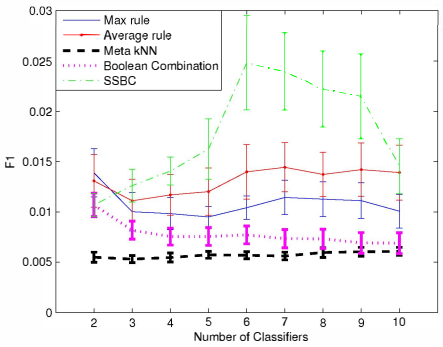


Fig. 3. Cross-cut of the overlapping data distributions for target (right-blue curves) and non-target (left-red curves) samples. This logarithmic scheme shows the imbalances $\Lambda_{GEN} = \{1 : 2^0, 1 : 2^1, 1 : 2^2, 1 : 2^3, 1 : 2^4\}$

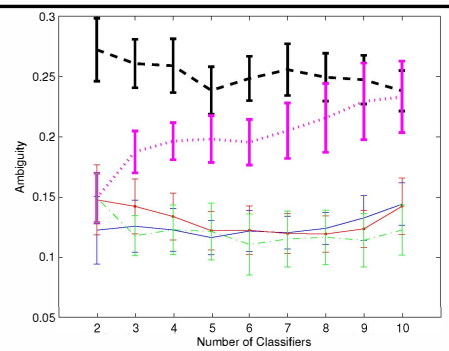
The standard hyperparameters of the PFAM classifiers were used (e.g. $[\alpha = 0.001, \beta = 1, \epsilon = 0.001, \bar{\rho} = 0, r = 0.60]$, [21]), and a hold-out validation process was employed to optimize the number of training epochs with different presentation orders. 10 target samples were maintained in training training and validation sets, which is typical of applications with limited training data. Similarly, the number of negative samples was variated according to the desired imbalances in Λ_{GEN}.

Five combination strategies are used for comparison. The MAX rule selects the maximum target score produced by the base classifiers in the pool. The AVG rule estimates the mean of the target scores. In meta kNN, the 1-NN classifier was trained on independent score-level validation data, and it was employed in test to produce output distance-based scores. For BC, the ten Boolean functions are applied to different pairs of classifiers, and the BC algorithm was run on an independent validation set to find the operation points that maximize the ROC convex hull [13]. Finally, the SSBC is applied with a validation set containing a profile with the same imbalance as the expected in test [2].

Imbalanced generation: This scenario provides a situation where the ensemble is deployed and an operational point at $fpr = 1\%$ provides the final decisions. The number of classifiers was variated from 2 to 10, adding one PFAM at a time in decreasing order according to AUC accuracy, evaluated on an independent validation set. The performance of all the approaches was evaluated on a test set with imbalance $\lambda_{GEN}^{max} = 1:1000$. The ambiguity and F_1 measure are shown for the comparison. The ambiguity was defined by Zenobi and Cunningham in [22], and include the responses of the base classifiers and the overall ensemble.



(a) F_1 measure



(b) Ambiguity

Fig. 4. Performance over the number of classifiers in the ensemble, using different combination strategies and adding the classifiers in decreasing order according to AUC evaluated on validation. Results are shown for the problem with total probability of error of 20%.

Figure 4 presents the resulting F_1 measure and ambiguity for the ensembles in the scenarios with total probability error of 20%. Regarding the F_1 measure, the maximum, average, BC and SSBC combinations perform better than meta-kNN at all times, and a significant increase in performance is shown by SSBC when the ensemble contains between 5 and 9 classifiers. The phenomenon was repeated for the other levels of overlap, becoming more evident as the total probability error grows. The ambiguity of the meta-kNN combination remains at a high compared to the other four approaches, which, combined with the low F_1 performance, allows to see that this approach exploits the diversity of opinions in a less efficient way. On the other hand, the ambiguity shown by SSBC remains low compared to the meta-kNN, reinforcing that useful diversity of opinions is exploited by this approach.

In Fig. 4a, it can be observed that the last value for SSBC in the curve, corresponding to 10 classifiers in the ensemble, is significantly higher than the first value (2 classifiers). This is related to the order in which base classifiers are added to the ensemble, and the limit of useful diversity of opinions provided by the base classifiers. The last (and least accurate) classifiers added to the ensemble, negatively affects diversity and thus, the global performance. This tendency is more evident in problems with a high level of total probability of error, in which the classifiers with less performance bias the ensemble towards the erroneous decisions. In general, the approaches that show a higher diversity tend to produce a lower performance, showing that there is a limit in the useful diversity, and beyond that limit, it damages the ensemble accuracy.

