Università degli Studi di Cagliari

PHD DEGREE

Electronic and Computer Engineering

Cycle XXXIII

TITLE OF THE PHD THESIS

Template update algorithms and their application to face recognition systems in the deep learning era

Scientific Disciplinary Sector

ING-INF/05

PhD Student: Giulia Orrù

Supervisor prof. Gian Luca Marcialis

Final exam. Academic Year 2019 – 2020
Thesis defence: February 2021 Session

UNIVERSITY OF CAGLIARI

Abstract

Faculty of Engineering and Architecture
Dept. of Electrical and Electronic Engineering

Template update algorithms and their application to face recognition systems in the deep learning era

by Giulia ORRÙ

Biometric technologies and facial recognition systems are reaching a very high diffusion for authentication in personal devices and public and private security systems, thanks to their intrinsic reliability and user-friendliness. However, although deep learning-based facial features reached a significant level of compactness and expressive power, the facial recognition performance still suffers from intra-class variations such as ageing, different facial expressions, different poses and lighting changes.

In the last decade, several "adaptive" biometric systems have been proposed to deal with this problem. Unfortunately, adaptive methods usually lead to a growth of the system in terms of memory and computational complexity and involve the risk of inserting impostors among the templates. The first goal of this PhD thesis is the presentation of a novel template-based self-update algorithm, able to keep over time the expressive power of a limited set of templates. This classification-selection approach overcomes the problem of manual updating and stringent computational requirements. In the second part of the thesis, we analyzed if and to what extent this "optimized" self-updating strategy improves the facial recognition performance, especially in application contexts where the facial biometric trait undergoes great changes due to the passage of time. In contexts of long-term use, in fact, the high representativeness of the deep features may not be enough and this is usually overcome with a re-enrollment phase. For this reason, one of our goals was to evaluate how much an automatic template updating system could compete with human-in-the-loop in terms of performance.

To simulate situations of long-term use in which the temporal variability of biometric data is high, we acquired a new dataset collected by using frames of some videos in YouTube related to Daily Photo Projects: people take a picture every day for a certain period of time, usually to show how their appearance is changing. The temporal information present in this new dataset allowed us to evaluate how long a facial feature can remain representative depending on the context and the recognition system.

Extensive experiments on different datasets and using different facial features are conducted to define the contexts of applicability and the usefulness of adaptive systems in the deep learning era.

Acknowledgements

I would like to thank my supervisor, prof. Gian Luca Marcialis, for his precious advice and support, and all the colleagues of the PraLab, in particular, prof. Fabio Roli, for his guidance and insights.

I am also very grateful to prof. Julian Fierrez and his Biometric Recognition Group (ATVS) and to Dr. Sébastien Marcel and the Idiap Research Institute for giving me the opportunity of doing two great internships.

I also would like to thank prof. Fernando Alonso-Fernandez and prof. Martin Drahansky for their reviews of this thesis.

Contents

List of Figures

List of Tables

Dedicated to my family

Chapter 1

Introduction

In the last few years, facial recognition (FR) systems have spread enormously in daily life, having been integrated into many mobile devices and public and private security systems. The deep learning and CNNs era has made possible to achieve a close-to-zero error rate on most state-of-the-art facial datasets. An extensive set of templates that cover all the "intra-class" variations in the facial appearance, such as poses, illuminations, expressions and style changes [1–3], is needed to obtain a performing FR system. In fact, although a pre-processing phase partially manages to follow the variations in pose and illumination, the variations and the ageing of the biometric traits remain an open problem, especially in systems with small storage spaces [4, 5].

To address these limitations, several "adaptive" biometric systems have been proposed as an alternative to a periodic re-enrolment phase [6–8]. Adaptive strategies, also called self-update methods, are based on the detection of novel templates to replace or update the template gallery without the need of the human-in-the-loop. Except for the first set of templates collected supervised during the enrolment phase, update data is collected unsupervised during system operation. Adaptive biometric systems may be affected by an increase in computational complexity due to the continuous insertion of novel templates into the gallery. Therefore, they do not fully meet the stringent requirements of devices and applications with small storage and computing resources. Moreover, adaptive algorithms may led to the introduction of impostors in the users' gallery, opening a severe breach from the security point of view.

1.1 Contributions

The first goal of my PhD research was to address the facial-based self-update problem by considering a very limited space for storing the templates per client. We believe that, besides the advantage of meeting the stringent hardware requirements of mobile devices, the limited number of templates allow to reduce the introduction of impostors into the system. We hypothesise that intra-class variations and inter-class variations, although large, make that feature space smoothly partitioned according to each user. In other words, the whole features space may be hypothesised as mono-modal when dealing with samples of the same user and multi-modal when different users are taken into account [9]. As a matter of fact, if the neutral expression characterises initial templates at a controlled lighting condition, adaptive algorithms tend to attract similar faces from the same users, drifting to other ones when expressions or environmental conditions are too far from the initial ones [10, 11].

We introduced two primary methodologies [9]: in the first one, we pushed our hypothesis at its extreme by adopting a clustering algorithm to identify the distribution's "centroid" for each user. Then, the nearest samples of that centroid are selected and stored in the user's gallery. In the second methodology, the best templates are selected according to a specific criterion to "optimise". In particular, the editing algorithms used for supervised template selection are exploited and adapted to the task [12, 13]. Both methodologies rely on the classification-selection paradigm proposed in [3]. Furthermore, we compared these two "optimised" selections with a random selection simulating a human supervisor. We evaluated the performance of these "optimised" template update methods on three public datasets using representations of the face extracted through handcrafted methods, such as BSIF [14], and the most well-known and powerful neural networks for face recognition, such as FaceNet [15], ResNet50 [16] and SeNet50 [17].

However, the datasets used for this evaluation and most of the public face datasets were collected over a short time span (from one to three years) or do not contain temporal information on the data.

For this reason, it is difficult to assess the performance of facial recognition systems (with or without adaptivity) when the variability of the biometric trait is high due to the passage of time. With the aim of simulating situations of use of a facial biometric system in the long term, we collected a dataset using the frames of YouTube videos related to Daily Photo Projects [18]. The temporal information and the long acquisition periods of this new dataset have allowed us to evaluate the performance of the compact and powerful representation of the facial deep-learning features across a large time lapse between enrolment and test [9, 19].

Chapter 2 of this thesis gives an overview of related work for face recognition and adaptive systems. Chapter 3 presents a novel classification-selection approach to update facial recognition systems and analyses its performance across three publicly available datasets. In Chapter 4 a new dataset, generated explicitly conceived to embed intra-class variations of users on a large time span, is presented. Chapter 5 concludes this manuscript by providing a summary of the major contributions.

Chapter 2

Biometric technologies and adaptive systems

Biometric recognition refers to the measurements of human biological and behavioural characteristics to automatically recognise or authenticate a person [20]. Any human characteristics which meet the requirements of *universality*, *uniqueness*, *permanence* and *collectability* [21] can be used as a biometric identifier. However, some biometric traits are better suited to be used to authenticate or recognise an individual. Among the physiological, the most common characteristics are the fingerprints, the hand geometry, the iris and the face; among the behavioural, we often find the signature, the voice, the pressure style of keyboard keys and the gait.

A biometric system is a pattern recognition system that operates by acquiring the user's biometric traits, extracting the feature set and comparing this set with the templates' features stored in a database.

Biometric systems can operate in two modes: *authentication* or *identification* [22](Fig.2.1).

- In authentication mode, the system verifies the user's identity by comparing the acquired biometric trait with the corresponding reference, called template, stored in the database. In this mode, the user declares his identity with a PIN, a smart card, a user name, etc..

- In identification mode, the system compares the biometric data in input with all the templates in the database in order to establish the user's identity.

Both operating modes are preceded by the *enrolment* phase in which a person's biometric data is captured and stored in the database. The stored data, the templates, are associated with the user's identity through an identification number or string.
In general, a biometric system consists of three or four modules:

- Acquisition module: a sensor captures the biometric trait.

- Pre-processing: an optional module used to enhance the quality of the raw data in order to extract more robust features over different acquisition scenarios.

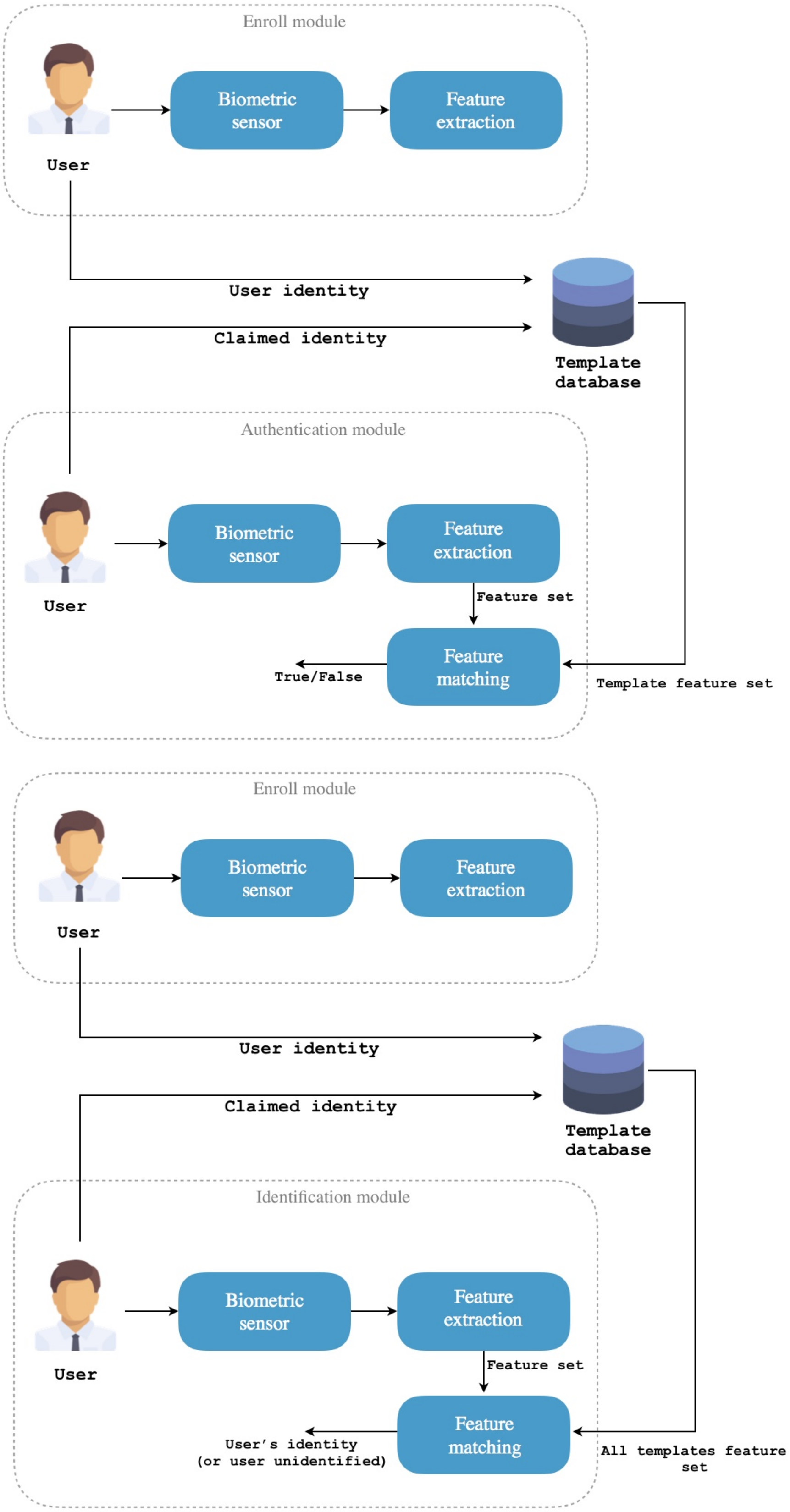


FIGURE 2.1: A biometric system can operate in two modes: authentication or identification. Depending on the mode, stored templates are matched against the templates of the declared user in authentication or against all the templates in the database in identification. A biometric system consists of four components: a sensor, a feature extractor, a matcher, and a database, containing the templates.

- Feature extraction: the acquired data are processed to extract important and discriminative information.
- Matching or classification: the extracted feature is compared with those present in the database to generate a matching score which represents the similarity between the references acquired in the enrolment phase and the input query. Based on the scores, the system predicts an identity.

2.1 Face recognition

The face is an essential biometric trait for authentication and it has been widely used in many security applications, being non-intrusive, intuitive and with a high level of uniqueness.

A three-stage process [23] can schematise a facial recognition system:

- Face detection: starting from a single image or video, a face detector should be able to identify and locate all the faces present regardless of their position, scale, pose, expression and lighting conditions [24]. Many techniques have been proposed to detect and locate the face, for example, the Viola–Jones detector [25], methods based on histogram of oriented gradient (HOG) [26] and convolutional networks [27].
- Feature extraction: extracts relevant information from the input faces encodes the identity information.
- Face recognition: consists of a comparison method and a classification algorithm and performs matching of the input face against one or more reference faces in the database.

In the last decades, several facial recognition algorithms have been developed [24]. These can be grouped into four significant technical flows (Table 2.1) that have taken place over the last thirty years [28]. In the 1990s and 2000s, the holistic approaches, also called appearance-based methods or global approaches, dominated the FR research. These approaches use the entire face as input of the recognition system, projecting it in a subspace of reduced dimensionality. Based on the technique used to sub-project the areas of the face, the holistic approaches can be divided into two subcategories, linear and non-linear [29]. Some examples of linear methods are Eigenface, based on the PCA (Principal Component Analysis) technique, and LDA [30, 31] based on the construction of a discriminant subspace to distinguish the faces. Non-linear methods include the Support Vector Machine (SVM) [32], based on the idea of finding a hyperplane that best divides a data set into two classes, and the Kernel Principal Component Analysis (KPCA) [33], a non-linear reformulation of the PCA. The main issue of holistic approaches is that these methods fail to address uncontrolled facial changes. In particular, they are ineffective in unconstrained conditions due to lack of robustness to lighting, pose, expression and image quality variations. To solve this prob-

Represent.	Holistic learning	Local handcraft	Shallow learning	Deep learning
Years	1990s	2000s	2010s	2012/2014
Issues	Fail to address uncontrolled facial changes.	Lack of distinctiveness and compactness.	Fails to address facial appearance variations.	Black-box, time-consuming, requires powerful hardware and a large number of images for training.

TABLE 2.1: Feature representation evolution in FR systems.

lem, in the early 2000s, local features were proposed to select discriminative features to identify individuals uniquely. These methods called local handcrafted methods or feature-based methods, are based on different types of features: some methods, called interest-point based, detect points of interest and extract features localized on these points, others, called local appearance-based, divide the input image into small regions (or patches). Some well-known local handcrafted approaches are Local Binary Patterns (LBPs) [34, 35], Gabor [36], BSIF [14], and their variants [37]. Unfortunately, these approaches exhibit as a limitation the lack of compactness and in some cases a low distinctiveness among individuals.

The introduction of shallow learning methods, in the early 2010s, allowed to face the problem of the low compactness level. These methods, also known as learning-based local descriptors, use unsupervised learning methods to encode the local micro-structures of the face into discriminative codes [38]. Although shallow learning methods achieve accuracy greater than 90%, they cannot handle the complex nonlinear facial appearance variations such as self-occlusion and variations in pose and illumination [39].

After 2012, the massive success of deep learning methods and convolutional neural networks allowed to achieve state-of-the-art results in many CV problems. Deep learning is based on the unsupervised learning of representations of data with multiple levels of abstraction [40]. From 2014 also the FR community adopted deep learning methods for face feature extraction and transformation. Among others, DeepFace [41], FaceNet [15], VGGFace [42] and VGGFace2 [43]. Some of them achieved state-of-the-art performance on the most challenging datasets known, such as LFW [44], IJB-A [45], IJB-B [46], etc.

The types of neural networks used for the facial recognition task are many. The most famous and frequent are Convolutional Neural Networks (CNN), but Generative Adversarial Networks (GAN), Deep Belief Networks (DBN) and Deep Boltzmann Machines (DBM) are also used [47, 48].

In recent years, the high performance achieved thanks to the era of deep learning, has allowed the scientific community of FR to investigate other important related issues such as algorithmic discrimination due to the high bias introduced during the training of deep models [49], the improvement of presentation attack detection performance [50], the introduction of the concept of “explainability” in such systems [51], and many others.

2.2 Adaptive biometric systems

Face recognition systems are sensitive to particular variations of the facial appearance due to changes in environmental conditions such as lighting changes, to individual physiological changes (ageing, scars, etc.) and to variations in the interaction between sensor and individual, such as the user's pose [3, 7]. For this reason, the templates acquired during the enrolment phase might become unrepresentative of the biometric data input.
Periodic supervised re-acquisitions partially overcome this problem. Moreover, it is almost impossible, for a human supervisor, to infer which template is more representative if several intra-class variations must be taken into account and the re-acquisition is invasive, expensive and time-consuming. Adaptive systems have been proposed as a solution to this problem with the aim to automatically update templates and adapt the facial model.
The adaptive biometric systems are based on a semi-supervised template updating, that exploits, in addition to labelled data acquired during the initial enrolment phase, unlabelled data, namely images or features sets acquired during system operation. In particular, the face image captured when a user requires the identity verification, is saved to evaluate if it can be added to the template gallery after processing and verification stage. The evaluation is based on criteria that vary according to the updating algorithm. However, all adaptive systems aim to ensure that the captured image belongs to the claimed user with a high probability and that does not bring redundancy of the information represented by the templates. It is easy to note that this approach is less expensive and invasive than a typical re-enrolment session [7]. In general, a simple addition of the "novel" template to the user's gallery would lead to the increase of the system complexity in terms of processing time and amount of memory required for storing data. Therefore, adaptive systems need filtering techniques of redundant information [12, 52] (eg., automatic replacement, pruning, etc.). This also holds for the standard supervised updating approach [12].

2.3 Related works

Adaptive biometric systems have become popular because of their ability to adapt to new input data. The adaptive attitude of a system is essential to have a good performance, especially to make it robust to changes.

We introduce here a common notation in order to explain the rest of this manuscript. The initial set of templates of k users stored in the gallery is named *gallery* $GT = \{t_{1,1}, t_{1,2}, ..., t_{k,1}, ..., t_{k,N}\}$, where N is the number of templates per user. For the sake of simplicity, we consider the same number of templates per user. Let $U = \{U_1, ..., U_n\}$ be the so-called *batches* [53], the set of samples used to simulate an update iteration and n the number of iterations. U contains n sets of faces collected during operation of the system, one for each iteration.

According to the terminology above, the primary ratio of a traditional self-update algorithm is described by Algorithm 1. Self-update estimates the threshold t^* through GT and, when the batch U_i is available for a certain claimed user, the system compute the distances between each U_i's sample and the user's template(s). Only the batch input samples that respect the threshold t^* are added into the user's gallery.

The traditional self-update is based on the semi-supervised learning theory [7], since the initial set of labelled templates and a set of unlabelled data obtained during the use of the biometric system are exploited [7]. Normally, self-update do not only rely on the addition of new templates to the gallery, but, where present, they re-set the system parameters, which may be represented by a classifier or a feature extraction algorithm (e.g. changing the weights of a neural network or some other kind of intermediate feature).

Algorithm 1 Traditional self-update algorithm

```
procedure TRADITIONAL SELF-UPDATE SYSTEM
    Let $GT$ be the intial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the $n$ batches of unlabeled samples
    Estimate the update threshold $t^*$ using $GT$
    for $i = 1$ to $n$ do
        for $e$ in $U_i$ do
            if $distance(e, GT) < t^*$ then
                $GT_{new} = GT \cup e$
            end if
        end for
        $GT = GT_{new}$
        Estimate the update threshold $t^*$ using $GT$
    end for
end procedure
```

The first adaptive algorithms for face recognition [54, 55] were based on the application of the Principal Component Analysis (PCA) [30]. PCA reduces the space of features by projecting the original space into a new one. The new space is composed of a set of orthogonal bases, called eigenvectors. In [10], a semi-supervised PCA version is reported: the PCA is applied for each batch. The semi-supervised PCA (Algorithm 2) is an off-line self-training method that updates the templates and the eigenspace of a PCA-based face recognition system incrementally. In the implementation of the PCA-based adaptive system, we selected the eigenvectors number such that the cumulative variance is constant, calculated as the sum of the most significant feature variance, equal to 80%. The i-th feature variance is calculated as

$$\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \tag{2.1}$$

where λ_i corresponds to the i-th eigenvalue (the eigenvalues represent the variance along with the principal components).

Algorithm 2 Self-update with semi-supervised PCA

```
procedure SEMI-SUPERVISED PCA
    Let $GT$ be the intial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the n batches of unlabeled samples
    Compute the matrix $W$ of the principal components using $GT$
    Project $GT$ to the eigenspace using the principal components matrix $W$
    Estimate the update threshold $t^*$ using $GT$
    Project $b_i$ to the eigenspace using the principal components matrix $W$
    for $i = 1$ to $n$ do
        for $e$ in $U_i$ do
            if $distance(e, GT) < t^*$ then
                $GT_{new} = GT \cup e$
            end if
        end for
        $GT = GT_{new}$
        Estimate the update threshold $t^*$ using $GT$
        Compute the matrix $W$ of the principal components using $GT_{new}$
        Project $GT_{new}$ to the eigenspace using the principal components matrix W
        Project $b_i$ to the eigenspace using the principal components matrix W
    end for
end procedure
```

Traditional adaptive biometric systems can not distinguish between samples that contain redundant information. For this reason, some methods try to filter the redundant information in order to mitigate the growth of the facial recognition system. One of these methods is the "context sensitive" method [52] which combines a traditional self-update method with a change detection module. The authors introduce an approach based on inserting a new template only if it presents new information. The context sensitive method, described in Algorithm 3, is based on the detection of changes in ROI illumination conditions. For this reason, a global luminance quality (GLQ) index is calculated for each input data and each template of the gallery.

Algorithm 3 Context sensitive self-update

```
procedure CONTEXT SENSITIVE SELF-UPDATE SYSTEM
    Let $GT$ be the intial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the $n$ batches of unlabeled samples
    Let $R$ be the intial ROIs template gallery
    Let $GLQ(x, y)$ be the function to compute the global luminance quality
    Estimate the update threshold $t^*$ using $GT$
    Estimate the capture condition threshold $t_c$ using $GLQ(GT, GT)$
    for $i = 1$ to $n$ do
        for $e$ in $U_i$ do
            if $distance(e, GT) < t^*$ then
                if $\sum GLQ(e, R_i) \geq t_c$ then
                    $GT_{new} = GT \cup e$
                end if
            end if
        end for
        $GT = GT_{new}$
        Estimate the update threshold $t^*$ using $GT$
    end for
end procedure
```

GLQ is a quality index based on image correlation, luminance, distortion and contrast distortion. A sufficiently high GLQ value indicates that the sample has new capture conditions in comparison to the stored templates, which justifies the increase in complexity of the system. Therefore, this context sensitive method allows inserting genuine templates that present many intra-class variants to update the gallery.

