Pattern Recognition 48 (2015) 3385–3406

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Adaptive skew-sensitive ensembles for face recognition in video surveillance

Miguel De-la-Torre [a,b,*], Eric Granger [a], Robert Sabourin [a], Dmitry O. Gorodnichy [c]

[a] Laboratoire d'imagerie de vision et d'intelligence artificielle, École de technologie supérieure, Université du Québec, Montréal, Canada
[b] Centro Universitario de Los Valles, Universidad de Guadalajara, Ameca, Mexico
[c] Science and Engineering Directorate, Canada Border Services Agency, Ottawa, Canada

ARTICLE INFO

*Article history:*
Received 25 November 2014
Received in revised form
29 March 2015
Accepted 8 May 2015
Available online 19 May 2015

*Keywords:*
Adaptive classifier ensembles
Boolean combination
Imbalance estimation
Video-to-video face recognition
Video surveillance
Adaptive multiple classifier systems

ABSTRACT

Decision support systems for surveillance rely more and more on face recognition (FR) to detect target individuals of interest captured with video cameras. FR is a challenging problem in video surveillance due to variations in capture conditions, to camera interoperability, and to the limited representativeness of target facial models used for matching. Although adaptive classifier ensembles have been applied for robust face matching, it is often assumed that the proportions of faces captured for target and non-target individuals are balanced, known a priori, and do not change over time. Recently, some techniques have been proposed to adapt the fusion function of an ensemble according to class imbalance of the input data stream. For instance, Skew-Sensitive Boolean combination (SSBC) is a active approach that estimates target vs. non-target proportions periodically during operations using Hellinger distance, and adapts its ensemble fusion function to operational class imbalance. Beyond the challenges of estimating class imbalance, such techniques commonly generate diverse pools of classifiers by selecting balanced training data, limiting the potential diversity produced using the abundant non-target data. In this paper, adaptive skew-sensitive ensembles are proposed to combine classifiers trained by selecting data with varying levels of imbalance and complexity, to sustain a high level the performance for video-to-video FR. Faces captured for each person in the scene are tracked and regrouped into trajectories. During enrollment, captures in a reference trajectory are combined with selected non-target captures 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 from the input trajectories using the HDx quantification method, and pre-computed histogram representations of imbalanced data distributions. This approach allows one to adapt pre-computed histograms and ensemble fusion functions based on the imbalance and complexity of operational data. Finally, the ensemble scores are accumulated of trajectories for robust spatio-temporal recognition. Results on synthetic data show that adapting the fusion function of ensemble trained with different complexities and levels of imbalance can significantly improve performance. Results on the Face in Action video data show that the proposed method can outperform reference techniques (including SSBC and meta-classification) in imbalanced video surveillance environments. Transaction-based analysis shows that performance is consistently higher across operational imbalances. Individual-specific analysis indicates that goat- and lamb-like individuals can benefit the most from adaptation to the operational imbalance. Finally, trajectory-based analysis shows that a video-to-video FR system based on the proposed approach can maintain, and even improve overall system discrimination.



## 1. Introduction

Video surveillance systems commonly rely on spatio-temporal face recognition (FR) to detect the presence of target individuals of interest in live or archived videos, either for watchlist screening or search and retrieval applications. Video-to-video FR systems commonly match input facial trajectories[1] from videos against the facial models of all target individuals enrolled to the system, and raise a warning in the case of positive detection. In this challenging scenario several persons may appear before a camera

* Corresponding author. Tel.: +1152 375 75 80 500x47291; fax: +1152 375 75 80 500x47218.
*E-mail addresses:* miguel.delatorre@profesores.valles.udg.mx (M. De-la-Torre), eric.granger@etsmtl.ca (E. Granger), robert.sabourin@etsmtl.ca (R. Sabourin), dmitry.gorodnichy@cbsa-asfc.gc.ca (D.O. Gorodnichy).

[1] A trajectory is 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.

http://dx.doi.org/10.1016/j.patcog.2015.05.008
0031-3203/

view point, and their appearance varies either abruptly or gradually due to, e.g., changes in illumination and pose. Changes in the capture conditions are associated with changes in the representation of the underlying class distribution of data in the face matching space. Uneven proportions between target and non-target individuals are related to the prior probability of occurrence for a given individual, and are commonly referred to as class imbalance or skew.

Facial models used for matching are composed of a set of reference samples (for template matching), or a statistical model estimated during training with reference samples (for statistical or neural classification). For instance, some recent systems for face re-identification applications successfully employ adaptive ensembles of 2-class (target vs. non-target) classifiers to design and update facial models based on new reference trajectories, yet avoiding the knowledge corruption [1,2]. And approaches to address the class imbalance problem in face recognition have also been proposed [3,4]. This paper focuses on the design of facial models based on adaptive skew-sensitive ensembles of 2-class classifiers.