Table I shows a comparison between the combination strategies, considering 7 classifiers trained on different levels of imbalance. It can be seen that SSBC provides the most accurate fpr in all cases, remaining always close to the desired $fpr = 1\%$ regardless of the total probability of error between classes. In contrast, the average rule and meta kNN provide the highest tpr at the expenses of increased fpr. This is undesirable since false alarms should be limited in video surveillance in an environment with numerous non-target individuals. Comparing the F_1 measures for the different combination methods, SSBC significantly outperforms all other approaches. Only the problem with an overlap of 1% that seems to be better addressed by the meta kNN. From results, one would suggest that traditional combination methods are suitable to be used in imbalanced environments when the classification problems are easy enough (e.g. present a lower total probability of error between classes, simple decision boundaries, etc). However, as the total probability of error grows, SSBC outperforms the others. Note that this global performance is affected by a low tpr.

Exploiting complexity: Using more than one classifier for each level of imbalance allows to exploit the complexity of data to generate more robust ensembles. In this experiment, ensembles were augmented by increasing the number of classifiers per imbalance. Variations in the classifiers was introduced by changing the presentation order in the training sets. A sensitivity analysis was conduced to observe the performance of ensembles after changing the size of these sub-pools from 1 to 3 classifiers per imbalance, resulting in pools of 7, 14 and 21 classifiers. The test set was kept with the maximum imbalance.

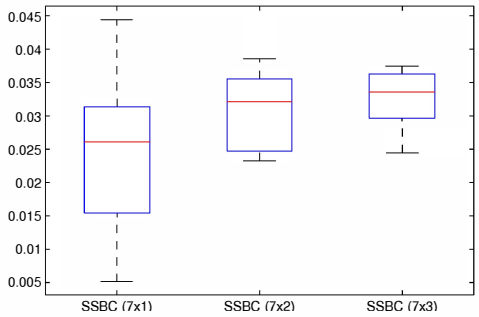


Fig. 5. F_1 box plots for the skew-sensitive ensembles with a pool of classifiers with 7 imbalances, problem with 20% total probability of error. A sub-pool for each of the imbalances was growth from one to three classifiers, resulting in pools of 7, 14 and 21 classifiers

Figure 5 presents the box plots for the F_1 performance achieved by skew-sensitive ensembles with different sizes of pools of classifiers. It can be seen that the median grows as the number of classifiers increases. The variations represented by

TABLE I. AVERAGE PERFORMANCE OBTAINED WITH DIFFERENT SYSTEMS. ENSEMBLES ARE COMPOSED OF 7 BASE CLASSIFIERS. THE BOLD NUMBERS REPRESENT THE PERFORMANCE VALUES SIGNIFICANTLY HIGHER THAN OTHER APPROACHES.