Other adaptive biometric methods aim to filter redundant information by a two-staged approach, classification and selection. In the first stage, the input samples are pseudo-labelled based on the probability of belonging to a given class. The second stage selects the best samples to update the system gallery. Ref. [11] (algorithm 4) proposes a two staged classification-selection approach based on harmonic function and risk minimization technique. In particular, in the classification stage, soft probabilistic labels are assigned to input data by calculating the adjacency matrix and by finding the harmonic function on the graphic representation of the samples. The calculation of the harmonic function is based on the computation of the adjacency matrix as match-score, in particular by the use of the Laplacian matrix. The harmonic function is used as minimum energy function and is interpreted as the posterior probability that a sample belongs to a genuine user. The second stage allows selecting, among the genuinely classified samples, those that have more information based on the risk minimization criteria and Bayesian risk theory. The risk minimization method was initially proposed to update a fingerprint recognition system. In this thesis experiments, we apply this method to face recognition.

Table 2.2 shows an overview of the methods described above. In summary, it is possible to categorise adaptive biometric systems into two types: a traditional update approach, drawn in Fig. 2.2a, that only classifies input

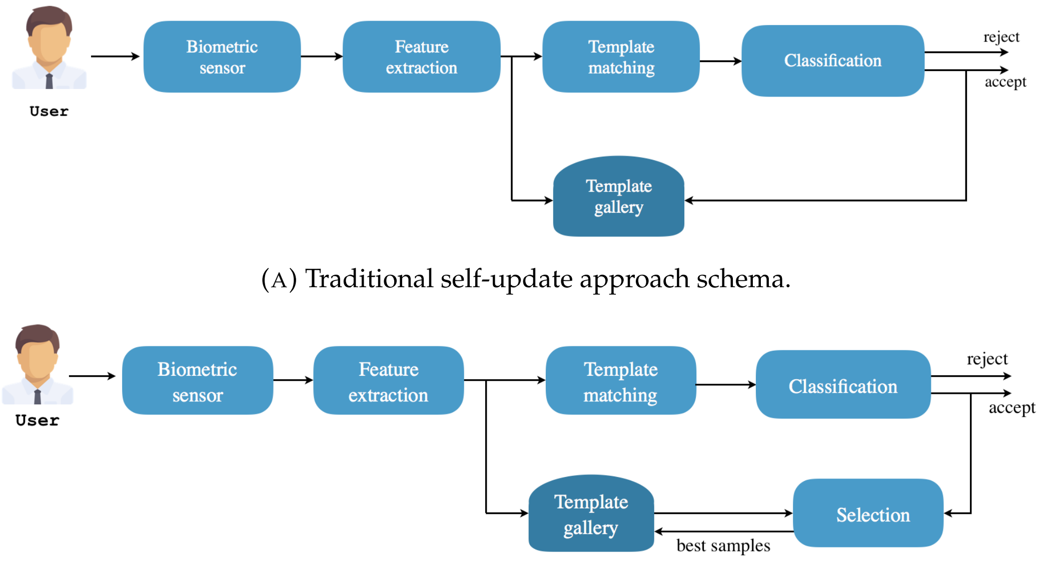
Algorithm 4 Risk minimization self-update

```
procedure RATTANI ALGORITHM
    Let $GT$ be the intial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the $n$ batches of unlabeled samples
    for $i = 1$ to $n$ do
        $N \leftarrow \{GT \cup u_i\}$
        Calculate the adjacency matrix $W$
        Calculate Laplacian matrix
        Compute the harmonic function $f_u$
        $P = \emptyset$
        for $x_j \in u_i$ do
            if $f_u(j) > 0.5$ then
                $P = P \cup x_j$
            end if
        end for
        Estimate the risk $R(f) = \sum_{i=1}^{n} min(f_i, 1 - f_i) \forall n \in N$
        while (risk is minimum) do
            Find $x_k \in P$ using $k = argmin^k(R(f^{+(x_k)})$
            Add $x_k$ to $GT$
            Remove $x_k$ to $P$
        end while
    end for
end procedure
```

samples and adds to the gallery if they meet the genuinity requirements and a two-step approach, drawn in Fig. 2.2b, also called classification/selection. The most relevant problem for methods that do not include a selection phase is the uncontrolled growth of complexity and galleries size. Actually, even for methods with a selection phase that do not set a limit on the number of templates per user this growth in complexity and size is present, thus making unrealistic the real application of adaptive biometric systems. To this aim, we propose two approaches described in the next chapter, which keep constant the number of templates per user. This gallery size limit allows controlling the computational complexity of the system in order to realistically evaluate the performance when the adaptation ability is added under limited computing and memory resources.

Self-update Method	Ref	Year	Description	Selection phase
Traditional	[3, 7]	∼ 2004	addition of new templates to the gallery and re-set of the system parameters	
Semi-supervised PCA	[10]	2006	based on space reduction through PCA	
Risk-minimization	[11]	2012	based on harmonic function and risk minimization technique	x
Context sensitive	[52]	2015	based on the detection of changes in ROI illumination conditions	x

TABLE 2.2: Overview of state-of-the-art self-update methods. Methods without a selection phase have an uncontrolled growth of complexity and galleries size. Methods with selection phase, without a limit in the number of samples per user, albeit in a more controlled manner, still have a growth of complexity and galleries size.



(A) Traditional self-update approach schema.

(B) Classification/selection approach schema

FIGURE 2.2: Scheme that highlights the different functioning of the classification-selection approach compared to the traditional self-update.

Chapter 3

A novel classification-selection approach for adaptive face recognition systems

In real applications, the use of traditional adaptive methods leads to an increase of the gallery size and the system complexity, that is, the processing time and the amount of data stored in the system.
Additional templates are added to the user's gallery, which makes the adaptivity ability not suitable for mobile devices and IoT applications. Furthermore, misclassified samples could be added, as well. In order to overcome these problems, new techniques to filter redundant information and to select the most significant templates for updating are needed. In this thesis, we present an "optimised" classification-selection approach that keeps the number of templates constant at each iteration, so that this systems can be used in applications with small storage and computing resources. Our approach

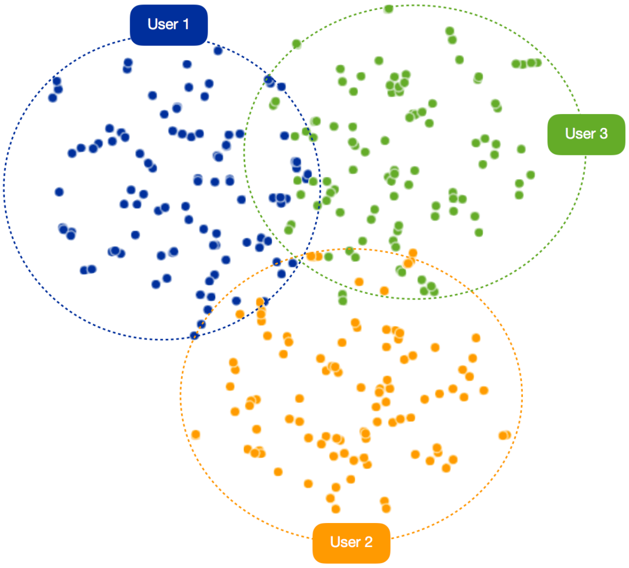


FIGURE 3.1: 2D graphic representation of the sample distribution on a data set of 3 users under the hypothesis presented. Each cluster is associated with a user. The clusters are partially overlapping.

is based on the hypothesis that the distribution of the facial features is mono-modal for samples of the same user and multi-modal for samples of different users. In other words, we can identify clusters, partially overlapped (Fig. 3.1), where borderline features belong to changes in the subject appearance (intra-class variations). This mono-modal behaviour, assuming that a face is represented by a feature space x sampled from the appearance of k enrolled subjects, can be formulated as follows:

$$p(x) = \sum_{i}^{k} p(x \mid l(x) = i) P(l(x) = i) \tag{3.1}$$

Where $l(x)$ is a labelling function such that $l(x) \in \{1,...,k\}$. In other words, starting from the theorem of total probability, the probability of x is the sum of the conditional probabilities $p(x|l(x) = i)$ weighted by the prior probability $P(l(x) = i)$. However, to be aware of all possible samples' labels *a priori* may be very difficult. For this reason we assume that $p(x)$ depends on an analogous number k of possible clusters of *unlabelled* samples $CL = \{CL_1, ...CL_k\}$:

$$p(x) = \sum_{i}^{k} p(x \mid x \in CL_i) P(x \in CL_i) \tag{3.2}$$

Fig. 3.1 shows what we mean. The depicted circles are the possible projections of $p(x \mid x \in CL_i)$, which are overlapped in some regions. This can be further modelled by the contribution of all known subjects. In other words, each CL_i can be seen as a set of k subsets $CL_i = \{CL_{i1}, \dots, CL_{ik}\}$, where each sample in CL_{ij} is labelled as the subject j, that is, $\forall x \in C_{ij}, l(x) = j$. Accordingly:

$$p(x \mid x \in CL_i) = \sum_{j}^{k} p(x \mid x \in CL_i, x \in CL_{ij}) P(x \in CL_{ij} \mid x \in CL_i) \tag{3.3}$$

Being $CL_{ij} \subseteq CL_i$, we have $P(x \in CL_{ij} \mid x \in CL_i) = P(x \in CL_{ij})$. If we hypothesise a large majority of samples of the subject i falling in CL_i, we have that $P(x \in CL_{ii})$ is much more than any other $P(x \in CL_{ij})$; Eq. 3.3 can be rewritten as:

$$p(x \mid x \in CL_i) \approx p(x \mid x \in CL_{ii}) \tag{3.4}$$

Each cluster concurring to the generation of $p(x)$ is dominated by the mode $p(x \mid x \in CL_{ii})$.

On the basis of this modelling, the individual contribution of user i to the whole feature space is given by $\cup_j CL_{ji}$, which corresponds to:

$$p(x \mid l(x) = i) = \sum_{j}^{k} p(x \mid x \in CL_{ji}) P(x \in CL_{ji}) P(x \in CL_j) \tag{3.5}$$

Consequently, templates should be selected from $\cup_j CL_{ji}$. We introduced two criteria to detect and select the templates "optimally".

In the first approach, based on clustering methods, we assume that $p(x \mid x \in CL_i)$ is further characterised by one centroid called c_i which approximates the centroid of CL_{ii} according to Eq. 3.4. The second approach is based on the patterns located over the region characterised by the probability in Eq. 3.5, i.e. in $\cup_j CL_{ji}$.

In both the "optimised" selection approaches, each partition $\cup_j CL_{ji}$ is estimated by the pseudo-labelling step used in the traditional self-update method. The core of the novel classification-selection adaptive method proposed consists in the template selection phase. The description of the two different criteria are reported in Sections 3.1-3.2.

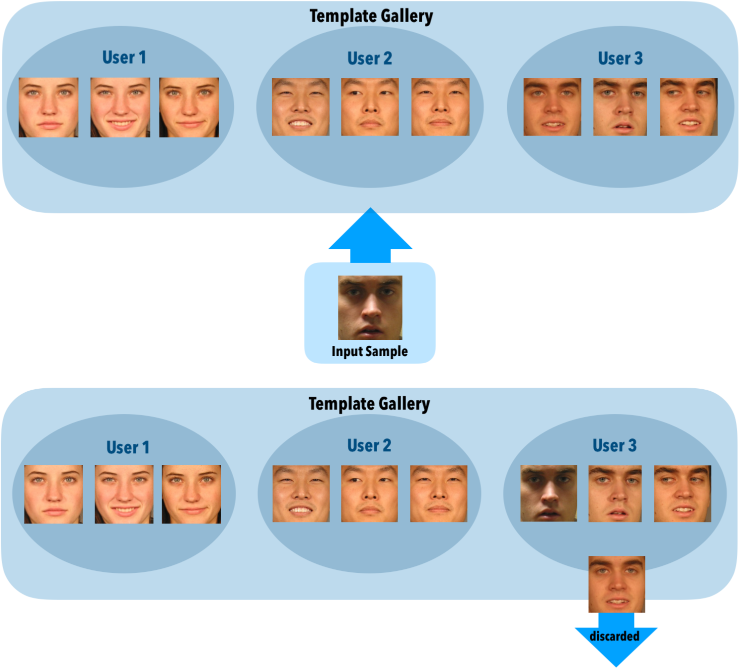


FIGURE 3.2: Example of gallery update on a system with three templates per user limit.

3.1 Classification-selection by clustering method

The first method is based on the detection of the "centroid" of each user's distribution as expected from the mono-modal appearance of $p(x \mid x \in CL_i)$ in Eq. 3.2. During the first step the input faces are classified by using the updating threshold t^* and a pseudo-label is assigned to each sample. In the second step, for each user, the algorithm selects, from the associated cluster, the p closest samples to the centroid. However, we do not consider only labelled and pseudo-labelled samples associated with the i user for identifying the related cluster and, thus, its centroid. The appearance depicted by Eq. 3.2 and Fig. 3.1 suggests that, if any, the centroid may be elsewhere in the feature space. As a matter of fact, the available templates could be "far" from the centroid, due to temporary and temporal variations in the input data.

Accordingly, if we consider only samples pseudo-labelled through the available templates, we could find a bad estimated centroid. Therefore, we chose to use the K-Means algorithm that is based on the search for natural clusters and works at the feature level applied to all labelled and pseudo-labelled samples available.

In other words, assuming that the cluster CL_i is characterised by a centroid c_i, if the generating function of facial samples is Gaussian for each subject, the centroid corresponds to $p(c_i \mid c_i \in CL_i) > p(x \mid x \in CL_i), \forall x \neq c_i$. By moving away from c_i, the distribution gradient is gradually negative and the points near to c_i are geometrically close to each other, as well as their probability of occurrence is high. Following the clustering-based model, it is highly likely that user i templates lie around c_i, since they are representative of CL_{ii}, that is, the main mode of CL_i.

The p most "representative" samples per user are selected around each centroid (Fig. 3.3). This working hypothesis is based entirely on the mono-modality reported in Eq. 3.2.

If the mono-modality hypothesis is not satisfied, two problems may arise:

1. A user may be associated with more than one cluster. A selection criterion for the most representative cluster should be chosen. In this case, some users could not have a main reference mode, and this leads to the impossibility to update their templates.

2. A cluster may be associated with more than one user. A selection criterion is needed for the users which the cluster must be referred to.

In order to verify at which extent this hypothesis is correct, we simulated different acquisition conditions to evaluate how feature distribution changes. We reported this evaluation in the Sec. 3.3.3.

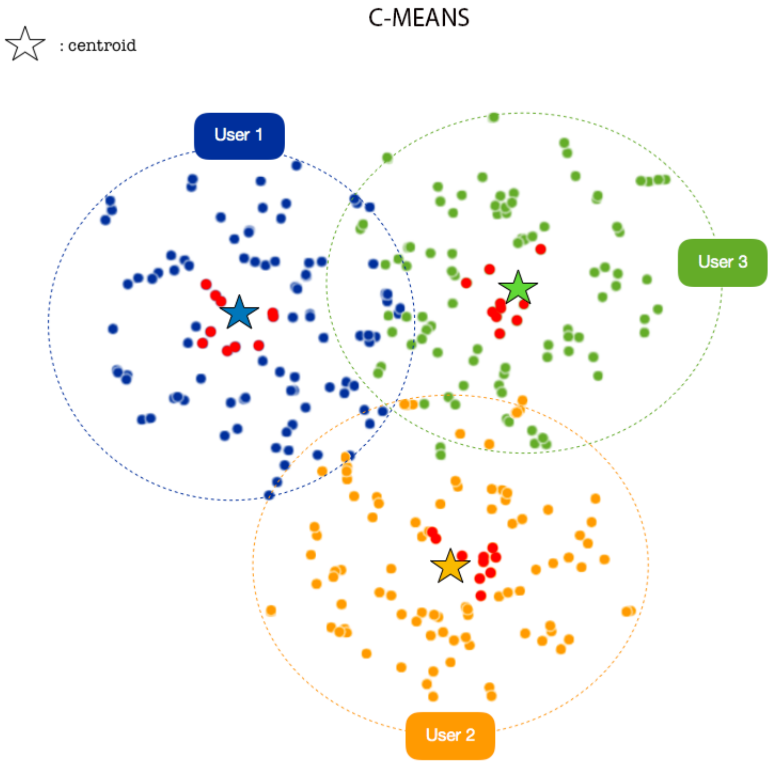


FIGURE 3.3: 2D graphical representation of the application of the K-Means algorithm on a dataset of 3 users. The stars symbolize cluster centers. The samples marked in red are those selected by the algorithm.

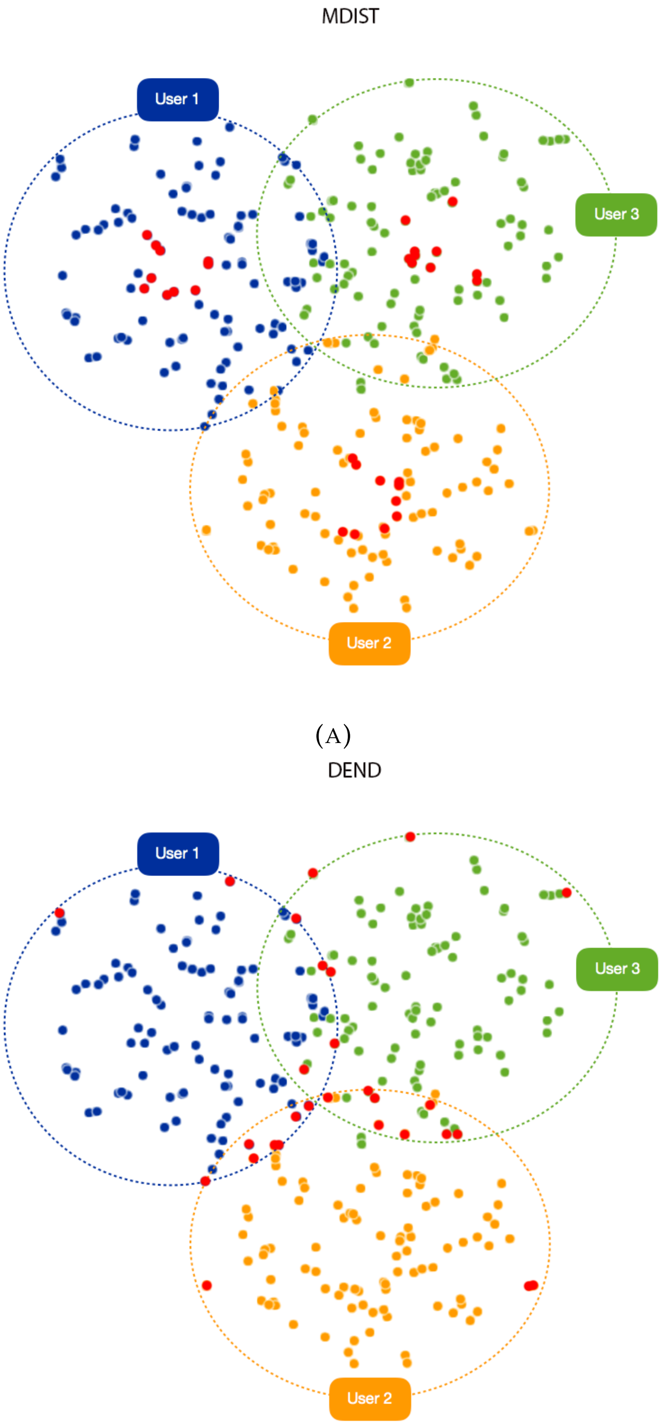
Algorithm 5 K-Means self-update

```
procedure PROPOSED SELF-UPDATE SYSTEM
    Let $GT$ be the initial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the $n$ batches of unlabelled samples
First stage
    estimate the update threshold $t^*$ using $GT$
    for $i = 1$ to $n$ do
        for $e$ in $u_i$ do
            if $distance(e, GT) < t^*$ then
                $GT_{new} = GT \cup e$
            end if
        end for
Second stage
        generate $k$ clusters by K-Means algorithm, where $C = k$ being $k$ the number of enrolled users
        for $i$ in $[1, .., l]$ do
            let $c_i$ be the $i - th$ cluster generated by K-Means
            let $u_j$ the user with the highest number of samples in $c_i$
            select the $p$ nearest samples to the centroid of $c_i$ and update the gallery of $u_j$ accordingly
        end for
        $GT = GT_{new}$
        estimate the update threshold $t^*$ using $GT$
    end for
end procedure
```

3.2 Classification-selection by editing algorithms

The second approach is based on the semi-supervised adaptation of editing algorithms adopted previously for supervised template selection.

In particular, MDIST and DEND algorithms were used because of their complementary [12]. Again, the system is two-staged: during the first phase, the input samples are pseudo-labelled by using the updating threshold t^*. During the second phase, the Euclidean distances of the feature vectors of all samples are calculated and p templates are selected. The MDIST algorithm selects the p templates averagely close in the features space (Fig. 3.4a), that is, facial images with small intra-class variations; the DEND algorithm selects the p templates with the biggest average distance (Fig. 3.4b), that is, facial images with relevant intra-class variations. Due to its definition, the DEND algorithm could lead to the insertion of a high number of impostors when selects samples from overlapping areas. On the contrary, MDIST selects a gallery that does not cover a large number of variations. This behaviour is exemplified in Figs. 3.4a-3.4b. We, therefore, expect that they have different performance depending on the environmental conditions. Probably, MDIST may work better in a more controlled environment where inputs are not much different from templates; DEND may work better on data presenting high intra-class variations. However, the overlap among cluster should



(A)

(B)

FIGURE 3.4: 2D graphical representation of the application of the MDIST and DEND algorithms on a dataset of 3 users. The samples marked in red are those selected by the algorithm.

be low, and this may depend on the features set adopted beside the intrinsic variations in the subjects' appearance due to temporal and temporary causes.

This approach relaxes the previous hypothesis of the Gaussian generating function. The MDIST method allows finding the p samples of $\cup_j CL_{ji}$ that are, averagely, the closest each other for the user i. This solves the possible centroid estimation errors due to lack of data, as we search in a space larger than that of CL_{ii}.

What we may expect from this approach is basically dependent on the effectiveness of the pseudo-labelling stage, on the features set adopted, on the temporal and temporary variations captured during system's operations and on the initial templates. However, the maximum number of possible templates, p, should limit the probability of impostors insertion, especially when using the DEND algorithm.

In addition to the MNIST and DEND methods, other selection methods can be used, such as the RANDOM method, which selects p random samples simulating what may happen by keeping a human in the loop to perform a periodic template update.