The effects of class imbalance on classifier performance have been shown by several authors [5–8], and pattern recognition literature presents several ensemble-based methods to train ensembles on imbalanced data [9]. Algorithms designed for environments with data distributions that change over time can be categorized according to the use of a mechanism to detect concept drift or change [10]. Approaches with active detection of changes in prior probabilities seek explicitly to determine whether and when a change has occurred in the prior probability before taking a corrective action [3,4,10]. Conversely, passive approaches assume that a change may occur at any time, or is continuously occurring, and hence classification systems are updated every time new data becomes available [10,11]. The advantage of active approaches mainly consists in the avoidance of unnecessary updates. However, they are prone to both false positive and false negative drift detections, with the respective false updates and false no-updates. Passive approaches avoid these problems at an increased computational cost due to the constant update.

A representative example of active approaches for changing imbalances is the skew-sensitive Boolean combination (SSBC) that continuously estimates the class proportions using the Hellinger distance between histogram representations of operational and validation samples [4]. Every time the operational imbalance changes, SSBC selects one of the pre-calculated fusion functions that correspond to a set of prefixed imbalances. However, the limited number of validation imbalance levels that can be used to approximate the imbalance in operations is a limiting factor for the estimation of operational imbalance. Rather than selecting the closest imbalanced histogram representations, more sophisticated estimation methods may be employed for accurate estimation of the class proportions. Moreover, although it is scarcely exploited, the abundant non-target samples in video surveillance allow one to produce training sets with different complexities and imbalances, and use them to generate diverse pools. A specialized combination and selection scheme of these diversified pools may lead to robust ensembles, considering both the different levels of complexity and imbalance [8].

In this paper, adaptive skew-sensitive classifier ensembles are proposed for video surveillance applications. The proposed approach allows to select training data with varying levels of imbalance and complexity to design ensembles of classifiers that provide enhanced accuracy and robustness. Face captures of each person in the scene are tracked and regrouped into trajectories, and a decision threshold is applied to the accumulation of positive predictions from base classifiers for robust spatio-temporal recognition. During enrollment, facial captures from a reference trajectory are combined with selected non-target captures from the universal and cohort models[2] to generate a pool of 2-class classifiers using data with various levels of imbalance and complexity (class overlap and dispersion). Training/validation sets with different imbalances and complexities are built through random undersampling, and cover a range of imbalances from 1:1 to a maximum imbalanced estimated according to experience $1:\lambda^{max}$. During operations, the operational level of imbalance is periodically estimated from the input data stream using the HDx quantification method, and pre-computed histogram representations of imbalanced data distributions. The HDx quantification allows one to estimate the prior probability of operational data based on the Hellinger distance between histogram representations of class distributions in the feature space, and employ a single validation set that is not required to provide a specific imbalance [12]. Finally, the proposed approach allows one to adapt pre-computed histograms and ensemble fusion functions based on the imbalance and complexity of operational data.

The proposed approach is validated with synthetic and video data, and compared against reference adaptive ensembles using BC, meta-kNN fusion and score-level average fusion. The synthetic problem was designed to observe the impact of different theoretical probabilities of error as well as distinct imbalance levels in the performance of the system (Gaussian distributions in a two-dimensional feature space). The Carnegie Mellon University Face In Action (FIA) video database was used to emulate face re-identification applications. The transaction-based performance evaluates face matching of the system using the ROC and precision–recall spaces, and individual-specific characterization allows one to analyze specific cases. Finally, trajectory-based analysis is employed to show the overall system performance over time.

The rest of this paper is organized as follows. Section 2 presents a brief review of techniques for ensemble design (generation, selection and fusion) techniques, and specifically ensemble techniques proposed to address the problem of class imbalance. Section 3 describes the adaptive skew-sensitive ensembles proposed for FR in imbalanced environments. Section 4 provides synthetic experiments that motivated the proposed approach. Section 5 presents the experimental methodology and results with the FIA video data for validation of the proposed approach in face re-identification applications.

## 2. Ensemble methods for class imbalance

Ensemble-learning techniques combine classifiers with diversity of opinions to increase classification performance. The design process can be divided into three main steps – generation of a pool of base classifiers, selection and fusion of classifiers [13–16]. The first step allows one to train base classifiers with diversity of opinions, and the last two take advantage of this diversity to produce more accurate predictions. Diversity can be created by employing distinct classifiers, train distinct instances of a classifier with different initial conditions (parameters), or using different training sets [14].

Representative examples of ensemble methods are bagging, boosting, random subspaces, which employ different training sets of data or features from the training set to build distinct base classifiers [14,17]. An example of diversity generation by various parameters is the work of Connolly et al. [18], which takes advantage of diversity in the hyperparameter space of classifiers to produce useful diversity of opinions. Examples of selection strategies are greedy search, clustering-based methods and ranking-based methods, and examples

[2] In this paper, a universal model (UM) is defined as a database containing ROI patterns from selected unknown people appearing in scene, and the cohort model (CM) is defined as a database with ROI patterns from other target individuals enrolled to the system.