PFAM				**Score-level AVG**				**Meta kNN**				**Proposed approach**			
$fpr(\downarrow)$	$tpr(\uparrow)$	$prec(\uparrow)$	$F_1(\uparrow)$	$fpr(\downarrow)$	$tpr(\uparrow)$	$prec(\uparrow)$	$F_1(\uparrow)$	$fpr(\downarrow)$	$tpr(\uparrow)$	$prec(\uparrow)$	$F_1(\uparrow)$	$fpr(\downarrow)$	$tpr(\uparrow)$	$prec(\uparrow)$	$F_1(\uparrow)$
Total probability error: 1%															
13.26%	94.30%	1.61%	0.0314	11.74%	**99.90%**	1.57%	0.0307	1.37%	97.70%	**16.04%**	**0.2496**	**0.81%**	58.50%	6.85%	0.1219
(4.19)	(4.94)	(0.40)	(0.0078)	(3.06)	**(0.10)**	(0.39)	(0.0075)	(0.51)	(0.56)	**(5.03)**	**(0.0623)**	**(0.07)**	(5.53)	(0.45)	(0.0077)
Total probability error: 5%															
13.92%	50.30%	0.79%	0.0153	16.62%	**92.30%**	0.93%	0.0183	8.86%	87.40%	2.37%	0.0441	**0.93%**	57.50%	**6.17%**	**0.1102**
(3.70)	(11.06)	(0.32)	(0.0061)	(4.64)	**(2.21)**	(0.18)	(0.0034)	(2.16)	(2.33)	(0.99)	(0.0174)	**(0.08)**	(5.51)	**(0.73)**	**(0.0118)**
Total probability error: 10%															
12.32%	39.50%	0.75%	0.0140	13.71%	75.80%	1.40%	0.0267	15.67%	**81.70%**	0.62%	0.0122	**1.24%**	36.80%	**3.50%**	**0.0625**
(4.48)	(10.02)	(0.30)	(0.0054)	(4.16)	(5.32)	(0.47)	(0.0087)	(2.11)	**(3.80)**	(0.10)	(0.0019)	**(0.20)**	(4.07)	**(0.66)**	**(0.0106)**
Total probability error: 15%															
14.52%	42.00%	0.38%	0.0075	10.44%	49.10%	1.35%	0.0234	23.12%	**78.20%**	0.39%	0.0078	**1.13%**	21.80%	**2.16%**	**0.0390**
(3.55)	(9.58)	(0.10)	(0.0020)	(3.97)	(10.47)	(0.38)	(0.0059)	(2.50)	**(2.72)**	(0.06)	(0.0013)	**(0.13)**	(2.50)	**(0.34)**	**(0.0059)**
Total probability error: 20%															
19.00%	51.50%	0.28%	0.0057	11.99%	54.50%	0.74%	0.0144	27.88%	**75.00%**	0.28%	0.0056	**1.12%**	14.20%	**1.32%**	**0.0240**
(3.06)	(9.33)	(0.04)	(0.0007)	(3.77)	(5.68)	(0.13)	(0.0024)	(2.30)	**(2.56)**	(0.02)	(0.0004)	**(0.10)**	(2.13)	**(0.22)**	**(0.0038)**
Total probability error: 25%															
12.62%	32.40%	0.47%	0.0083	10.92%	42.20%	0.60%	0.0115	31.27%	**68.60%**	0.23%	0.0046	**1.22%**	8.10%	**0.67%**	**0.0123**
(3.78)	(6.52)	(0.15)	(0.0020)	(3.30)	(7.13)	(0.09)	(0.0017)	(2.56)	**(2.55)**	(0.02)	(0.0003)	**(0.10)**	(1.03)	**(0.07)**	**(0.0012)**

the upper and lower bars becomes narrower as the number of classifiers grows. The difference in performance between the second (7x2 classifiers) and the third (7x3 classifiers) boxes is small. Other criteria like spatial complexity may be used to select the size of the sub-pools.

VI. EXPERIMENTS WITH VIDEO SURVEILLANCE DATA

Assume that a surveillance camera continuously captures videos from a scene and feeds them to the segmentation module (Viola-Jones face detector) that isolates the facial regions of interest (ROIs) in each consecutive frame. Discriminant features are extracted using a face descriptor (multi-block local binary patterns), and the 32 principal components are selected after PCA. The tracking module (incremental visual tracking [23]) follows the face of each individual and regroups ROIs from a same individual in trajectories, whereas the classification module produces consecutive identity predictions for each ROI. Finally, the spatio-temporal decision fusion module accumulates target predictions and applies individual-specific thresholds for enhanced spatio-temporal FR [1]. In the reference system, the EoDs are co-jointly trained using a DPSO learning strategy to generate a diversified pool of PFAM classifiers [20]. In the proposed approach, the base classifiers are independently trained data sets with different levels of imbalances, and hyperparameters are optimized with the aforementioned DPSO learning strategy.

Videos from the CMU-Face in Action (FIA) database are employed. Each video corresponds to 20 seconds sequences from one of 244 individuals in a passport checking scenario [24]. Six cameras capture the scene at 640×480 pixels, with 30 frames per second. During each trial, 10 individuals of interest were randomly selected for enrollment, and the remaining non-target individuals were split into two independent subsets for training and test.

During enrollment, a pool of PFAM classifiers was generated according to $\Lambda_{GEN} = \{1:1, 1:10^{1/3}, 1:10^{2/3}, 1:10, 1:10^{4/3}, 1:10^{5/3}, 1:100\}$. The DPSO algorithm was initialized with a population size of 20 particles, a maximum of 6 subswarms of 5 particles maximum, and a maximum of 10 iterations [21]. The global best classifier was selected for each imbalance in Λ_{GEN}. Test videos were concatenated one after the other, emulating a passport checking scenario. Four blocks of 30 minutes were prepared (D_1, D_2, D_3 and D_4), with different imbalances. The first two blocks are composed of trajectories from capture session 2, and the last two blocks are composed of trajectories from capture session 3. Trajectories from blocks D_1 and D_3 were captured with an unzoomed camera, and trajectories from blocks D_2 and D_4 were captured with a zoomed camera. Learning is performed following 4x6-fold cross-validation for 24 independent trials. The first two folds are merged for training (D^t), and the rest are distributed in validation sets to stop training epochs (D^e), for fitness evaluation (D^f), estimation of fusion function (D^c) and selection of the operational point (D^s).