Algorithm 6 Self-update with editing algorithm (semi-supervised)

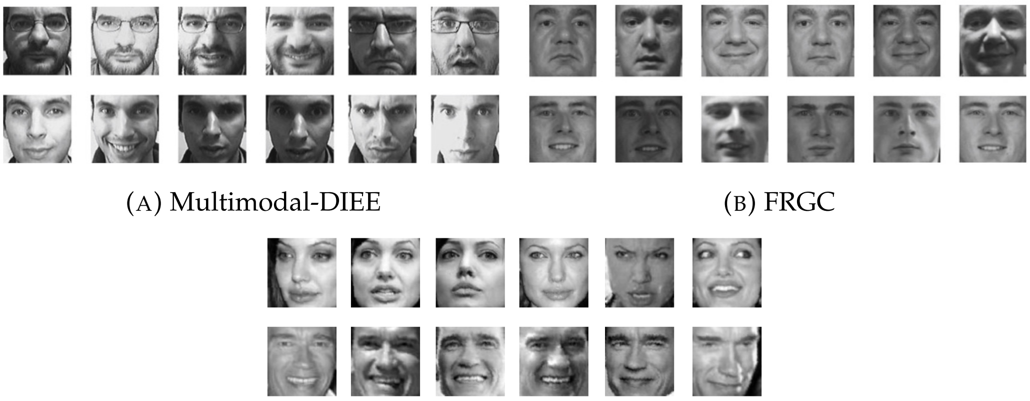
```
procedure SELF-UPDATE WITH EDITING ALGORITHM
    Let $GT$ be the intial template gallery
    Let $U = \{U_1, ..., U_n\}$ be the $n$ batches of unlabelled samples
    $p \leftarrow$ maximum number of template  per client
    $U = \{u_1, ..., u_n\}$ $\leftarrow$ unlabelled samples
    select(T,p $\leftarrow$ select p template from T (MDIST, DEND or others)
    estimate the update threshold $t^*$ using $GT$
    for $i = 1$ to $n$ do
        for $e$ in $u_i$ do
            if $score(e, GT) > t^*$ then
                if $size(GT) < p$ then
                    $GT_{new} = GT \cup e$
                else
                    $GT_{new} = select(GT \cup e, p)$
                end if
            end if
        end for
        $GT = GT_{new}$
        estimate the update threshold $t^*$ using $GT$
    end for
end procedure
```

3.3 Experiments and Analysis

3.3.1 Dataset

The purpose of this experimentation is to evaluate the performance of the novel classification-selection approach under different environmental conditions. We performed experiments on three publicly available datasets: the Multimodal-DIEE, the FRGC and a subset of the LFW. They were used, respectively, to simulate a fully controlled, partially controlled and uncontrolled application environment.

The Multimodal-DIEE dataset [53] (Fig. 3.5a), acquired by the University of Cagliari, includes 59 users (60 faces per user). The acquisition time is approximately 1.5 years divided into 6 acquisition sessions. The acquisition protocol is fully controlled: the dataset images are captured in the frontal pose, at a fixed distance from the sensor. Some variations in lighting conditions are present.



(A) Multimodal-DIEE

(B) FRGC

(C) LFW

FIGURE 3.5: Example of variation of faces in the dataset used to evaluate the performance of the novel classification-selection approach under different environmental conditions.

The FRGC [56], acquired by the University of Notre Dame, is composed by the faces of 222 users acquired on 16 sessions (Fig. 3.5b). Some of these sessions contain uncontrolled captures. The dataset, used to simulate a partially controlled environment, presents variations of expressions and lighting conditions. In this thesis experiments, 187 users have been selected with about 100/200 faces per subject.

The LFW dataset [44] is composed of the collection of more than 13,000 web images for a total of 5,749 users. Whereas for 4,069 people is present only a single image and, for many other users, the images are not enough to be used to simulate a process of template update for which several images are needed, only a subset of 16 people with about 15/30 faces per user for a total of 390 faces was selected from this dataset. Due to the nature of the dataset, the acquisition protocol is entirely uncontrolled. The images contain intra-class variations, in particular, variations of pose, lighting, age and expression (see Fig. 3.5c). The dataset was, therefore, used to simulate an uncontrolled environment.

3.3.2 Face representation

In this first evaluation and in the rest of the manuscript we used a handcrafted representation extracted through the BSIF [14] algorithm and three deep representations extracted through the neural networks FaceNet, ResNet50 and SeNet50[15–17].

The BSIF algorithm [14] calculates a binary code string for the pixels of a given image. The value of each pixel is considered as a local descriptor and may be used to build the histograms that allow characterising the textural properties within sub-regions of the image. Each bit in the binary code string is calculated binarising the response of a linear filter with a threshold at zero. For face recognition, facial image is divided into $n \times m$ non-overlapping regions and a descriptor is computed for each of these regions. Finally, the descriptors are concatenated for a comprehensive description of the face. We applied the BSIF algorithm with 10 filters of size 9×9 pels and a subdivision into 6×6 non-overlapping regions.

FaceNet [15] is a state-of-the-art deep convolutional neural network based on a triplet loss function. It maps each face into a 128-dimensional Euclidean space in which distances directly correspond to a measure of facial similarity. We used an open-source implementation based on TensorFlow [1] trained on the model 20170512-110547, deriving a 128B feature vector. This model has been trained on the MS-Celeb-1M dataset [57].

The Residual Network (ResNet-50) is a convolutional neural network 50 layers deep. We used the ResNet50 auto-encoding network [16], pre-trained on MS-Celeb-1M dataset and then fine-tuned on VGGFace2 dataset [58], to derive a 2,048B feature vector.

The Squeeze-and-Excitation Network (SeNet50) [17] is a exceptionally performing neural network, able to generalise extremely well across challenging datasets. We used this auto-encoding network to derive a 2,048B feature vector. The pre-trained model is trained on MS-Celeb-1M dataset and then fine-tuned on VGGFace2 dataset [58].

Before extracting the features, we applied some pre-processing steps (Fig.3.6). Faces are rotated in order to align eyes, guaranteeing a pre-set inter-ocular distance, scaled, cropped and saved in grayscale. The images are scaled to a 100×100 size for the Multimodal-DIEE and LFW datasets and 150×150 for the FRGC dataset.

3.3.3 A preliminary view on the feature space representativeness

Before evaluating the performance of facial recognition systems with the new classification-selection methods, we pointed out how the handcrafted and

[1] https://github.com/davidsandberg/facenet

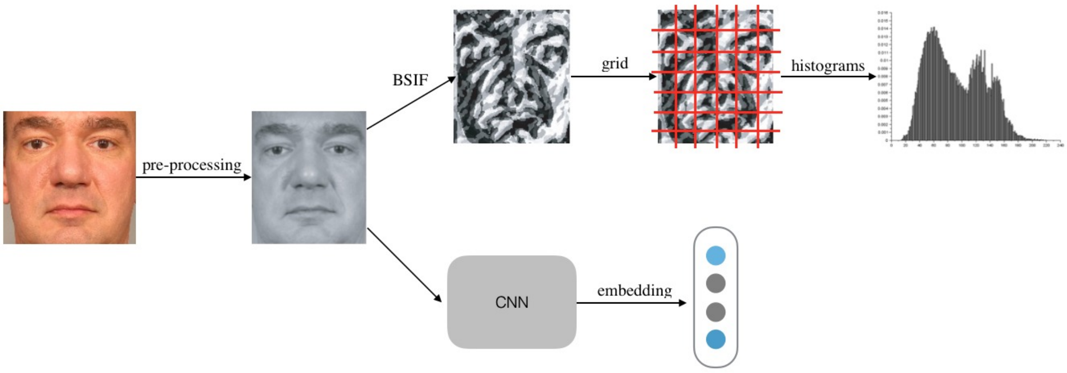


FIGURE 3.6: Block system of features extraction.

auto-encoded features spread the genuine users' and the impostors' matching scores over the three data sets. The methods proposed are based on the hypothesis that the distribution of the facial features is mono-modal for samples of the same user and multi-modal for samples of different users. In other words, we expect to identify clusters, partially overlapped, representing the users of the dataset. Viewing and analysing datasets' genuine and impostors matching scores distribution is useful for understanding how much the conditions of data acquisition affect performance.

We reported the matching scores sets for each subject of the datasets in Figs. 3.7-3.9, where the x axis is the subject identifier. Each graph relates to a dataset and a representation of the face and shows the matching scores relating to the images of the same user in blue and those relating to the comparison with images of impostors in red. In the plots we have drawn the average threshold on all users, calculated by minimizing the total classification error. Since the two classes, genuine and impostors, are unbalanced, this threshold is useful only to give an idea of the overlap of the two classes.

The Multimodal-DIEE shows an accentuated "separation" between the genuine users' and the impostors' matching scores, compared to the other two datasets, in particular for what concerns the deep features (Fig. 3.7). Although the BSIF representation tends to overlap the two distributions, proving to be the least effective facial representation, the distribution of the impostors does not completely overlap with that of the genuine, as is the case for the LFW dataset with the same method (Fig. 3.9). The smaller overlap of the distributions is due to the fact that the Multimodal-DIEE dataset images contain little variation in the appearance of each individual's face.

For the FRGC images, a higher degree of overlapping can be noticed even in the case of the auto-encoded features. This confirms that the dataset can simulate a partially controlled environment for face recognition due to the highest presence of face appearance variations. From the graphs, it is evident that the population of the FRGC dataset is much larger than the other two datasets.

In the LFW results, a strong overlap emerges between the scores relating to the handcrafted features of the genuine and those of the impostors

(Fig. 3.9). The dataset, being collected from images with very different acquisition conditions from each other, represents a completely uncontrolled environment. The high representativeness of the auto-encoded features still manages to separate the two distributions, even if a slight overlap is present.

3.3.4 Experimental protocol

In this evaluation, each data set was randomly divided into seven batches. We used the first batch as the initial gallery and the last batch as the test set to evaluate the performance of the system. The other five were used as adaptation sets in order to simulate periodic system's update. According to [10], an independent test set is useful to have the same reference for all update cycles. In particular, for these experiments, having no temporal information to exploit, an approach with a dependent test set [53] would not have contributed in the analysis of the results.

The parameter p defines the initial number of templates in the gallery and the number of samples per user present in each batch. This value must be chosen on the basis of the application and the technical characteristics of the devices in use. In this work we used small values of p to simulate the worst case, i.e. applications with stringent requirements from the point of view of computational and storage resources.

For the Multimodal-DIEE and FRGC data sets we used $p \in \{5, 6, 7\}$, and $p = 4$ for the LFW data set due to the low number of samples per user.
As performance evaluation metrics, the Equal Error Rate (EER) and the percentage of impostors wrongly added were calculated.
We repeated each experiment 10 times and averaged the results over those runs. All the experiments were performed with a desktop PC with operating system Windows 7 Professional 64bit, a Intel Xeon E5 2630 v3 processor, HDD, 32 GB RAM and using MATLAB v.R2013a.

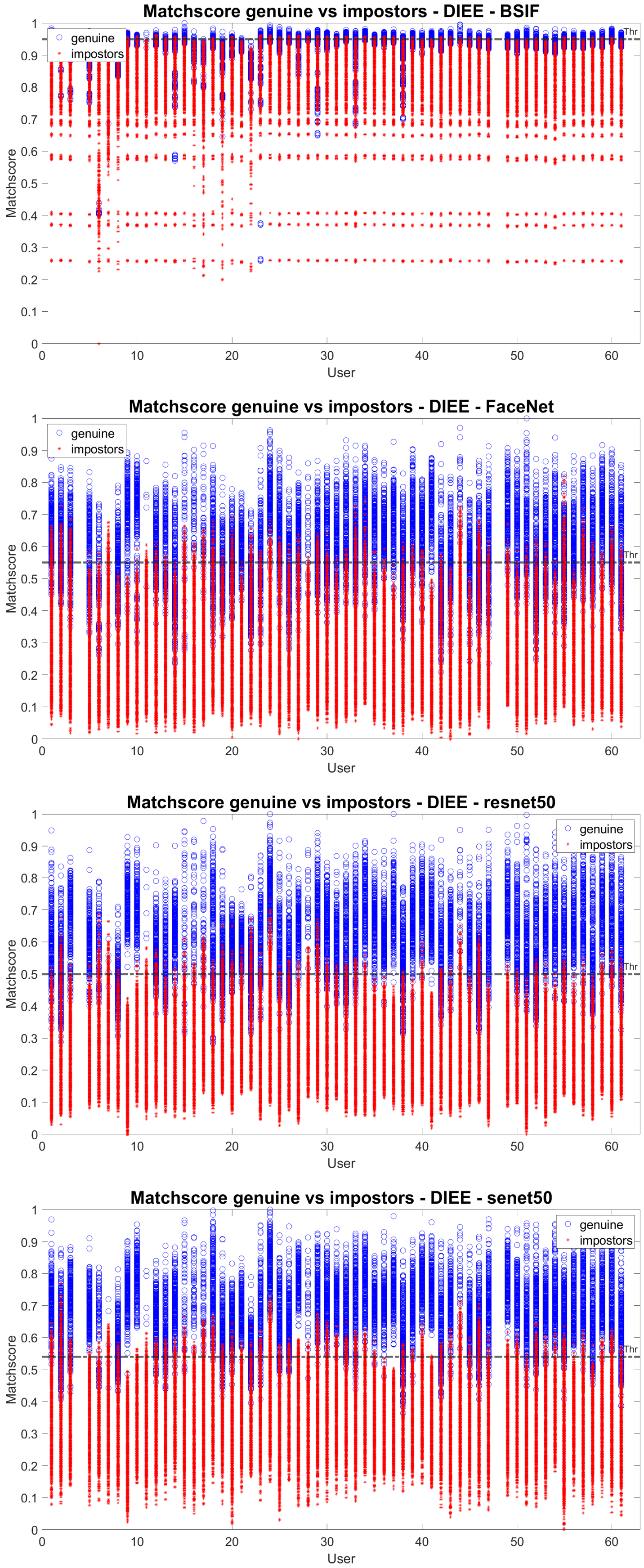


FIGURE 3.7: BSIF, FaceNet, ResNet50 and SeNet50 matching scores for the DIEE data set. The plots report in the x axis the subject identifier and in the y-axis the matching scores among the first five templates (genuine, in red) and the other samples (impostors, in blue).

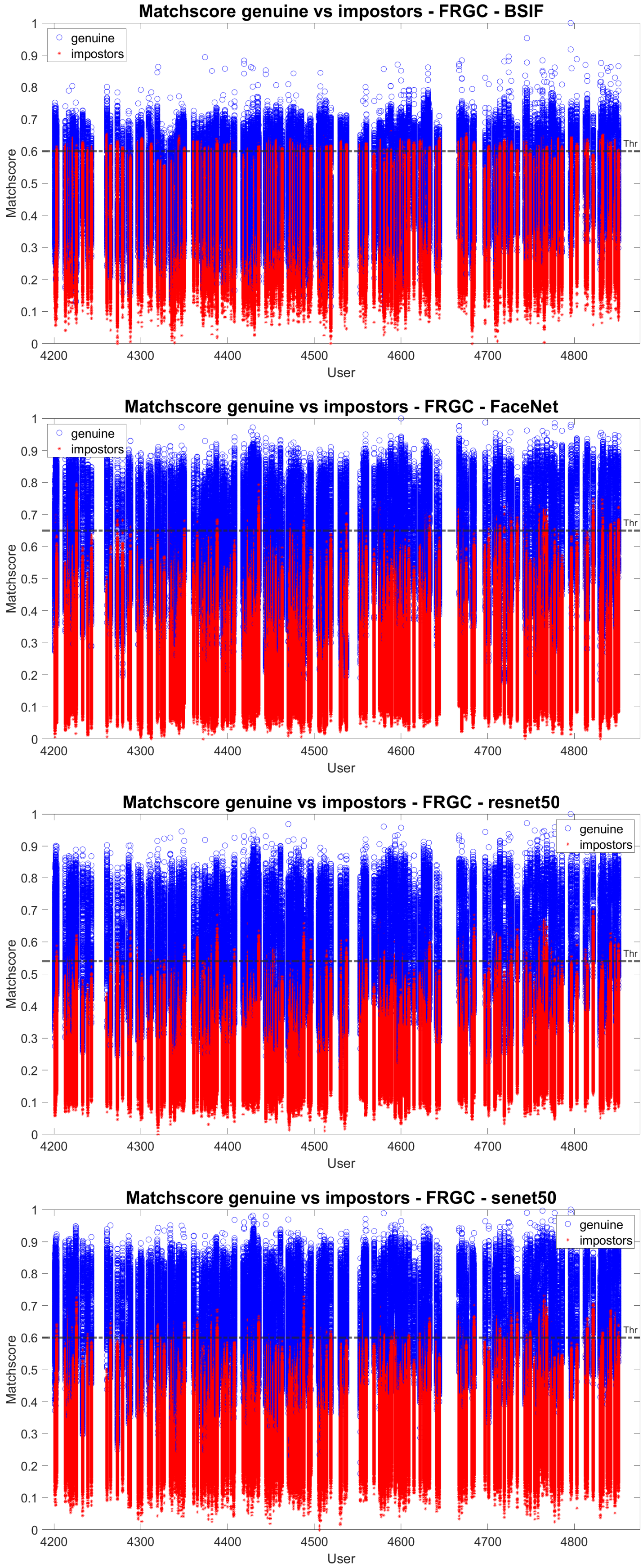


FIGURE 3.8: BSIF, FaceNet, ResNet50 and SeNet50 matching scores for the FRGC data set. The plots report in the x axis the subject identifier and in the y-axis the matching scores among the first five templates (genuine, in red) and the other samples (impostors, in blue).

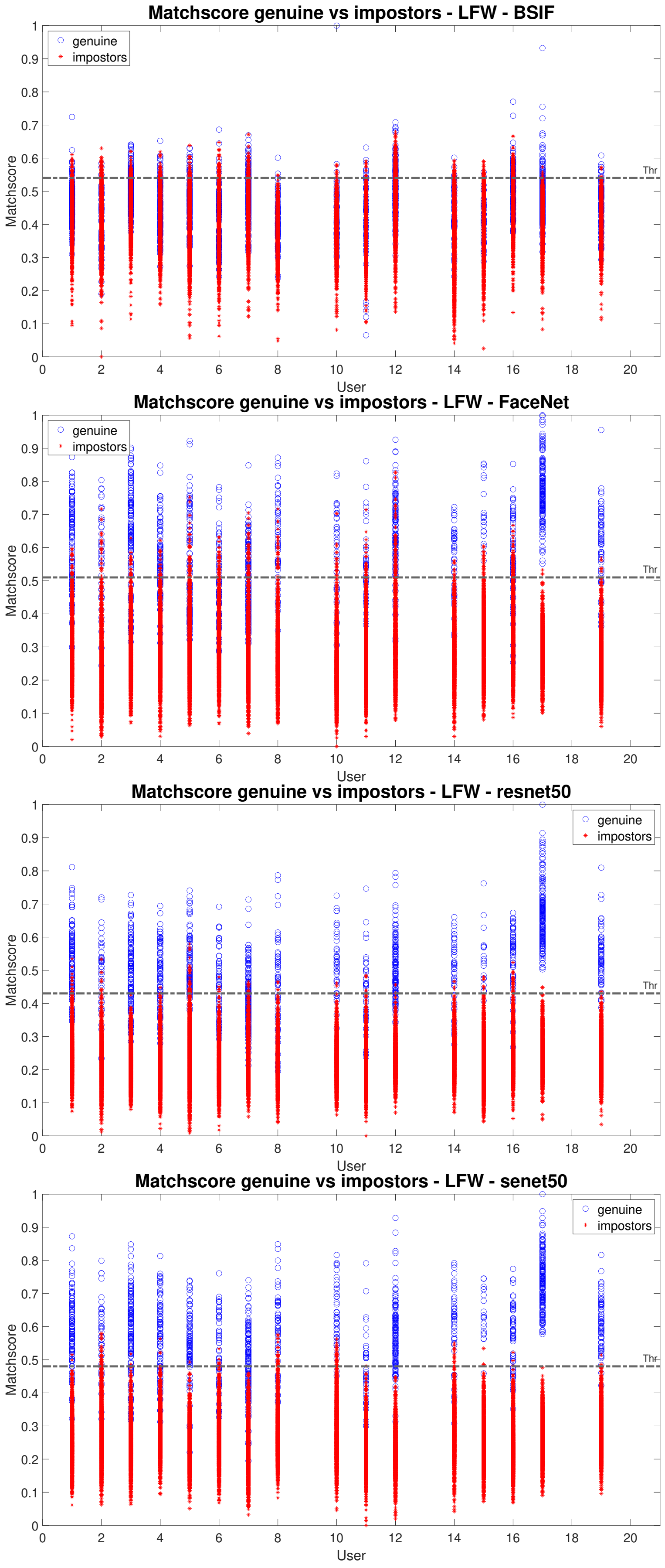


FIGURE 3.9: BSIF, FaceNet, ResNet50 and SeNet50 matching scores for the LFW data set. The plots report in the x axis the subject identifier and in the y-axis the matching scores among the first five templates (genuine, in red) and the other samples (impostors, in blue).

3.3.5 Results

In this section, the performance of the proposed approaches and the other state-of-the-art adaptive methods using handcrafted (BSIF) and auto-encoded (FaceNet, ResNet50 and SeNet50) features are presented. In particular, the traditional self-update [59], the method based on risk-minimization [11], the semi-supervised PCA [10], the K-means [60], and the editing classification/selection methods MDIST, DEND [9, 19] were implemented and tested in terms of EER, impostors percentage and processing time.
For Multimodal-DIEE and FRGC data sets three possible values for p, 5, 6 and 7 are tested, representing different constraints in terms of storage size. As the results did not differ significantly for different p values, we reported only $p = 6$ results. The plots show the performance measurements, in the y-axis, against the batch numbers, in the x-axis. In addition to the state-of-the-art updating methods presented in the Section 2.3, each EER plot also shows the performance of the system without updating (grey line, constant as it is not affected by the update).

Figs. 3.10,3.11,3.12,3.13 show the average values of impostors' percentage, the EER and the processing time for the Multimodal-DIEE data set using, respectively, BSIF, FaceNet, ResNet50 and SeNet50 features.

Of particular interest is that the "optimised" classification-selection algorithms, namely K-means and MDIST, exhibit an impostors percentage inserted in the gallery close to zero for all batches. Consequently, they manage to exploit the initial hypothesis on main modes. As counter-proof, the DEND method leads to a high number of impostors. The novel approach proposed, despite providing a limited and low number of templates, manage to maintain a high degree of representativeness and a very low EER. For all investigated feature sets, we can notice a significant improvement over the basic performance of the system (in grey). This points out the usefulness of this novel self-updating approach. In fact, in addition to a very low percentage of impostors into the galleries, they outperform the state-of-the-art adaptive methods.

In Figs. 3.14,3.15,3.16,3.17 the same metrics are reported for the FRGC data set. The results confirm a low percentage of impostors and high performance for the "optimised" classification-selection algorithms. Also for this dataset, as expected, the DEND method is not performing and introduces a large number of impostors into the system.

As reported in the Sec.3.3.4, for the LFW data set we used as initial gallery size and as number of samples per user present in each batch the value $p = 4$, due to the small size of the subset, (Figs. 3.18,3.19,3.20,3.21,3.22)

The results show us that, in a completely uncontrolled environment, a more significant discrepancy between BSIF and neural network features is pointed out. In fact, we have a significant loss in performance for the BSIF handcrafted features 3.18. This drop does not occur with auto-encoded features. However, maybe due to the small user population, K-Means is still the best approach to self-updating for BSIF features.