Table II shows the average performance using the reference system and SSBC using PFAM and ensembles with AVG score level fusion, meta-kNN, and the proposed approach, after selecting the operations point at $fpr = 1\%$. The performance of the proposed approach is significantly higher after adapting ensembles to the operational imbalance. It can be seen that the proposed approach maintains fewer false alarms after the operations point is adapted. This capacity is related to the use of the abundant non-target samples to establish the decision frontier at the fusion function, enhancing the discrimination between target and non-target classes. However, this is achieved at the expense of a lower tpr.

TABLE II. AVERAGE PERFORMANCE OF DIFFERENT APPROACHES AT AN $fpr = 1\%$ ON TEST BLOCKS AT DIFFERENT t TIMES. THE STANDARD ERROR IS DETAILED BETWEEN PARENTHESIS, AND BOLD NUMBERS SYMBOLIZE SIGNIFICANT DIFFERENCE IN TERMS OF F_1 MEASURE WITH RESPECT TO OTHERS.

Approach	**Measure**	$t=1$	$t=2$	$t=3$	$t=4$
SSBC [2]	fpr	5.15% (0.025)	4.15% (0.024)	4.71% (0.023)	3.30% (0.014)
	tpr / $recall$	61.54% (0.171)	56.94% (0.234)	59.74% (0.283)	59.41% (0.313)
	$precision$	23.19% (0.077)	24.67% (0.099)	30.61% (0.154)	34.43% (0.171)
	F_1	**0.300 (0.094)**	0.307 (0.135)	0.363 (0.183)	0.383 (0.217)
Proposed approach	fpr	5.15% (0.025)	1.47% (0.010)	1.61% (0.013)	1.11% (0.006)
	tpr / $recall$	61.54% (0.171)	54.60% (0.327)	49.79% (0.341)	54.40% (0.354)
	$precision$	23.19% (0.077)	40.82% (0.158)	48.82% (0.251)	48.13% (0.247)
	F_1	**0.300 (0.094)**	**0.422 (0.204)**	**0.434 (0.238)**	**0.477 (0.285)**

Facial trajectories were built using the IVT face tracker to regroup target facial regions, and used for trajectory-based

TABLE III. AVERAGE PERFORMANCE FOR THE REFERENCE AND PROPOSED APPROACHES, CONSIDERING THE 10 INDIVIDUALS OVER 24 TRIALS. THE STANDARD ERROR IS SHOWN IN PARENTHESIS.

	$t=1$	$t=2$	$t=3$	$t=4$
Average Imbalance	1:15	1:16	1:10	1:15
Number of target ROIs	85.3 (7.07)	102.7 (6.56)	79.3 (5.35)	95.0 (6.44)
SSBC [2] (AUC-5%)	67.87 (2.21)	67.67 (2.40)	71.41 (2.36)	73.36 (2.28)
Proposed approach (AUC-5%)	67.87 (2.21)	79.45 (1.98)	78.61 (2.14)	74.07 (2.57)

analysis of the system a passport-checking scenario. When a new face is captured in a video sequence, the location of the facial region is employed to initialize the tracker, and follows it until the individual leaves the scene. Target predictions produced by the system were accumulated over time for full trajectories to provide overall decisions, and the detection threshold was applied to these accumulations.

Table III shows the average operational imbalance, the number of target ROIs, as well as the average overall AUC for the ROC curves obtained over $0 \leq fpr \leq 0.05$ (AUC-5%). The AUC performance of the system the first test block, when the operational imbalance is not considered, is significantly lower compared to the performance after adapting the fusion function.

VII. CONCLUSION

In this paper, an adaptive skew-sensitive ensemble was proposed for video-to-video FR systems applied to face re-identification. The proposed approach adaptively combines 2-class classifiers trained by selecting data with varying levels of imbalance and complexity. Results on synthetic problems show that the adaptive skew-sensitive ensembles lead to a significant increase in ensemble diversity, robustness and performance. Similarly, results on CMU-FIA videos show that the proposed method can outperform reference techniques in imbalanced environments for face re-identification applications. The proposed approach maintains a low level of false positives and a high precision on synthetic and real data with imbalance, at the expense of a lower level of true positives.