To sum up, the EER values of the novel classification-selection method are the only ones that never exceed the line relating to the system without

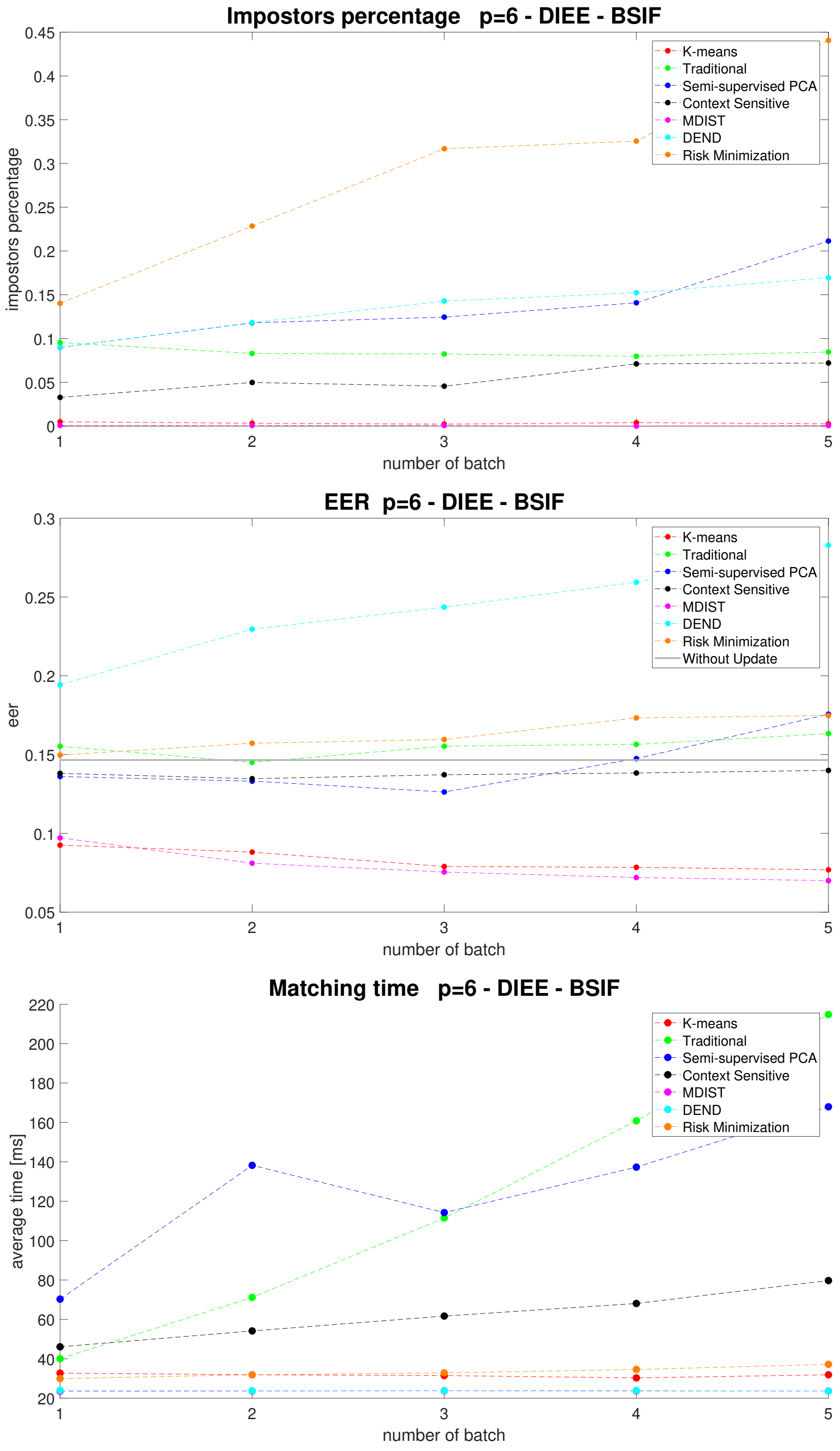


FIGURE 3.10: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for Multimodal-DIEE using BISF handcrafted features. On the x axis is shown the number of the batch and on the y-axis the performance index.

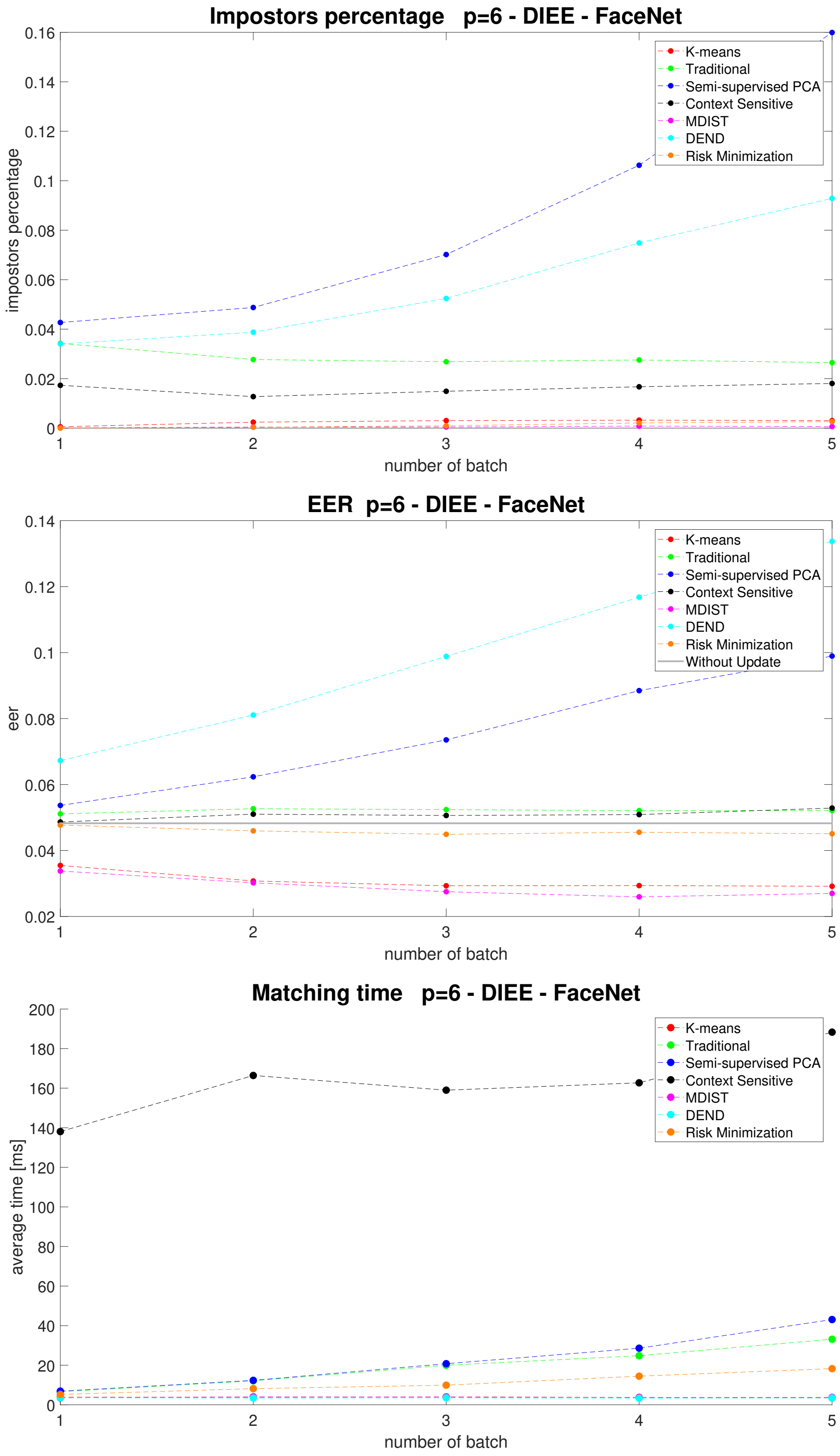


FIGURE 3.11: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for Multimodal-DIEE using FaceNet auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

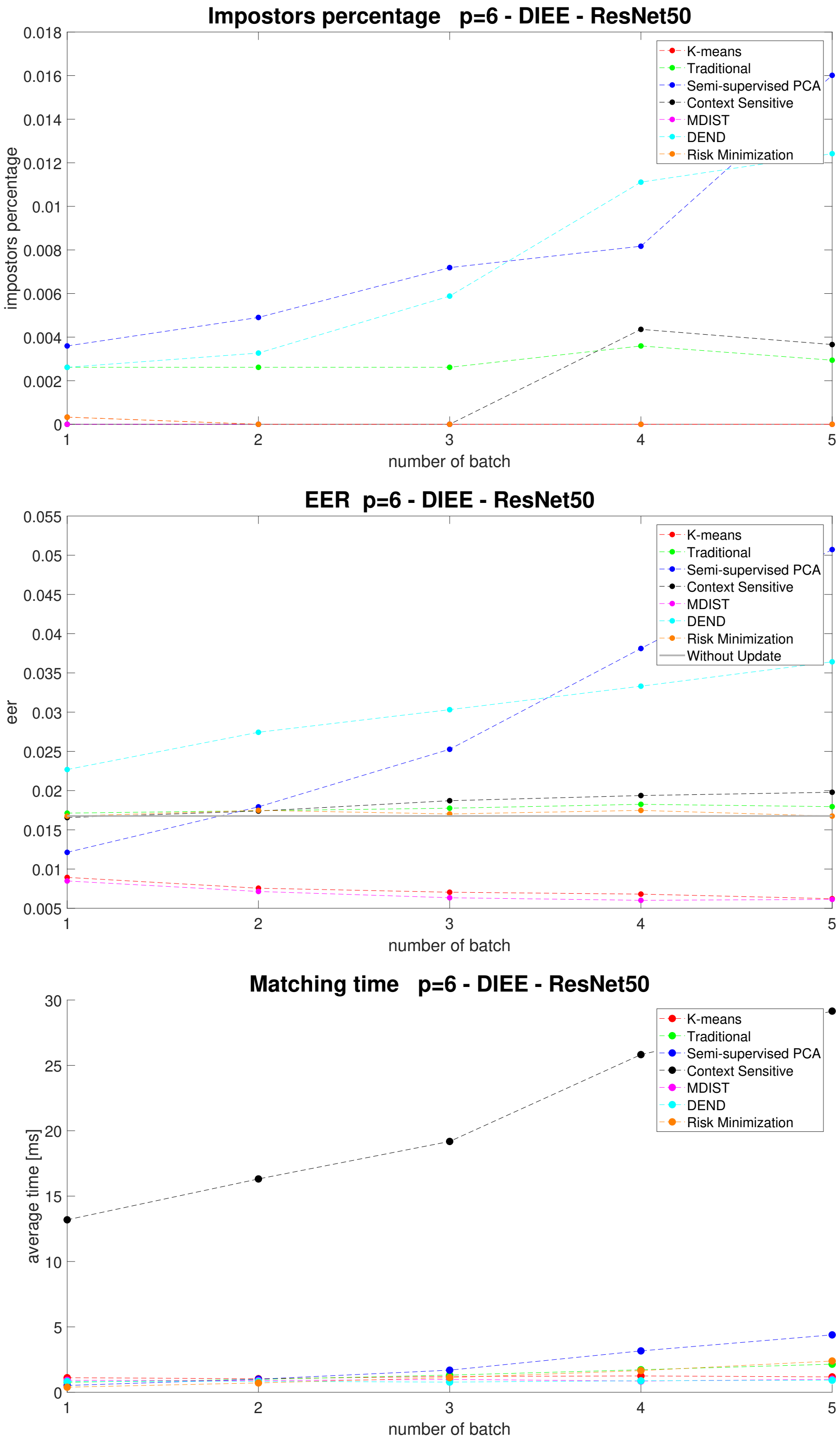


FIGURE 3.12: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for Multimodal-DIEE using ResNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

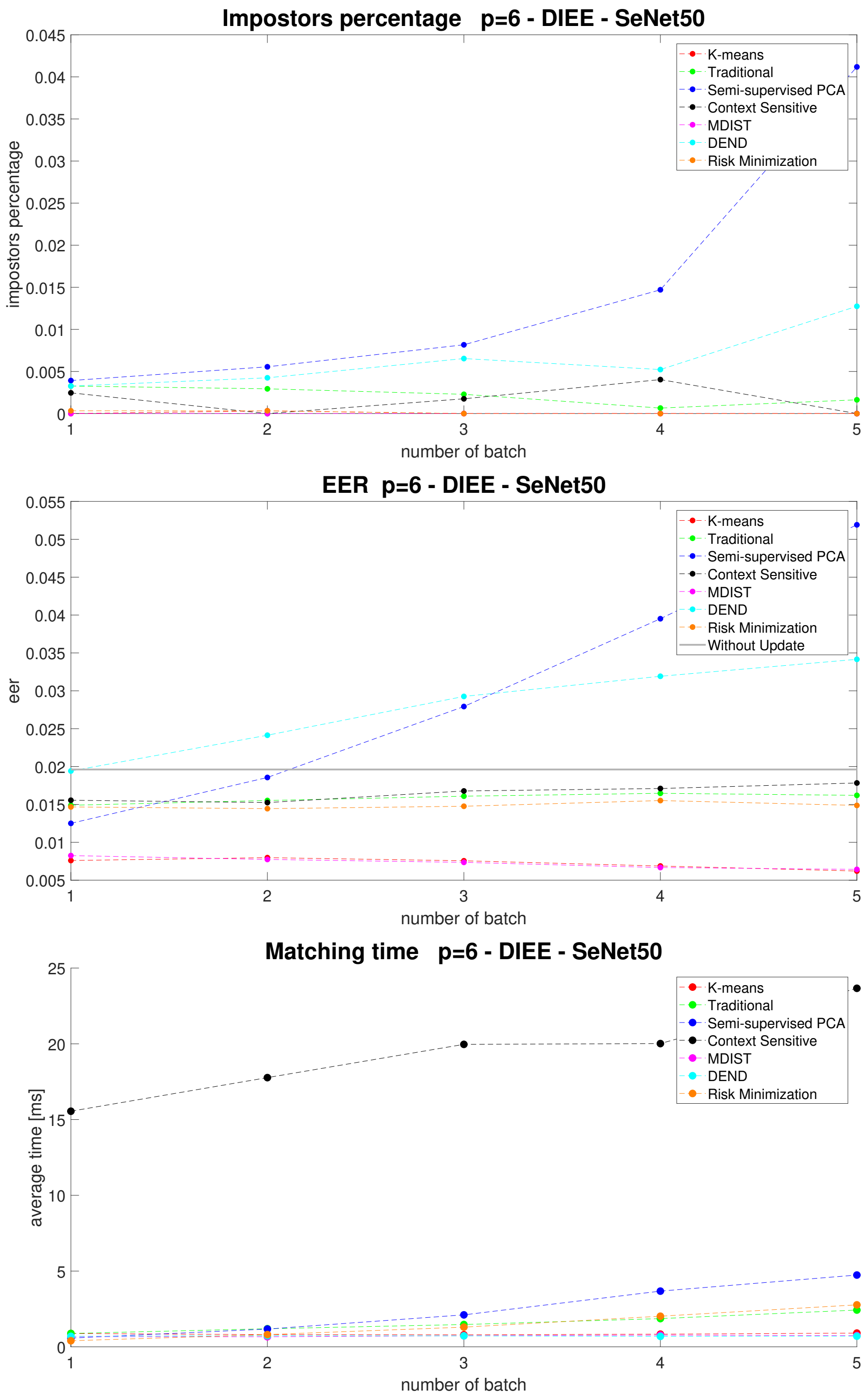


FIGURE 3.13: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for Multimodal-DIEE using SeNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

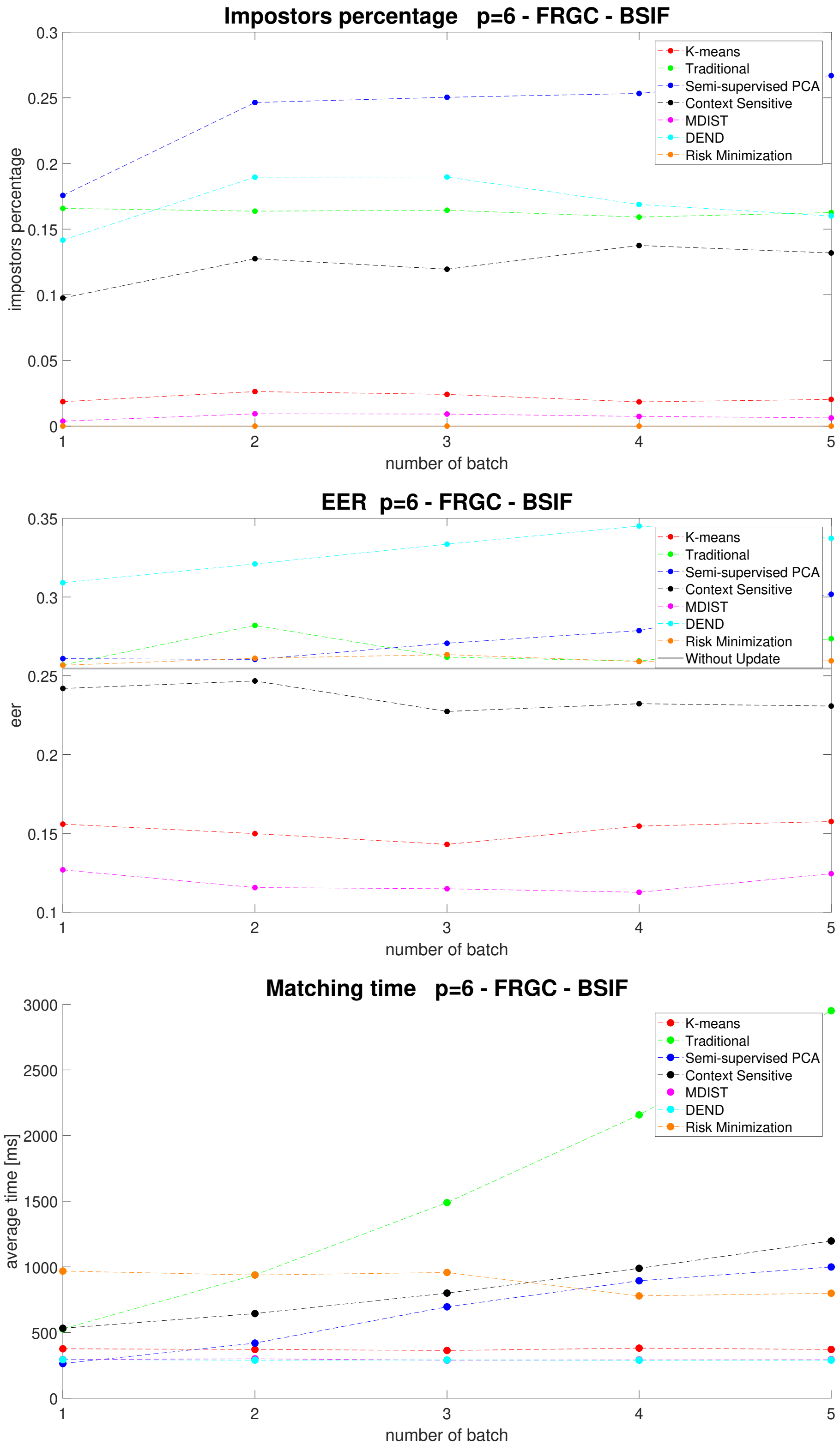


FIGURE 3.14: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for FRGC using BISF handcrafted features. On the x axis is shown the number of the batch and on the y-axis the performance index.

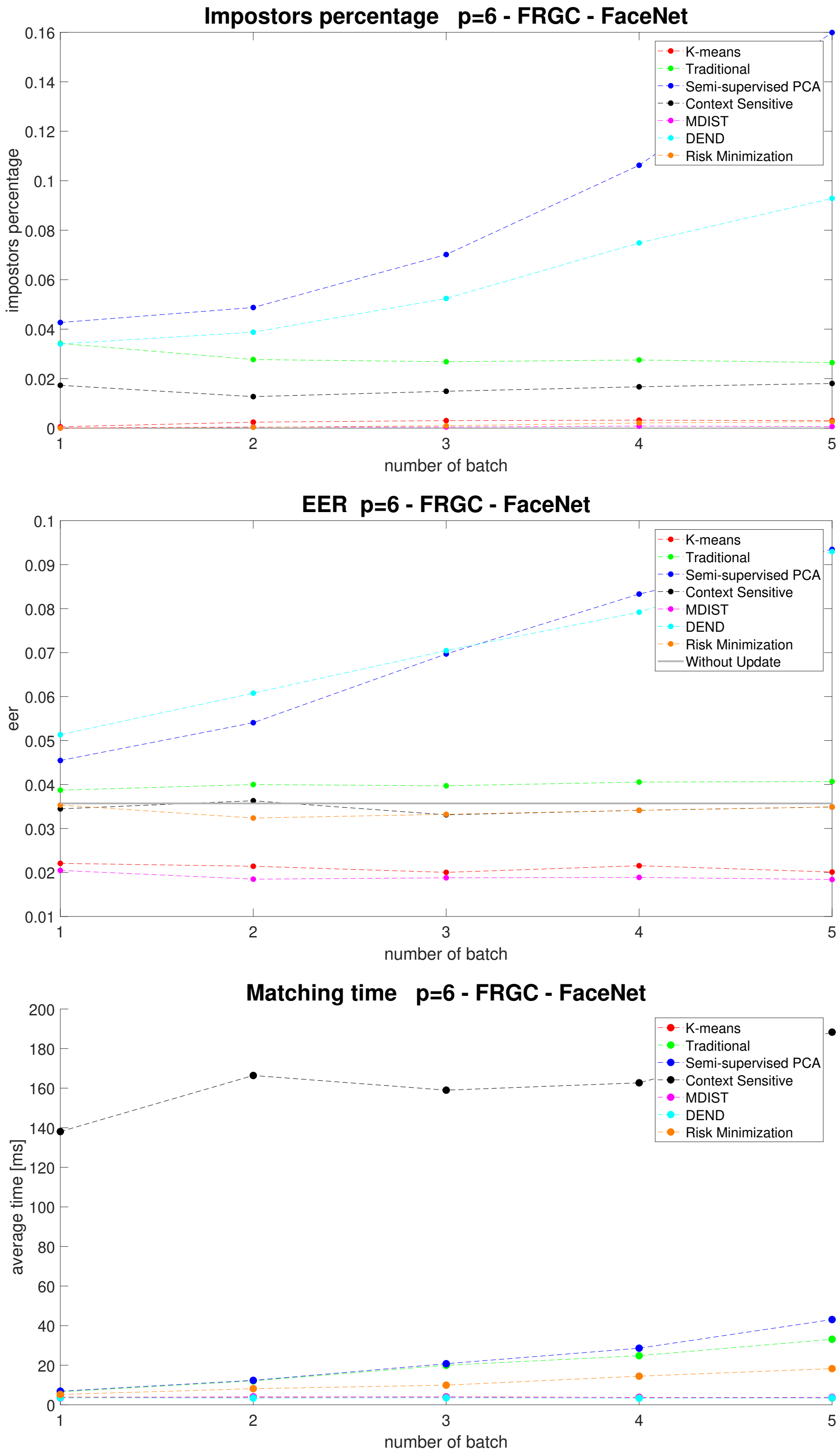


FIGURE 3.15: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for FRGC using FaceNet auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

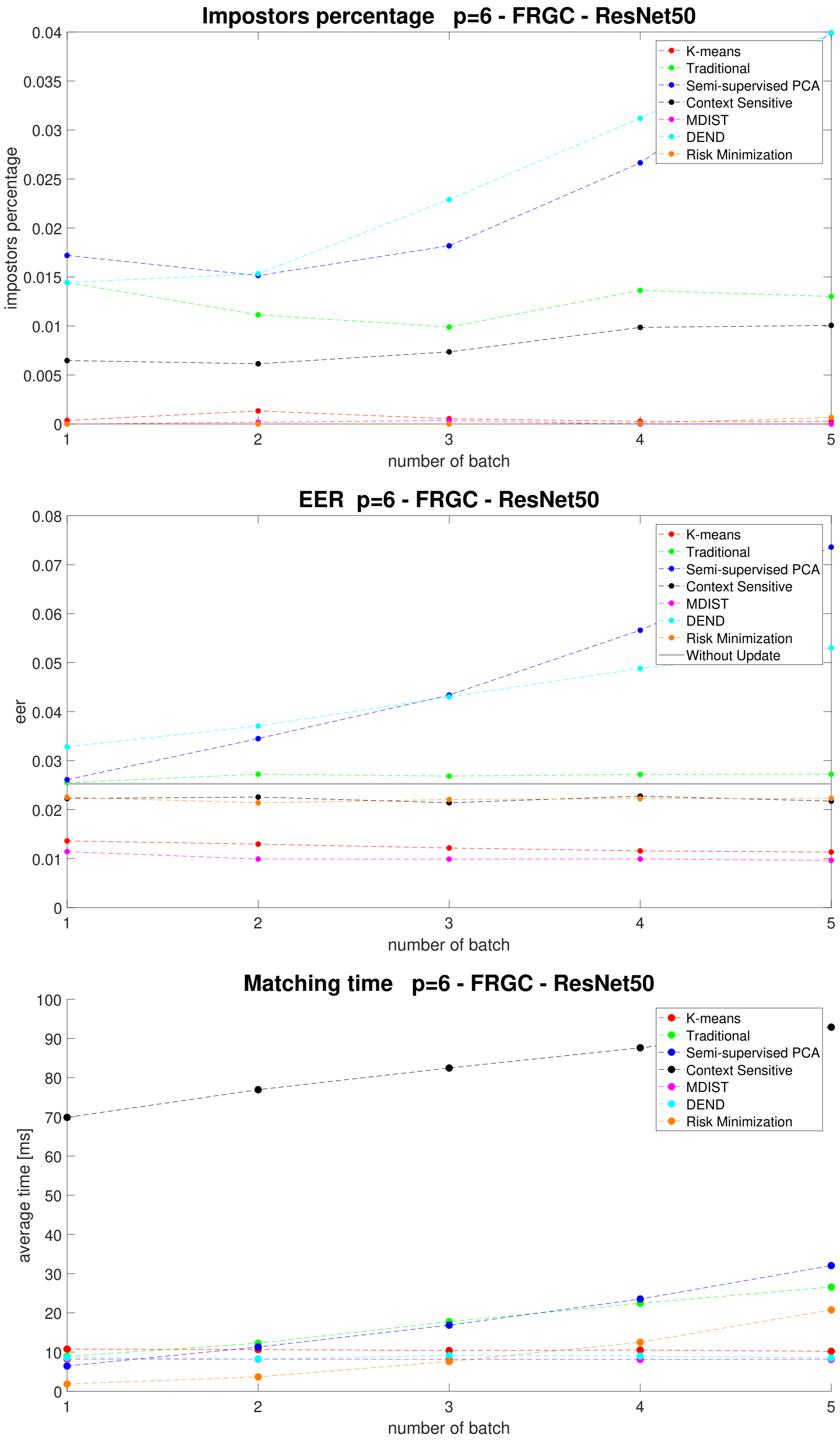


FIGURE 3.16: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for FRGC using ResNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

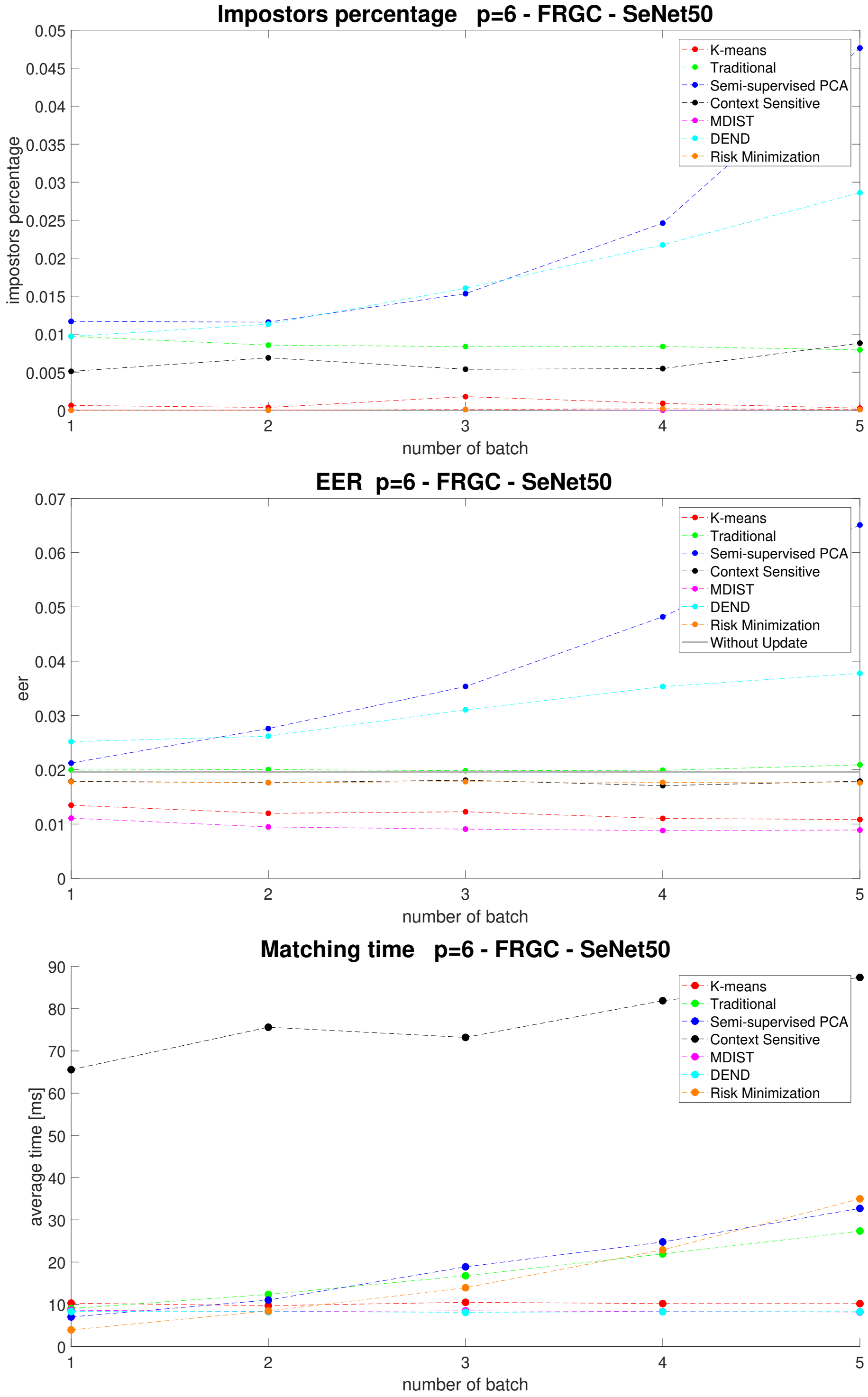


FIGURE 3.17: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=6 for FRGC using SeNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

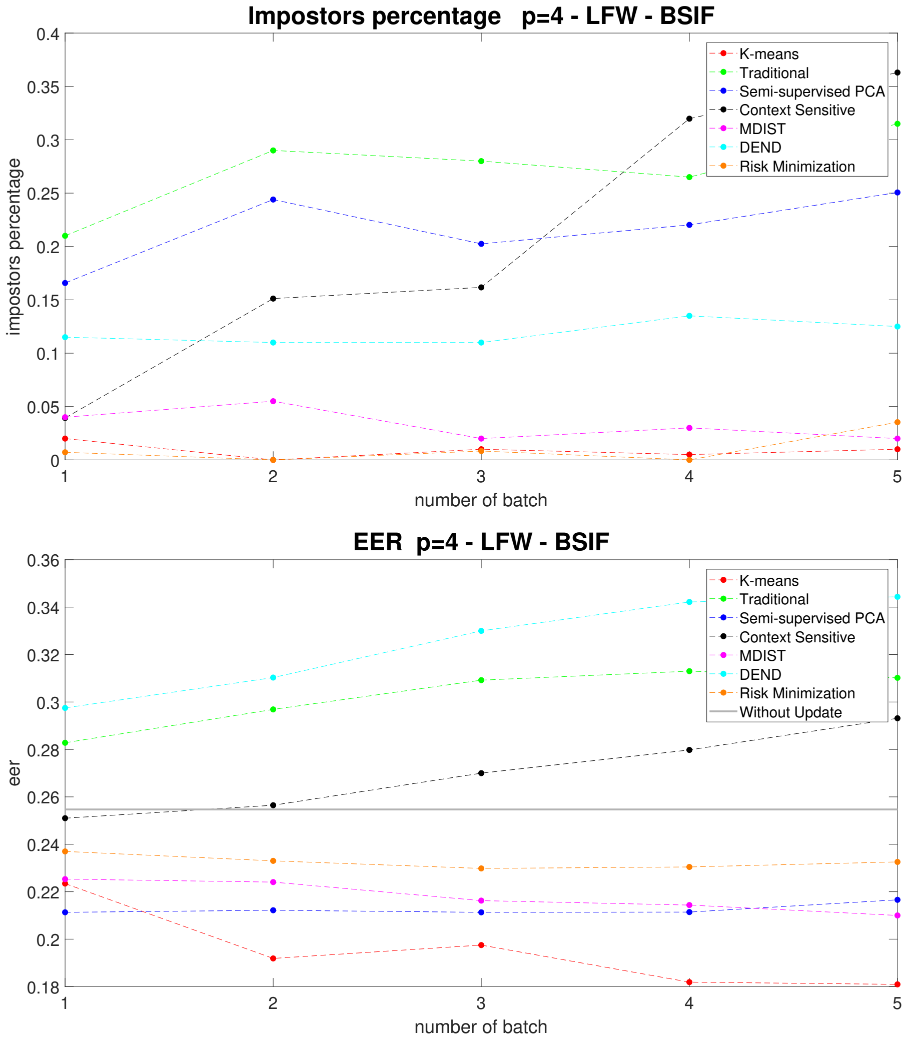


FIGURE 3.18: EER, percentage impostors comparison among the state of the art and the new proposed method with p=4 for LFW using BISF handcrafted features. On the x axis is shown the number of the batch and on the y-axis the performance index.

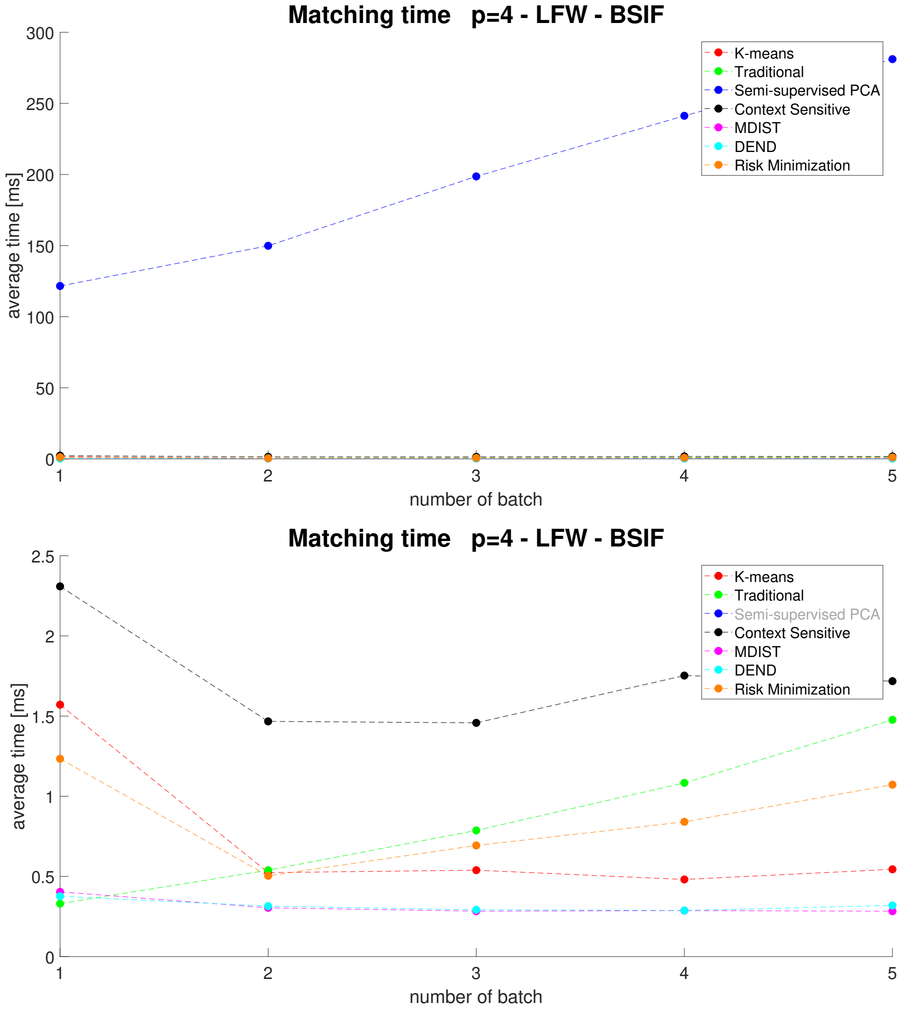


FIGURE 3.19: Matching time comparison among the state of the art and the new proposed method with p=4 for LFW using BISF handcrafted features. On the x axis is shown the number of the batch and on the y-axis the average time in milliseconds. In the second plot, the curve of the Semi-supervised PCA method has been eliminated to highlight the times of the other methods.

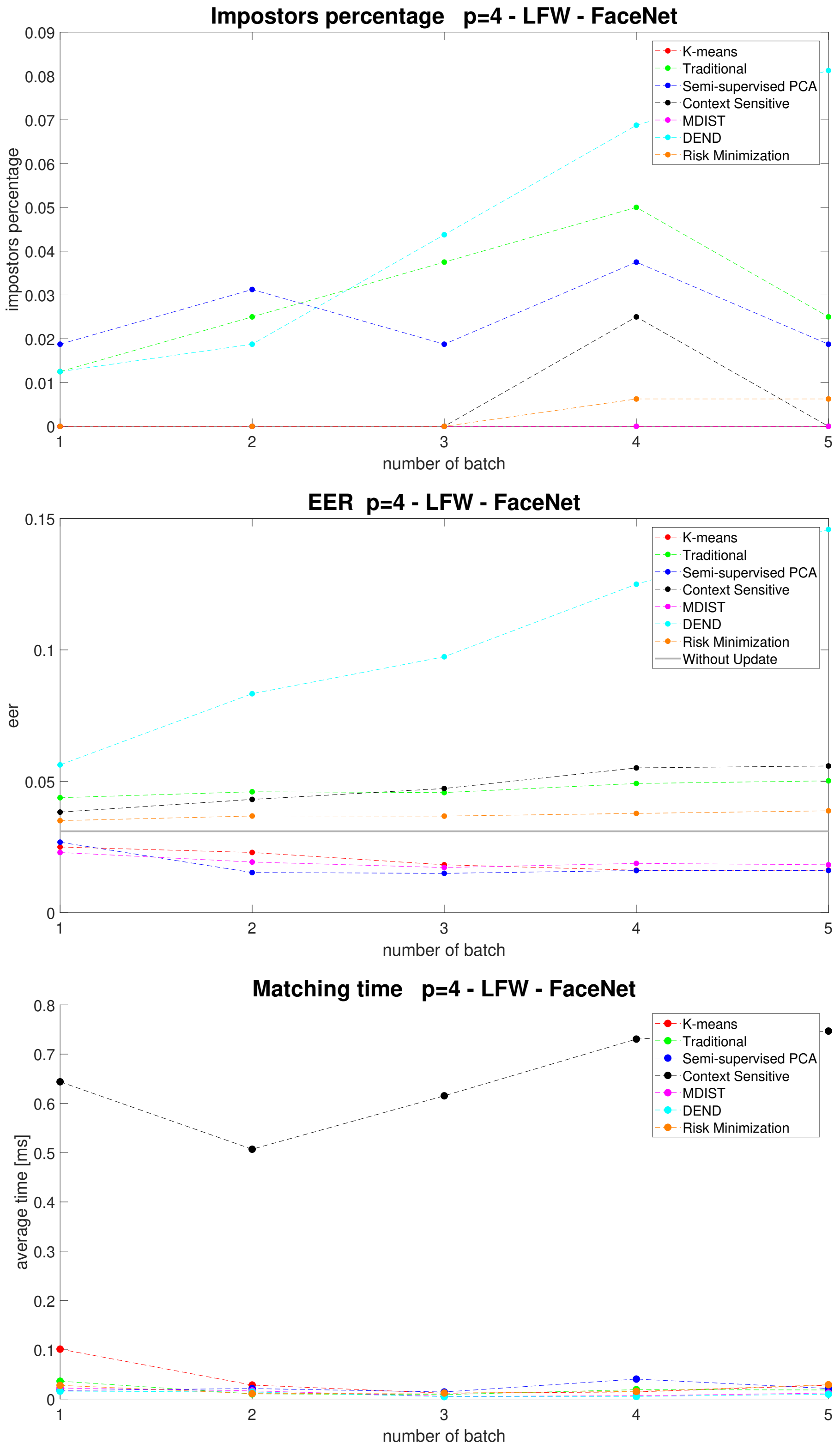


FIGURE 3.20: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=4 for LFW using FaceNet auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

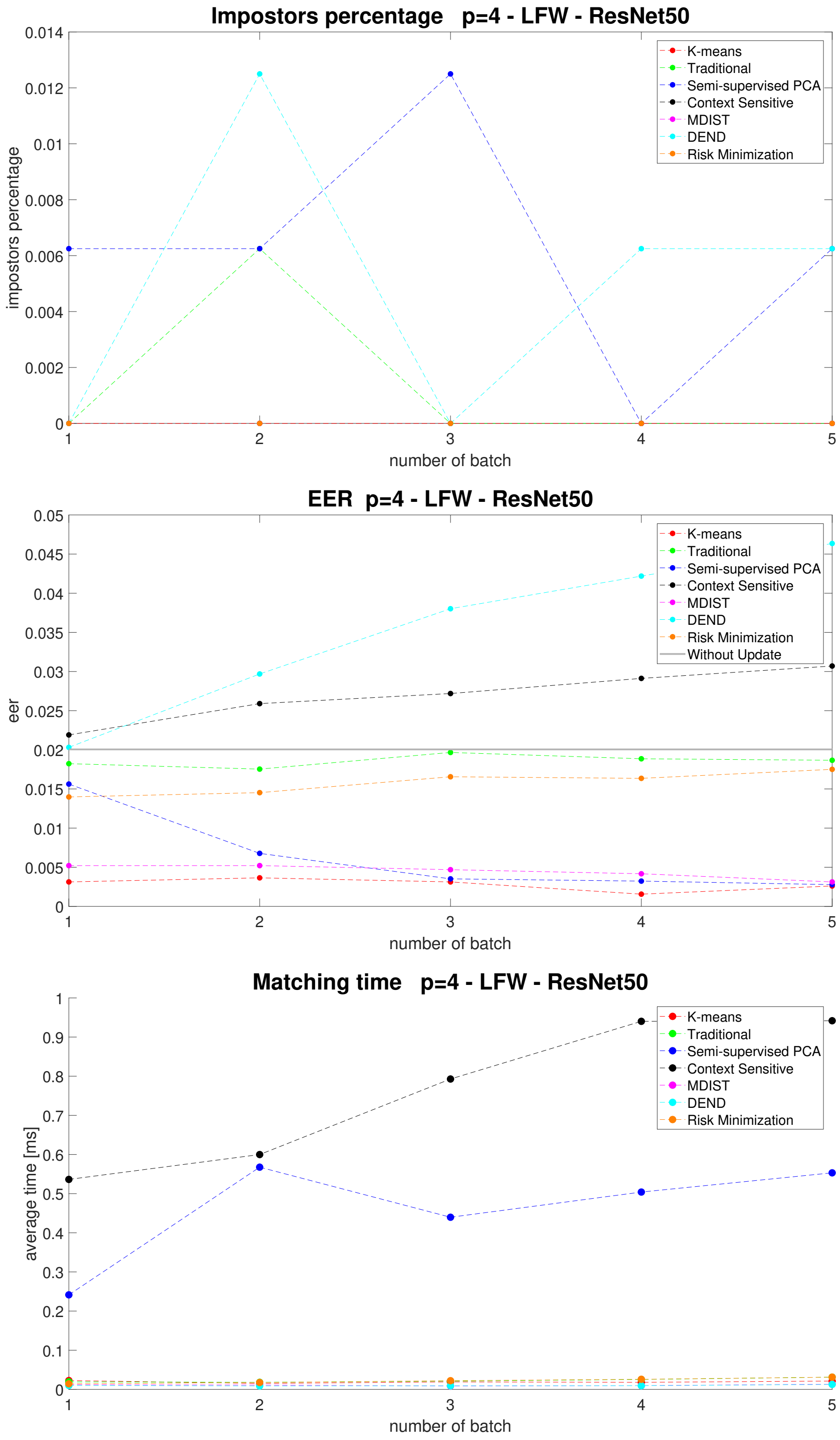


FIGURE 3.21: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=4 for LFW using ResNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

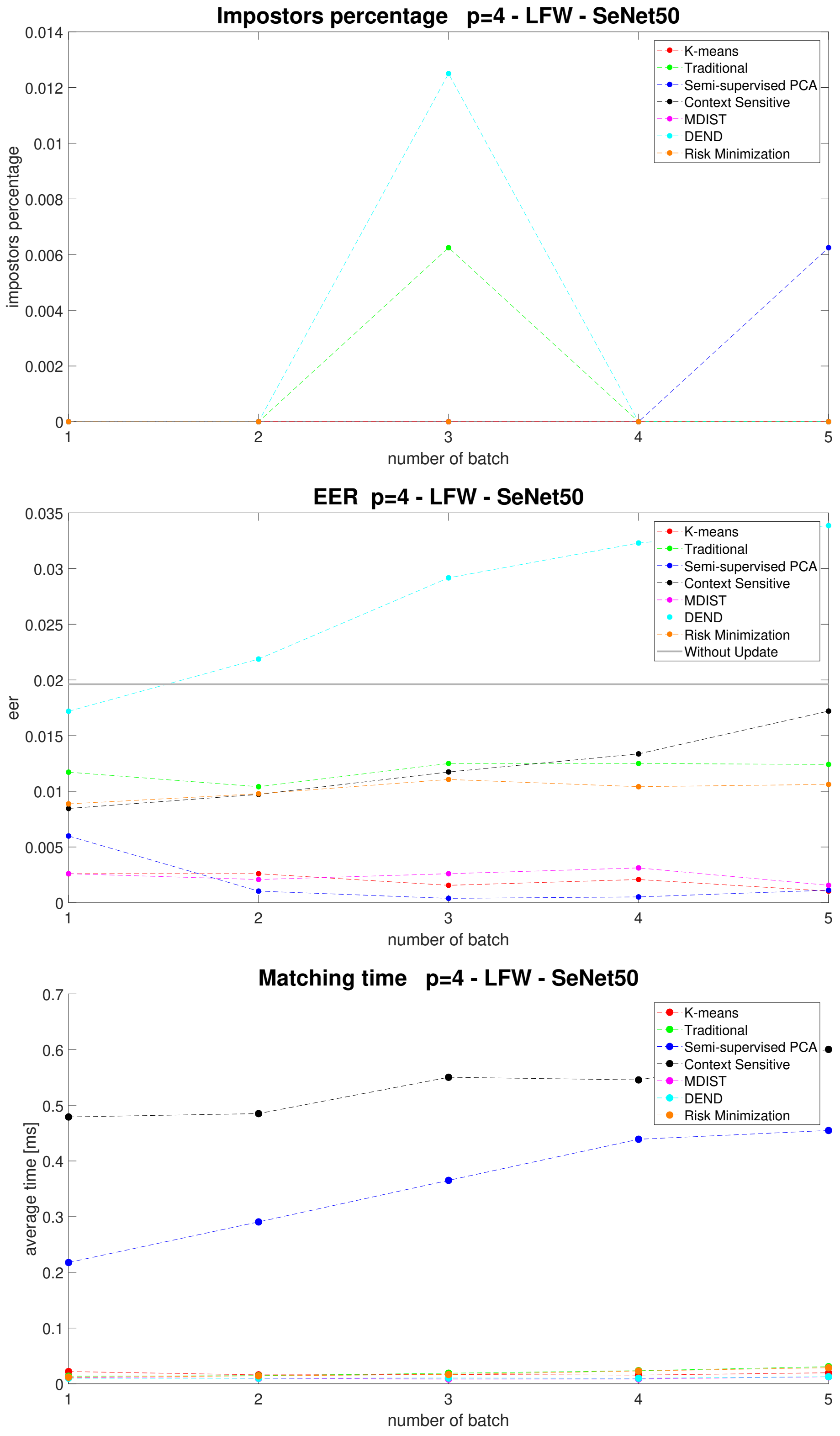


FIGURE 3.22: EER, percentage impostors and matching time comparison among the state of the art and the new proposed method with p=4 for LFW using SeNet50 auto-encoded features. On the x axis is shown the number of the batch and on the y-axis the performance index.

updating. In particular, although the LFW data set simulates an uncontrolled and hostile environment for self-updating, the hypothesis behind our model appears to be enough for guaranteeing the performance improvement and the good choice of the updated templates. These results are possible thanks to the very low and almost zero number of impostors who "dirty" the gallery.