Future research should consider exploiting adaptive skew-sensitive ensembles with imbalance-specific thresholds at decision fusion level. The characterization of the proposed system should also be performed in more challenging scenarios with uncontrolled capture conditions (e.g., crowded and outdoor scenes). Finally, adaptation of the proposed system to gradual or abrupt changes in the probability distribution of operational data due to varying capture conditions for facial appearance may be addressed employing self-update techniques, leading to further improvement.

REFERENCES

[1] M. De-la Torre, E. Granger, P. V. Radtke, R. Sabourin, and D. O. Gorodnichy, "Partially-supervised learning from facial trajectories for face recognition in video surveillance," *Information Fusion*, vol. 24, pp. 31–53, 2015.

[2] P. V. Radtke, E. Granger, R. Sabourin, and D. O. Gorodnichy, "Skew-sensitive boolean combination for adaptive ensembles –an application to face recognition in video surveillance," *Information Fusion*, vol. 20, pp. 31–48, 2013.

[3] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004.

[4] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 42, pp. 463–484, 2011.

[5] V. Lopez, A. Fernandez, S. Garcia, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp. 113 – 141, 2013.

[6] V. Gonzalez-Castro, R. Alaiz-Rodriguez, and E. Alegre, "Class distribution estimation based on the hellinger distance," *Information Sciences*, vol. 218, pp. 146–164, 2013.

[7] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Analysis and Applications*, vol. 1, pp. 18–27, 1998.

[8] J.-F. Connolly, E. Granger, and R. Sabourin, "Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition," *Pattern Recognition*, vol. 45, no. 7, pp. 2460 – 2477, 2012.

[9] Q. Tao and R. Veldhuis, "Hybrid fusion for biometrics: combining score-level and decision-level fusion," in *IEEE CVPR Workshops*, Piscataway, USA, 2008, pp. 1 – 6.

[10] G. Ditzler and R. Polikar, "Incremental learning of concept drift from streaming imbalanced data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 10, pp. 2283–2301, 2013.

[11] S. Oh, M. S. Lee, and B.-T. Zhang, "Ensemble learning with active example selection for imbalanced biomedical data classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 2, pp. 316–325, 2011.

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

[13] W. Khreich, E. Granger, A. Miri, and R. Sabourin, "Adaptive roc-based ensemble of hmms applied to anomaly detection," *Pattern Recognition*, vol. 45, pp. 208–230, July 2012.

[14] S. Wang, L. Minku, D. Ghezzi, D. Caltabiano, P. Tino, and X. Yao, "Concept drift detection for online class imbalance learning," in *IJCNN*, Dallas, USA, Aug 2013, pp. 1–10.

[15] M. C. du Plessis and M. Sugiyama, "Semi-supervised learning of class balance under class-prior change by distribution matching," in *ICML*, Edinburgh, UK, 2012, pp. 1–26.

[16] A. Bella, C. Ferri, J. Hernandez-Orallo, and M. J. Ramirez-Quintana, "Quantification via probability estimators," Sydney, Australia, 2010, pp. 737 – 742.

[17] G. Forman, "Quantifying counts and costs via classification," *Data Mining and Knowledge Discovery*, vol. 17, no. 2, pp. 164 – 206, 2008.

[18] Y. S. Chan and H. T. Ng, "Estimating class priors in domain adaptation for word sense disambiguation," in *Proc. Int. Conf. on Computational Linguistics and 44th Annual Meeting of the Asoc. Comp. Linguistics*, Stroudsburg, USA, 2006, pp. 89–96.

[19] M. De-la Torre, E. Granger, R. Sabourin, and D. O. Gorodnichy, "An individual-specific strategy for management of reference data in adaptive ensembles for face re-identification," in *ICDP*, IET, Ed., London, UK, December 2013, pp. 1–7.

[20] C. Lim and R. Harrison, "An incremental adaptive network for on-line supervised learning and probability estimation," *Neural Networks*, vol. 10, no. 5, pp. 925–939, 1997.

[21] E. Granger, P. Henniges, R. Sabourin, and L. S. Oliveira, "Supervised learning of fuzzy ARTMAP neural networks through particle swarm optimization," *J of PR Research*, vol. 2, pp. 27–60, 2007.

[22] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *Machine Learning*, 2001, vol. 2167, pp. 576–587.

[23] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. Journal of Computer Vision, Special Issue: Learning for Vision*, vol. 77, pp. 125–141, 2008.

[24] R. Goh, L. Liu, X. Liu, and T. Chen, "The CMU Face In Action Database," in *Analysis and Modelling of Faces and Gestures*. Carnegie Mellon University, 2005, pp. 255–263.