In order to evaluate if the method satisfies the requirement under which it was modelled, i.e. to keep under control the computational complexity of the recognition system, we investigated the analysis of processing time of the proposed algorithms. The time graphs show the increase of state-of-the-art methods of matching time as the number of batches increases. The methods with template limit allow instead to keep the processing time constant at each update iteration. The template limit per user means that the storage size can not exceed $(p \cdot k \cdot S)$ B. As previously indicated, p is the maximum number of template per user. We indicated with k the users number in the gallery and with S the feature vector size (byte). This value is constant after reaching the limit of p samples per user and independent of the iterations number. In the classification-selection systems without a limit of templates per user, the storage size depends on the iterations number i and it is $(\beta \cdot i \cdot \bar{m} \cdot k \cdot S)$ B. In the previous formula, we indicated with $\bar{m}$ the average number of templates added per updating iteration and with β the rate of selected templates in the range [0,1]. This value is obviously less than the overall acceptance rate set in terms of threshold t^*. Worth noting, $\beta = 1$ for the traditional self-update systems. Without the limit in the number of templates per user, the required storage memory tends to ∞ with i.

Chapter 4

A long time span dataset: the APhotoEveryday

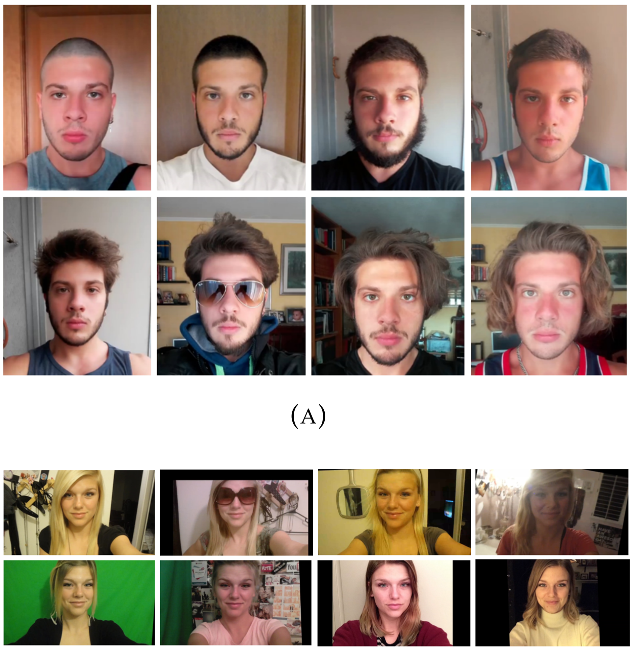
The evaluations carried out so far consider two datasets collected over a short time span, the Multimodal DIEE and the FRGC datasets, and a dataset, the LFW, that, although present medium/long-term changes, contains few samples per user. All three of these datasets, like most of the existing face datasets publicly available, do not contain temporal information on the data. Table 4.1 shows the characteristics of the facial datasets used so far to evaluate adaptive systems. It is possible to notice that they do not contain enough samples per user and were acquired in a medium-short term time. Accordingly, these datasets are not suitable to analyse properly how adaptive biometric systems follow the typical intra-class variations of ageing.

Dataset	**Ref.**	**# users**	**Avg.# samples/user**	**Avg. time**	**Highlights**
FIA	[61]	153	∼ (3 sessions)	10 months	expression, illumination, eye glasses, sequences
FRGC	[56]	200	133	2 years	expression, illumination, background, sequences
Mult.DIEE	[53]	59	60	1.5 years	frontal pose, expression, illumination, sequences
LFW	[44]	5,749	2	n/a (long period)	expression, illumination, occlusions, background, eye glasses
AR	[62]	116	26	2 weeks	frontal pose, expression, illumination, occlusions, eye glasses
IJB-A	[63]	500	11	n/a (long period)	expression, illumination, occlusions, background, eye glasses
APE	[18]	98	825	4.1 years	frontal pose, expression, illumination, occlusions, background, eye glasses

TABLE 4.1: A list of the existing datasets used to evaluate adaptive biometric systems with characteristics related to the number of users, the number of samples per user and the acquisition time span.

In an attempt to simulate situations of long-term use in which the temporal intra-class variability of the face appearance is high, we collected a new dataset using the frames of some YouTube videos within Daily Photo Projects [1] (Fig.4.1). The videos that make up this dataset, called "APhotoEveryday" or APE, contain a very high variability, as they aim to show a change in the person's appearance such as ageing, beard or hair growth, the transition between adolescence and adulthood, etc.

[1] Example: https://www.youtube.com/watch?v=iPPzXlMdi7o



(A)

(B)

FIGURE 4.1: Example of face appearance variation over time from the APE data set. These users have granted us the permission to use the images extracted from their YouTube videos.

Exploiting the high variability and the temporal information allows us to carry out an analysis that realistically simulates the normal adaptive facial recognition system operation by analysing how long-term use influences the performance in terms of recognition accuracy and percentage of impostors added to the gallery.

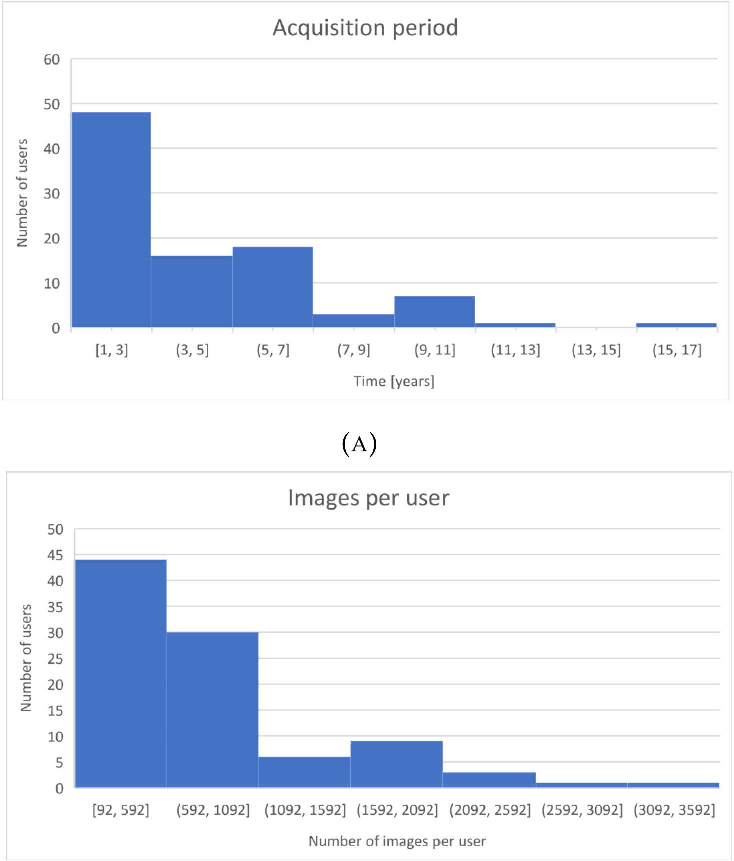
Fig.4.2 shows the APE details: the number of images per user varies between 92 and 3,578 with a median value of 661 and the acquisition time varies between less than one year and sixteen years with a median value of 3 years. Each image is labelled with a number that indicates the temporal progression of the user sequence. The first version of the APE dataset published in [18] contained 46 users.

4.1 Experiments and Analysis

4.1.1 Experimental Protocol

The experimental protocol was designed in order to exploit the characteristics of the APE dataset, in particular the high number of samples per user, the intra-class variations, the long acquisition time span and the temporal information.

For this reason, the APE users were divided on the basis of acquisition times into three categories: short time (less than two years of acquisition),



(A)

(B)

FIGURE 4.2: Details on the number of images per user and on the acquisition time of each video. In the first histogram o the acquisition times (a) and in the second the numbers of images per user (b).

medium time (between two and five years) and long time (more than five years). The percentages of users for each category are shown in Fig. 4.3.

As a preliminary analysis, the representativeness of the Facenet auto-encoded templates was evaluated using the Euclidean distance of the features. More details are given in the Sec. 4.1.2. The metrics used to evaluate the recognition performance on the single user are the "maximum recognised sequence", the "maximum sequence unrecognised" and the "recognition accuracy". The first parameter indicates the maximum percentage value of consecutive correct positive recognition, the second indicates the maximum percentage value of consecutive erroneous recognition, i.e. how many times in a row an incorrect classification has been found. Finally, the accuracy indicates the percentage of correct total identifications for the user, i.e. the percentage of images correctly recognised as belonging to the individual considered.

The comparison between adaptive systems was carried out using the three neural networks described in the Sec. 3.3.2, FaceNet, ResNet50 and SeNet50. As for the experiments on adaptive systems, the experimental protocol is summarised as follows:

- We subdivided the APE dataset into ten batches maintaining the sequence time progression: the first batch is the initial gallery and is composed of the first $p = 5$ samples per user; the remaining nine parts, composed of $\frac{\#samples - p}{\#adaptationsets}$ images per user, are the adaptation sets.
- As performance evaluation metrics, the Equal Error Rate (EER) and the percentage of impostors wrongly added were calculated. The updating threshold was estimated at FAR=1%.
- The adaptation sets were used to simulate the periodic sets of batches collected during the system operations individually analysed to update the users' templates.
- To simulate a real application and exploit changes in appearance over time we used the $(i+1)^{th}$ batch as test set of the i^{th} batch as suggested in [64].

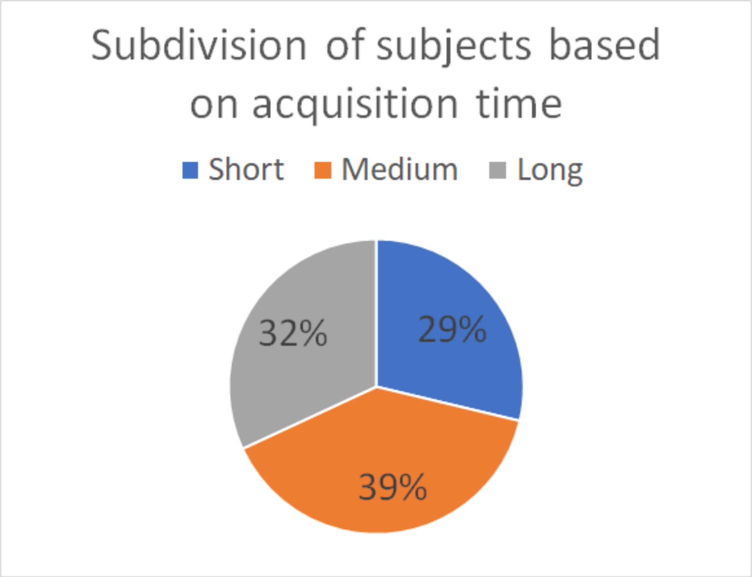


FIGURE 4.3: Graph showing the percentage of subjects falling into the categories short time (less than two years of acquisition), medium time (between two and five years) and long time (more than five years).

4.1.2 Representativeness of autoencoded templates over time

To investigate to what extent deep features are able to follow all the intra-class variations that the face can present over time, we carried out a preliminary analysis on the robustness of deep face recognition systems to temporal variations. First, we analysed the evolution of the euclidean distance between samples of the same user and samples of different users. In Fig. 4.4 we reported the evolution of the distances between the first user template and the other genuine samples in green, between the first template and the

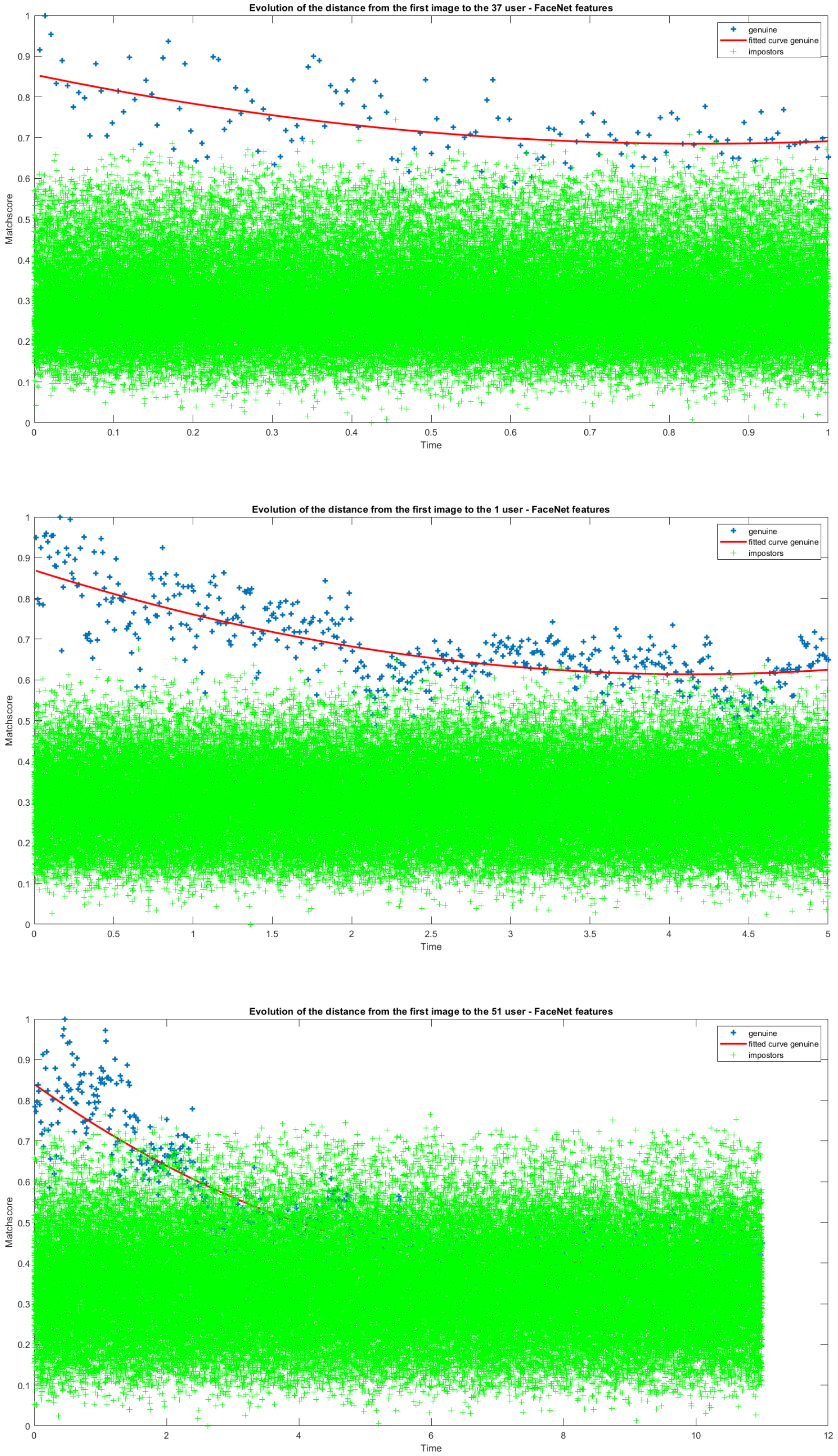


FIGURE 4.4: Evolution of the distances between the first user template and the other genuine samples in green, between the first template and the samples of the other users (impostors) in red using the FaceNet features of the APE dataset.

	Max recognized seq.	Max unrecognized seq.	Recognition acc.
Short time	47.70%	0,84%	92.13%
Medium time	37.47%	0.80%	93,58%
Long time	28.90%	1.71%	82.33%

TABLE 4.2: Average percentages of recognised and unrecognised sequences and the recognition accuracy of the samples of the same category (short, medium, long acquisition time).

samples of the other users (impostors) in red, using the FaceNet features of the APE dataset for 3 users, one for short, medium and long category.

Although the evolution of distances changes from user to user based on more or less accentuated intra-class variations, it is noted that as the acquisition time increases, the first acquired template loses its representativeness and is unable to follow the random and non-gradual manifestation of changes in the facial appearance.

We then evaluated the performance of a facial recognition system based on the minimum Euclidean distance of the incoming sample with those of the gallery (five samples per user). Figure 4.5 shows the analysis of the recognition sequences over time for the three users chosen to represent the three acquisition categories (short, medium and long time). In the graphs, the presence of a rectangle is equivalent to a continuous sequence of recognition. The height of this triangle indicates the number of correctly identified samples belonging to the sequence. The absence of the triangle is equivalent to a sequence of wrong classifications. The presence of small consecutive triangles is due to a succession of identifications and non-identifications. These graphs confirm that the templates do not maintain their representativeness over very long periods of time and therefore need to be replaced with more recent data. This is less true for users with short acquisition times, whose templates partially follow intra-class variations in face appearance.

Tab. 4.2 shows the average percentages of recognised and unrecognised sequences and the recognition accuracy of the samples of the same category. Recognition accuracy drops by ten percentage points for users with long acquisition times (more than five years). This is expected, as in these videos the variations of the facial appearance are multiple and very pronounced. On average, users with a medium capture time behave like those with short times. This makes us speculate that the variations under five years are controlled enough to be more easily followed by the compact and representative deep-features.

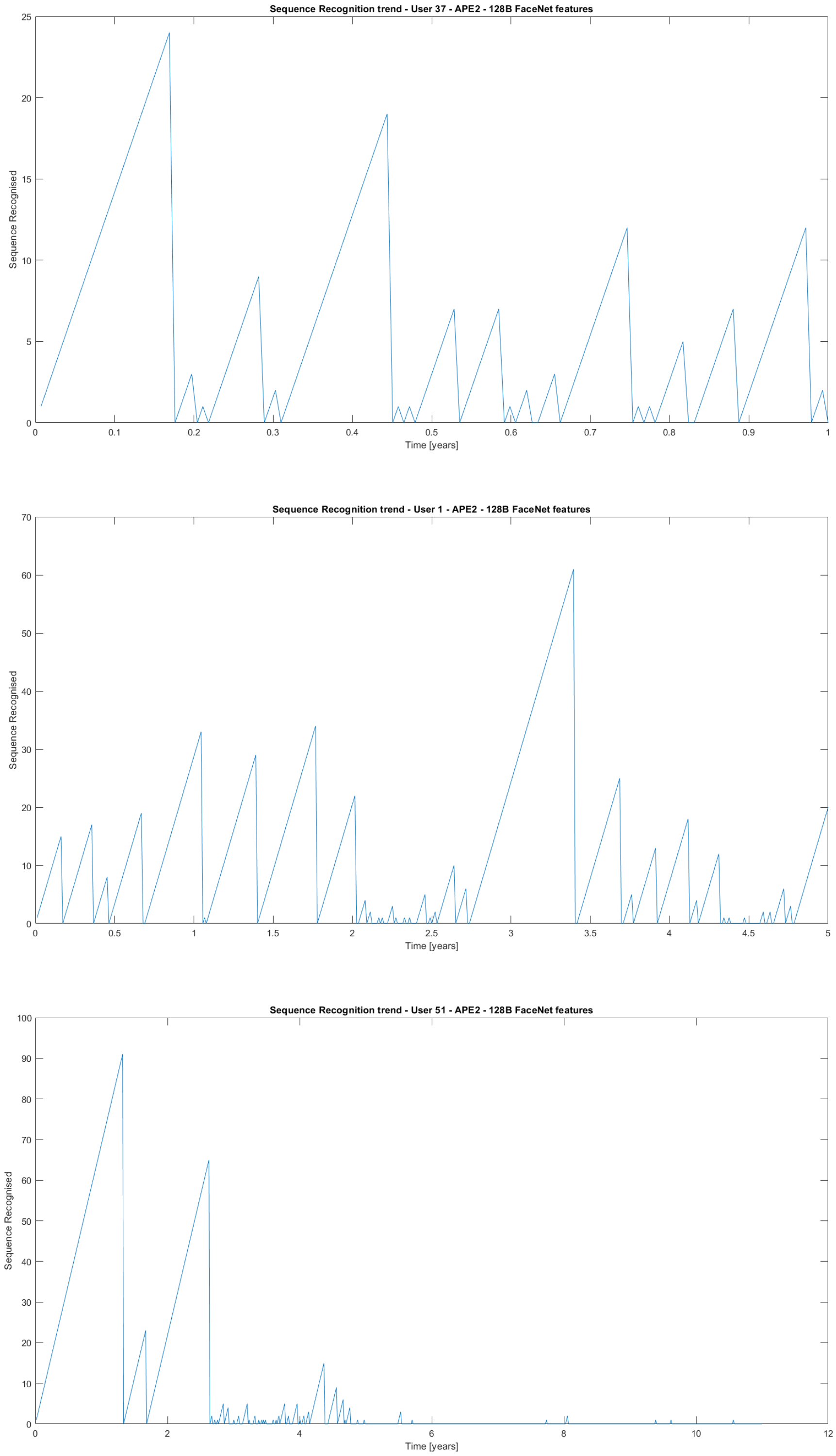


FIGURE 4.5: Analysis of the recognition sequences by Euclidean distance for three users of different categories (short, medium, long acquisition time). The base of each triangle represents a sequence of correct identifications, while the empty spaces represent a sequence of incorrect identification. The number of recognised sequences is indicated on the ordinate axis, which can therefore be associated with the vertex of the triangle, while the user acquisition period of the images is indicated on the abscissa axis. The presence of many triangles over a period of time implies that there can be even one misclassification that interrupts the series of correct identifications.

4.1.3 Results

The APE dataset allows us to evaluate the novel classification selection approach based on the clustering and the editing methods selection on data simulating realistic conditions of use.

In this evaluation, we added another classification-selection editing method, the RANDOM method, which simply selects p random samples of each user. This method allows us to simulate what may happen by keeping a human in the loop to perform a periodic template update. Indeed, the selection of a human supervisor is unpredictable and changes depending on the individual involved.

The traditional self-update [59], the method based on risk-minimization [11], the semi-supervised PCA [10], the K-Means [60], and the editing classification/selection methods MDIST, DEND and RANDOM [9, 19] were implemented and tested in terms of EER (Fig. 4.6) and impostors percentage (Fig. 4.7 using feature vectors extracted with FaceNet, ResNet50 and SeNet50 embeddings. These graphs allow us to analyse the performance of adaptive systems on heterogeneous data with different acquisition times, number of samples per user and presence of intra-class variations. Although the networks are particularly powerful and performing, the optimised update methods (red and black lines) present an improvement over the system without updating (magenta line) and the traditional self-update methods (green line).

The results confirm those found through the datasets analysed in Section 3.3.5, also as regards the percentage of impostors who are inserted into the system.

However, these graphs do not show us how updating the system differs based on time of use. One of the goals of this work was in fact to understand if and to what extent an optimised update would improve a facial recognition system. For this reason, we analysed how the EER changes depending on the category of users on the basis of acquisition times for FaceNet (Fig. 4.8), ResNet50 (Fig. 4.9) and SeNet50 (Fig. 4.10) embeddings.

The results confirm that for all the deep features analysed, despite their compactness and representativeness, "optimised" classification/selection template update methods, namely K-Means and MDIST, allow to keep an error lower than other adaptive methods, including human supervision, or system without update (magenta line).

It is important to note that, in long-term situations, state-of-the-art methods show a deterioration in performance. To better highlight this finding, we reported in Fig.4.11 a direct comparison between the K-Means and system without updating for the three categories of acquisition times, short, medium, long. Despite being highlighted by all three graphs, the great utility of the classification-selection method emerges particularly from the graph relating to the FaceNet features. In fact, in addition to the lowest error of the K-Means already of the first batches, this remains constant for all batches, while the system without updating shows an increase in error as the number of batches increases. This aspect shows that without updating the template gallery loses its representativeness over time.

In Figures 4.14 we reported how the gallery changes with the K-Means and random classification-selection methods of the same users that were used in the Section 4.1.2 analysis. The images of the impostors entered into the system show a red box around them. The random method, used to simulate what may happen by keeping a human in the loop for selecting the best templates to update, allows the insertion of impostors into the system. Templates selected with K-Means are similar to each other and, being close to the centroid, do not belong to impostors. The graphs allow appreciating how these methods with a template limit per user, even if very low as in this case with $p = 3$, perfectly follow the natural change in the appearance of the face. With each update, the templates in the gallery show the ageing of the biometric trait and do not contain informationally "old" samples.

In conclusion, the evidence from this study shows the benefits of update methods with "optimised" classification-selection in situations where the face appearance presents many intra-class variations.

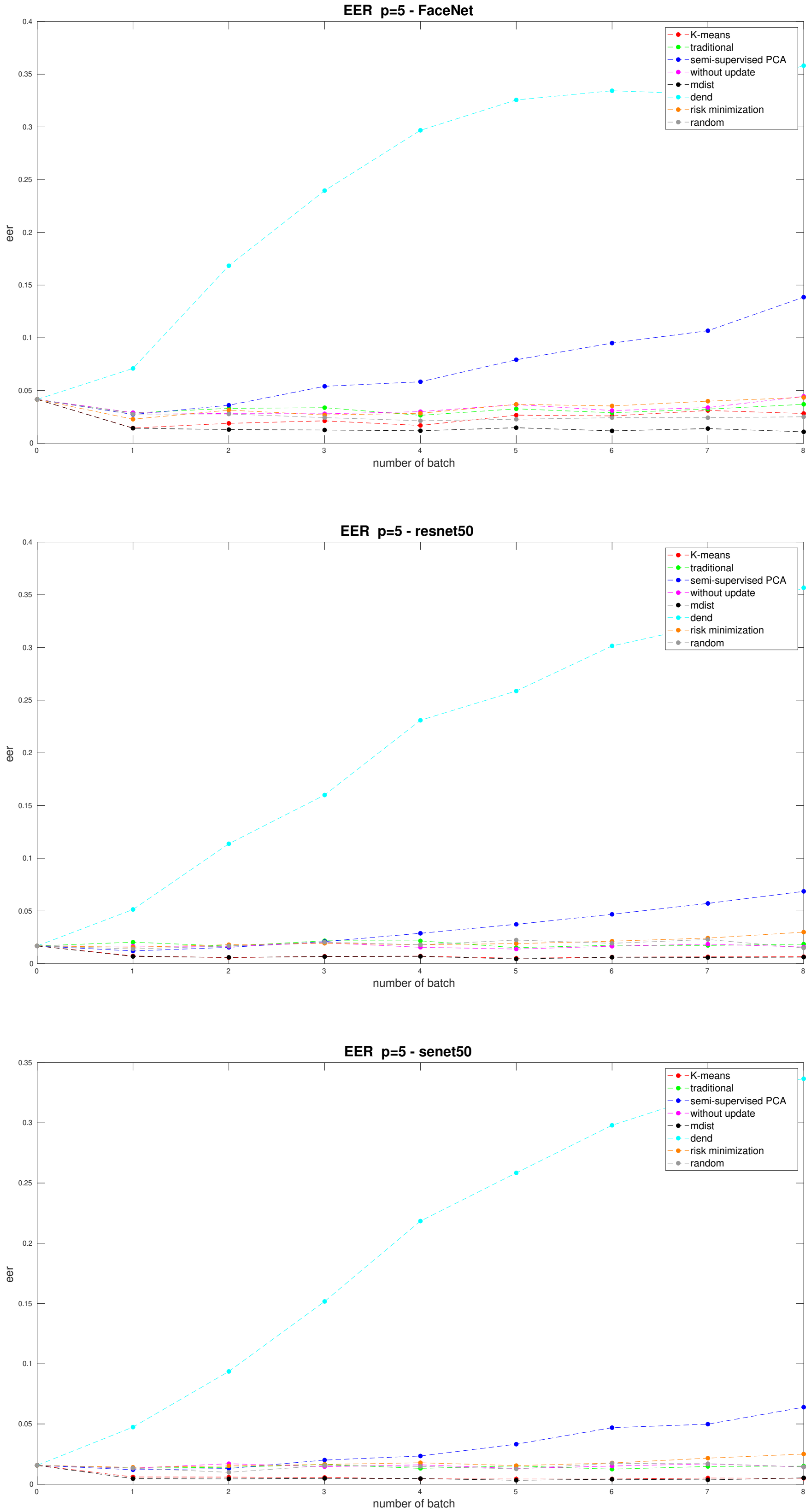
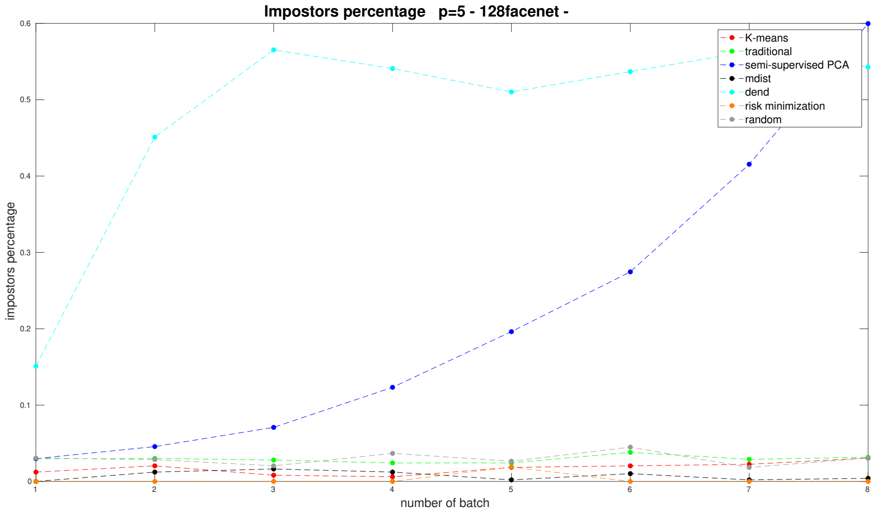
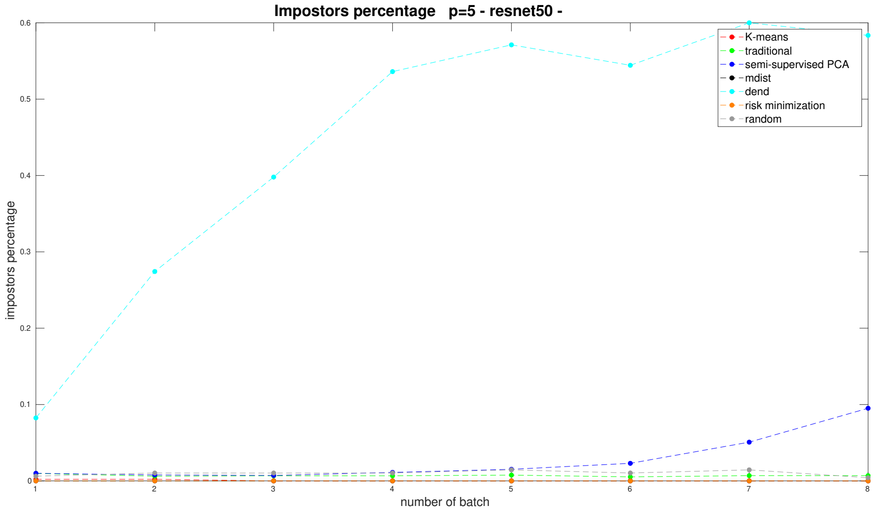


FIGURE 4.6: EER for different template update methods with p=5 using FaceNet/ResNet50/SeNet50 auto-encoded features on the APE dataset.





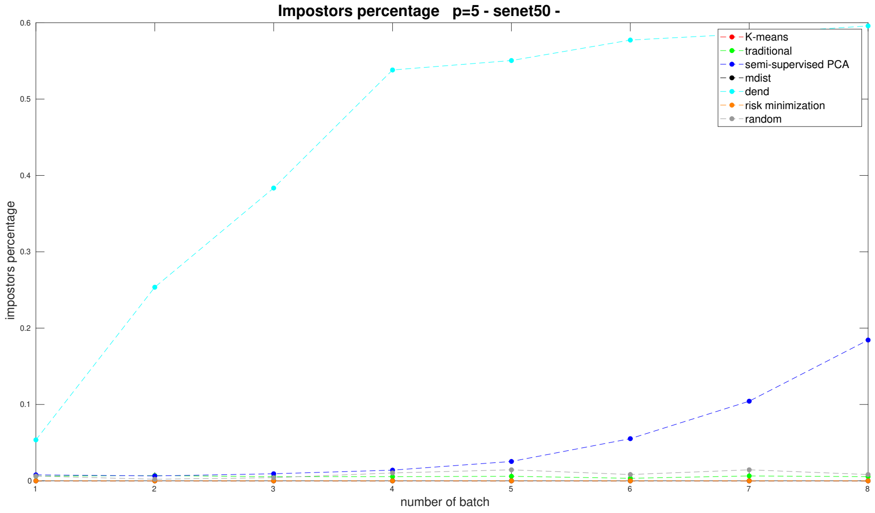
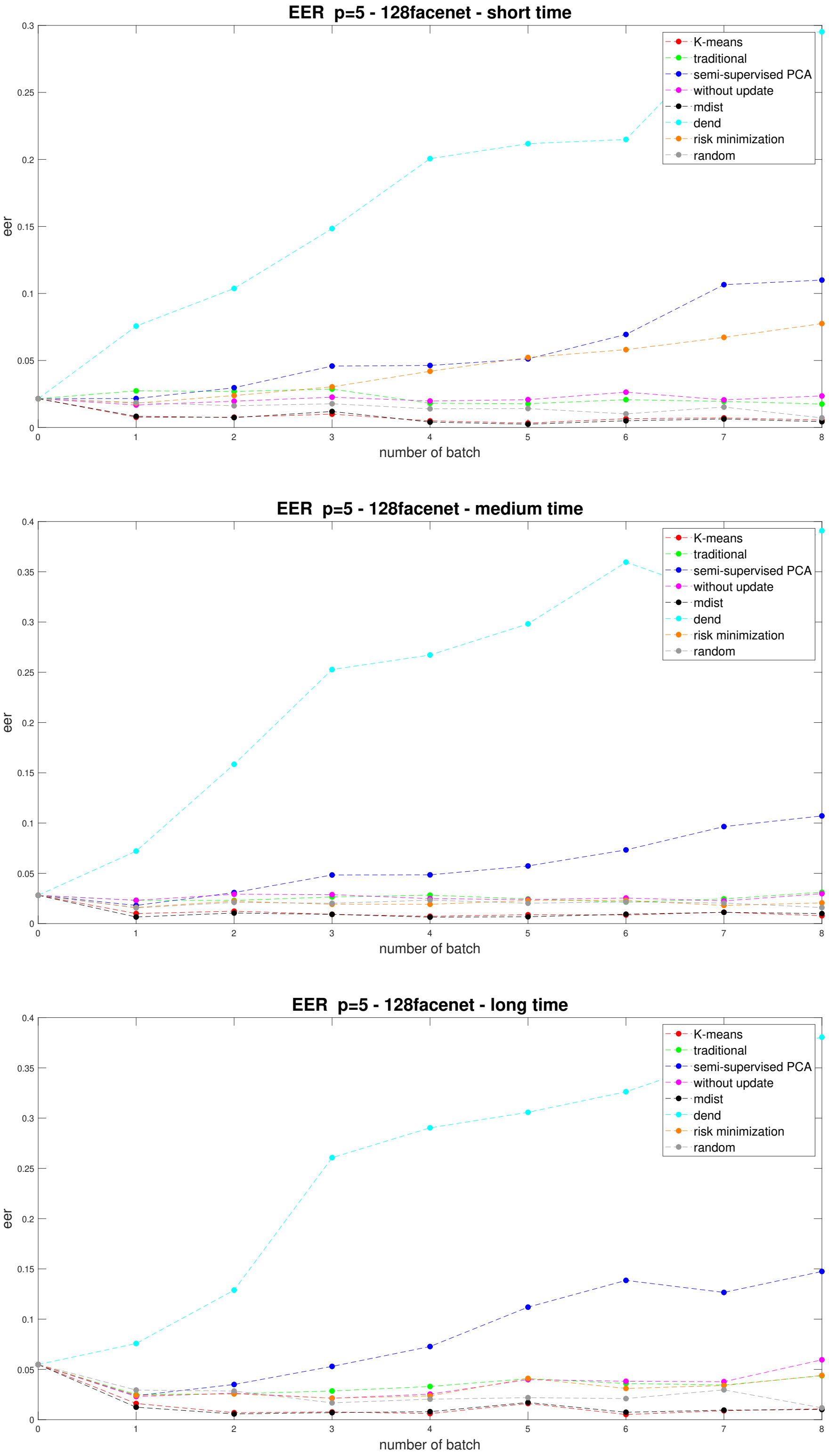


FIGURE 4.7: Impostors percentage for different template update methods with p=5 using FaceNet/ResNet50/SeNet50 auto-encoded features on the APE dataset.



EER p=5 - 128facenet - short time

EER p=5 - 128facenet - medium time

EER p=5 - 128facenet - long time

FIGURE 4.8: EER for different template update methods with p=5 using FaceNet auto-encoded features on the APE dataset divided on the basis of acquisition times.

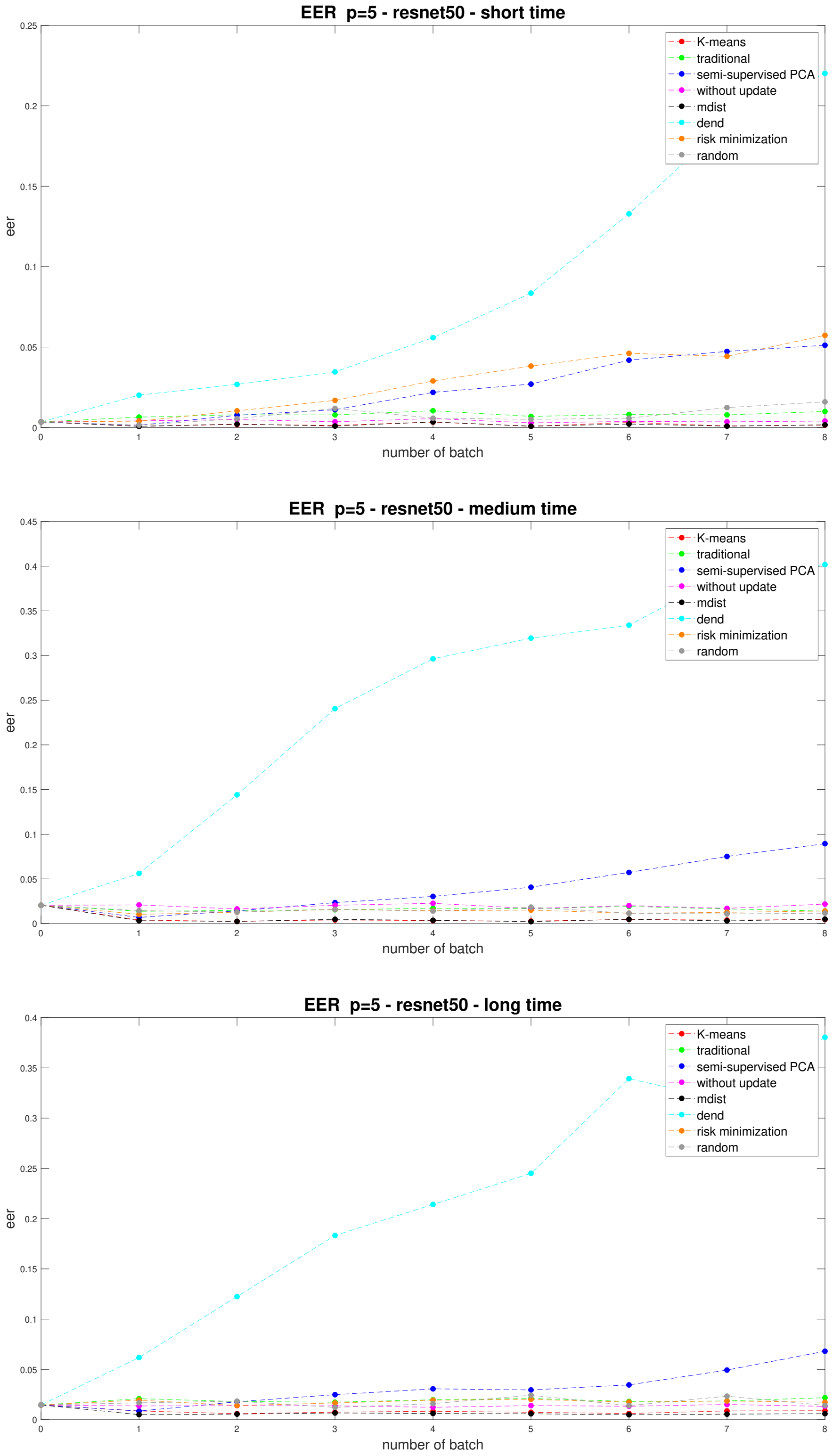


FIGURE 4.9: EER for different template update methods with *p*=5 using ResNet50 auto-encoded features on the APE dataset divided on the basis of acquisition times.

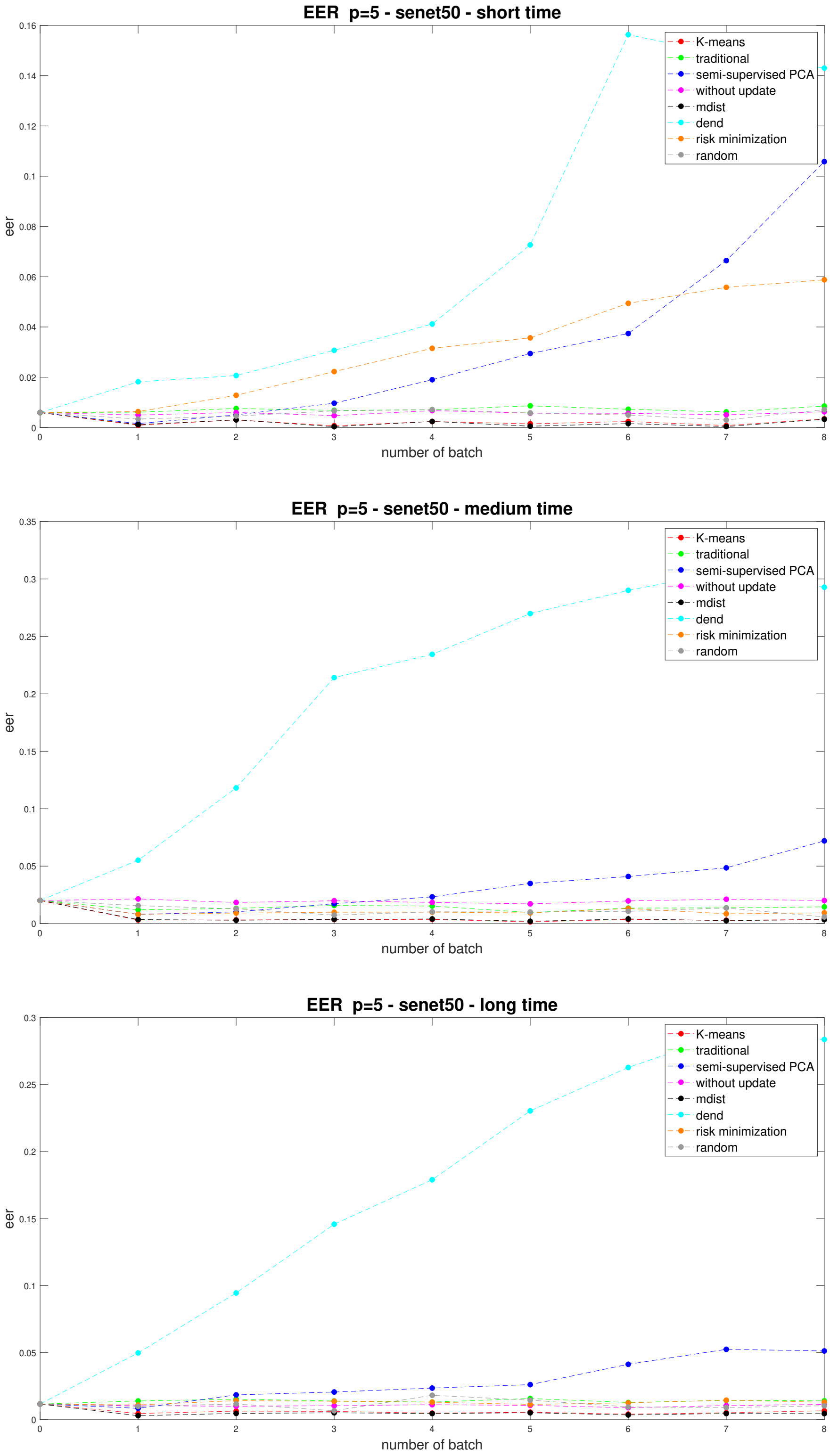
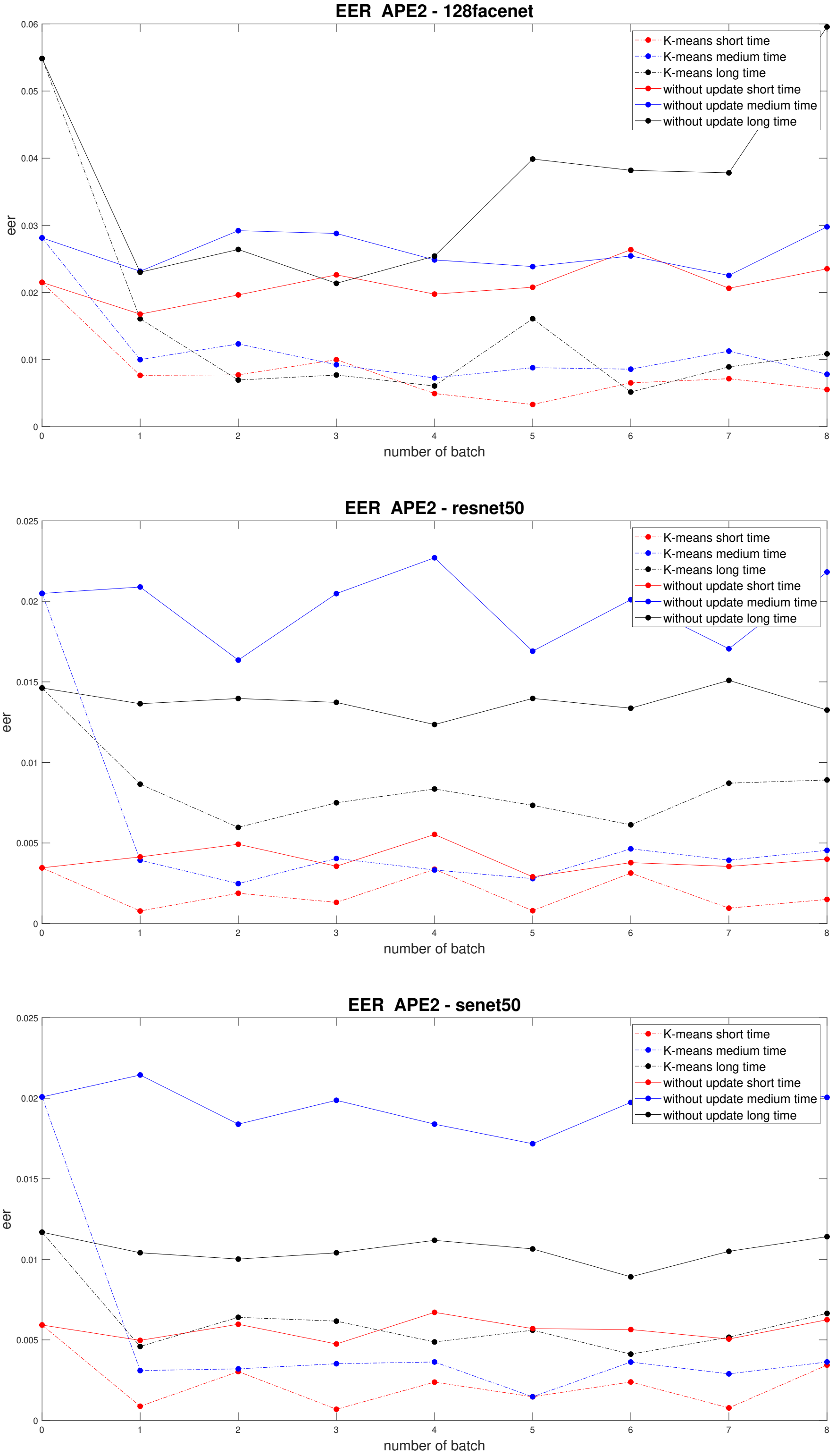


FIGURE 4.10: EER for different template update methods with p=5 using SeNet50 auto-encoded features on the APE dataset divided on the basis of acquisition times.


EER APE2 - 128facenet

EER APE2 - resnet50

EER APE2 - senet50

FIGURE 4.11: Comparison of EER between K-Means and system without updating for the three categories of acquisition times. For all the three auto-encoded features the use of "optimised" template update allows a substantial improvement in the performance compared to systems without updating.

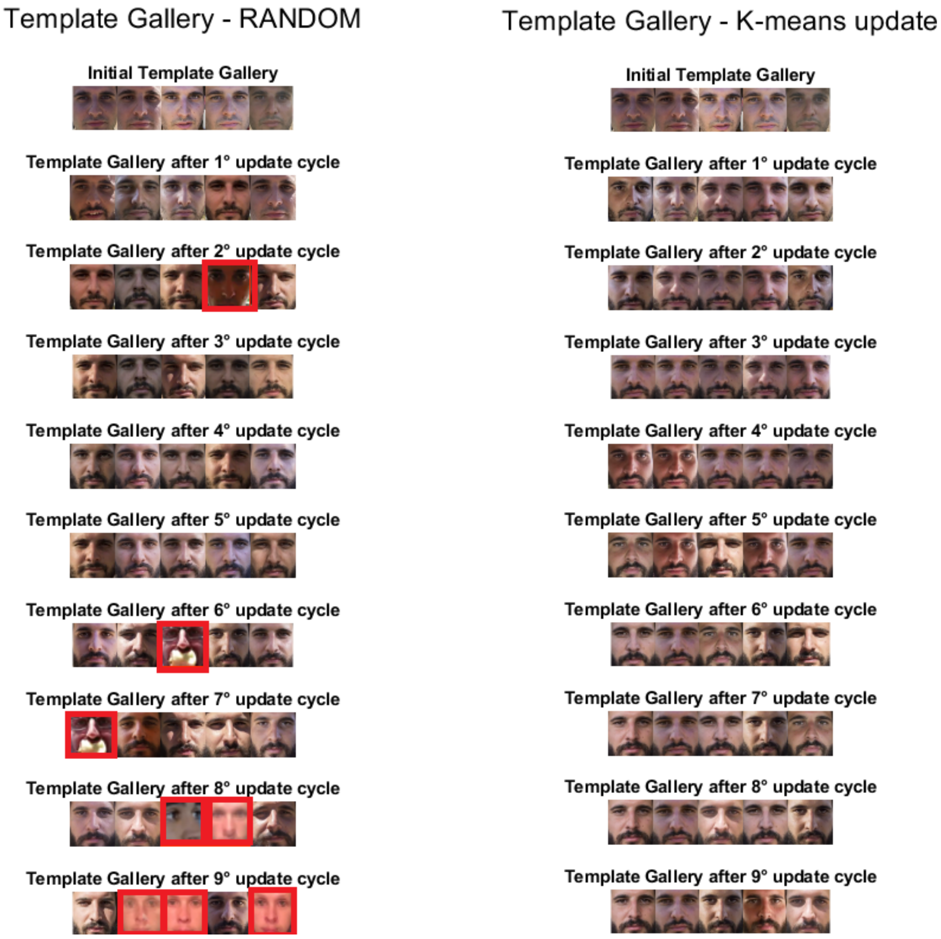

Template Gallery - RANDOM
Template Gallery - K-means update
Initial Template Gallery
Template Gallery after 1° update cycle
Template Gallery after 2° update cycle
Template Gallery after 3° update cycle
Template Gallery after 4° update cycle
Template Gallery after 5° update cycle
Template Gallery after 6° update cycle
Template Gallery after 7° update cycle
Template Gallery after 8° update cycle
Template Gallery after 9° update cycle

FIGURE 4.12: Update of the template gallery by K-Means and random classification-selection for user 37. Since the user's images were acquired in one year, he is part of the "short time" category.

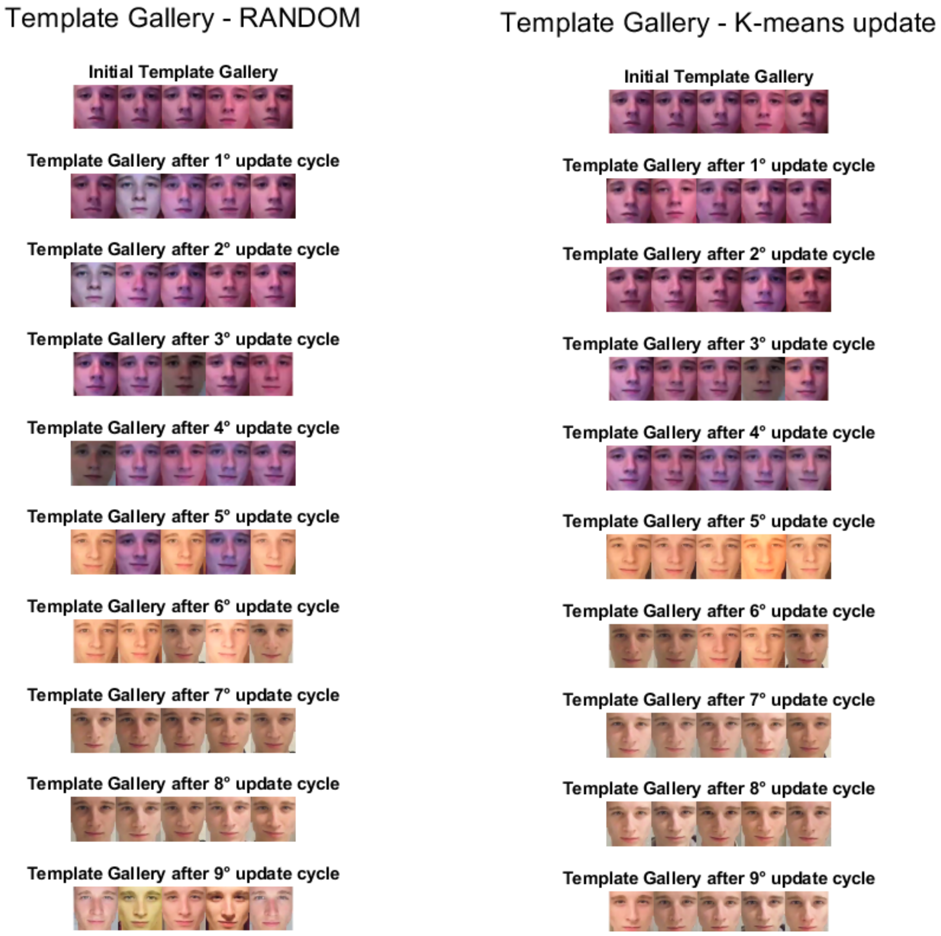

Template Gallery - RANDOM
Initial Template Gallery
Template Gallery after 1° update cycle
Template Gallery after 2° update cycle
Template Gallery after 3° update cycle
Template Gallery after 4° update cycle
Template Gallery after 5° update cycle
Template Gallery after 6° update cycle
Template Gallery after 7° update cycle
Template Gallery after 8° update cycle
Template Gallery after 9° update cycle
Template Gallery - K-means update
Initial Template Gallery
Template Gallery after 1° update cycle
Template Gallery after 2° update cycle
Template Gallery after 3° update cycle
Template Gallery after 4° update cycle
Template Gallery after 5° update cycle
Template Gallery after 6° update cycle
Template Gallery after 7° update cycle
Template Gallery after 8° update cycle
Template Gallery after 9° update cycle

FIGURE 4.13: Update of the template gallery by K-Means and random classification-selection for user 1. Since the user's images were acquired in five years, he is part of the "medium time" category.

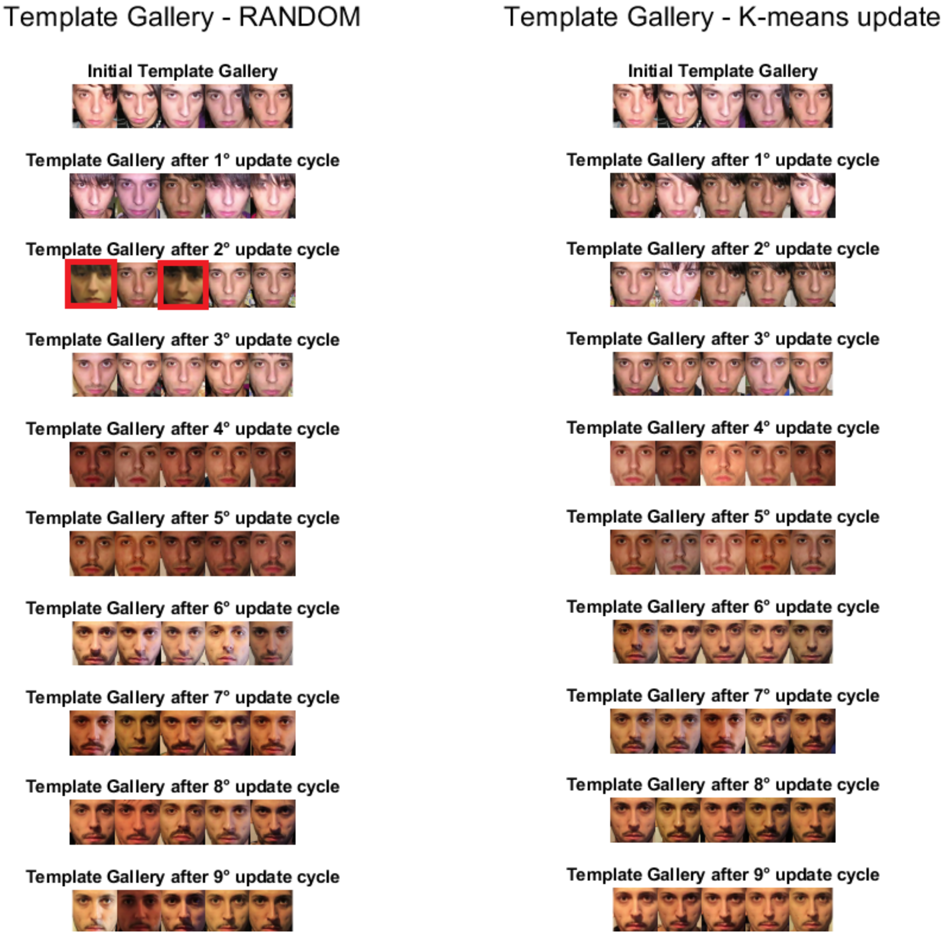

Template Gallery - RANDOM
Template Gallery - K-means update
Initial Template Gallery
Template Gallery after 1° update cycle
Template Gallery after 2° update cycle
Template Gallery after 3° update cycle
Template Gallery after 4° update cycle
Template Gallery after 5° update cycle
Template Gallery after 6° update cycle
Template Gallery after 7° update cycle
Template Gallery after 8° update cycle
Template Gallery after 9° update cycle

FIGURE 4.14: Update of the template gallery by K-Means and random classification-selection for user 51. Since the user's images were acquired in eleven years, he is part of the "long time" category.

Chapter 5

Conclusions

Compactness and expressiveness are the main strength points of the most feature sets extracted through modern deep-learning-based face recognition systems. In spite of the high progress, it is not yet clear to what extent deep features can embed all possible variations of the users' face. In this thesis, we investigate the performance improvement of face recognition systems by adopting self-updating strategies of the face templates and we propose a novel classification-selection approach with a maximum number of p templates per user to keep limited the storage and computational requirements. Based on the working hypothesis that the statistical distribution of the facial features exhibits a dominating mode around which the templates can be searched, we propose two different criteria to perform the templates selection, one based on clustering methods and one on editing methods.

The novel classification-selection self-update method showed excellent performance on different state-of-the-art datasets and using both handcrafted and auto-encoded feature sets. In particular, the proposed approach showed there is no need for re-training for auto-encoded features, which is computationally more expensive.

Although the results obtained are promising, the datasets used for this experimentation, as well as for previous works, present medium-short acquisition periods, few images per user and do not contain temporal image information. During my PhD, I acquired a novel facial dataset for evaluating the performance of several representative facial adaptive algorithms, called APhotoEveryday (APE). It contains temporary and temporal variations collected every day in extended periods, keeping the chronological order of the samples. The APE images are characterised by many variations because the videos from which they are extracted have the aim to demonstrate a change in the appearance of the individual. The APE dataset differs from the existing ones for the very high number of images per user, many variations of the faces and a very high time span.

These characteristics make it more suitable for simulating the continued use of adaptive face recognition algorithms. By exploiting its potentialities, we evaluated how some of the state-of-the-art neural networks worked under the random and non-gradual manifestation of changes in the facial appearance. Experimental results show the effectiveness of "optimized" adaptive methods concerning systems without an update or random selection of templates.

5.1 Related Publications

During the three years of my PhD, we have published the following publications:

5.1.1 Journal

- G. Orrù, G.L. Marcialis, F. Roli, "A novel classification-selection approach for the self updating of template-based face recognition systems", Pattern Recognition, Volume 100, 2020, 107121, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2019.107121.

5.1.2 Conference

- V. Mura, G. Orrù, R. Casula, A. Sibiriu, G. Loi, P. Tuveri, L. Ghiani, G.L. Marcialis, "LivDet 2017 Fingerprint Liveness Detection Competition 2017," 2018 International Conference on Biometrics (ICB), Gold Coast, QLD, 2018, pp. 297-302, doi: 10.1109/ICB2018.2018.00052.

- G. Orrù, G. L. Marcialis and F. Roli, "An experimental investigation on self adaptive facial recognition algorithms using a long time span data set," 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, 2018, pp. 1-6, doi: 10.1109/IPTA.2018.8608134.

- M. Micheletto, G. Orrù, I. Rida, L. Ghiani and G. L. Marcialis, "A multiple classifiers-based approach to palmvein identification," 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, 2018, pp. 1-6, doi: 10.1109/IPTA.2018.8608150.

- G. Orrù et al., "LivDet in Action - Fingerprint Liveness Detection Competition 2019," 2019 International Conference on Biometrics (ICB), Crete, Greece, 2019, pp. 1-6, doi: 10.1109/ICB45273.2019.8987281.

- G. Orrú et al., "Personal Identity Verification by EEG-Based Network Representation on a Portable Device," 2019 In: Vento M., Percannella G. (eds) Computer Analysis of Images and Patterns. CAIP 2019. Lecture Notes in Computer Science, vol 11679. Springer, Cham, https://doi.org/10.1007/978-3-030-29891-3_15.

- G. Orrù, P. Tuveri, L. Ghiani, G.L. Marcialis "Analysis of "User-Specific Effect" and Impact of Operator Skills on Fingerprint PAD Systems," 2019 In: Cristani M., Prati A., Lanz O., Messelodi S., Sebe N. (eds) New Trends in Image Analysis and Processing – ICIAP 2019. ICIAP

2019. Lecture Notes in Computer Science, vol 11808. Springer, Cham, https://doi.org/10.1007/978-3-030-30754-7_6.

- R. Casula, G. Orrù, D. Angioni, X. Feng, G.L. Marcialis, F. Roli, "Are spoofs from latent fingerprints a real threat for the best state-of-art liveness detectors?," 2020 25th International Conference on Pattern Recognition (ICPR 2020), Milan, in press.

- G. Orrù, M. Micheletto, J. Fierrez, G. L. Marcialis, "Are Adaptive Face Recognition Systems still Necessary?", 2020 Fourth IEEE International Conference on Image Processing, Applications and Systems (IPAS 2020), in press.

- G. Orrù, M. Micheletto, F. Terranova, G. L. Marcialis, "Electroencephalography signal processing based on textural features for monitoring the driver's state by a Brain-Computer Interface" In: 2020 IEEE 4th International Con-ference on Image Processing, Applications and Systems (IPAS). 2020, pp. 77–82.DOI:10.1109/IPAS50080.2020.9334946.

- G. Orrù, D. Ghiani, M. Pintor, G.L. Marcialis, F. Roli, "Detecting Anomalies from Video-Sequences: a Novel Descriptor" 2020 25th International Conference on Pattern Recognition (ICPR 2020), Milan, in press.

- S. Marrone, R. Casula, G. Orrù, G.L. Marcialis, C. Sansone, "Fingerprint Adversarial PresentationAttack in the Physical Domain" ICPR 2021 Workshops, LNCS 12666, pp. 1–14, 2021, doi: 10.1007/978-3-030-68780-9_42.

Bibliography

[1] Mostafa Mehdipour-Ghazi and Hazim Kemal Ekenel. "A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2016), pp. 102–109.

[2] N. Poh, A. Rattani, and F. Roli. "Critical analysis of adaptive biometric systems". In: *IET Biometrics* 1.4 (Dec. 2012), pp. 179–187. ISSN: 2047-4938. DOI: 10.1049/iet-bmt.2012.0019.

[3] Umut Uludag, Arun Ross, and Anil Jain. "Biometric template selection and update: a case study in fingerprints". In: *Pattern Recognition* 37.7 (2004), pp. 1533–1542. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2003.11.012. URL: http://www.sciencedirect.com/science/article/pii/S0031320304000081.

[4] Hugo Proença et al. "Trends and Controversies". In: *IEEE Intelligent Systems* 33.3 (2018), pp. 41–67. DOI: https://doi.org/10.1109/MIS.2018.033001416.

[5] E. Gonzalez-Sosa et al. "Exploring Facial Regions in Unconstrained Scenarios: Experience on ICB-RW". In: *IEEE Intelligent Systems* 33.3 (May 2018), pp. 60–63. DOI: https://doi.org/10.1109/MIS.2018.033001416.

[6] J. Fierrez-Aguilar et al. "Adapted user-dependent multimodal biometric authentication exploiting general information". In: *Pattern Recognition Letters* 26.16 (Dec. 2005), pp. 2628–2639.

[7] Fabio Roli, Luca Didaci, and Gian Luca Marcialis. "Adaptive Biometric Systems That Can Improve with Use". In: *Advances in Biometrics: Sensors, Algorithms and Systems*. Ed. by Nalini K. Ratha and Venu Govindaraju. Springer London, 2008, pp. 447–471. ISBN: 978-1-84628-921-7. DOI: 10.1007/978-1-84628-921-7_23.

[8] Julian Fierrez et al. "Multiple Classifiers in Biometrics. Part 2: Trends and Challenges". In: *Information Fusion* 44 (Nov. 2018), pp. 103–112. DOI: https://doi.org/10.1016/j.inffus.2017.12.005.

[9] Giulia Orrù, Gian Luca Marcialis, and Fabio Roli. "A novel classification-selection approach for the self updating of template-based face recognition systems". In: *Pattern Recognition* 100 (2020), pp. 107–121. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.107121.

[10] Fabio Roli and Gian Luca Marcialis. "Semi-supervised PCA-Based Face Recognition Using Self-training". In: *Structural, Syntactic, and Statistical Pattern Recognition*. Ed. by Dit-Yan Yeung et al. Springer Berlin Heidelberg, 2006, pp. 560–568. ISBN: 978-3-540-37241-7.

[11] A. Rattani et al. “A dual-staged classification-selection approach for automated update of biometric templates”. In: *Proceedings of the 21st International Conference on Pattern Recognition*. Nov. 2012, pp. 2972–2975.

[12] Biagio Freni, Gian Luca Marcialis, and Fabio Roli. “Replacement Algorithms for Fingerprint Template Update”. In: *Proceedings of the 5th International Conference on Image Analysis and Recognition*. ICIAR ’08. P=voa de Varzim, Portugal: Springer-Verlag, 2008, pp. 884–893. ISBN: 978-3-540-69811-1. DOI: 10.1007/978-3-540-69812-8_88.

[13] Biagio Freni, Gian Luca Marcialis, and Fabio Roli. “Template Selection by Editing Algorithms: A Case Study in Face Recognition”. In: *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, SSPR & SPR 2008, Orlando, USA, December 4-6, 2008. Proceedings*. Ed. by Niels da Vitoria Lobo et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 745–754. ISBN: 978-3-540-89689-0. DOI: 10.1007/978-3-540-89689-0_78. URL: http://dx.doi.org/10.1007/978-3-540-89689-0_78.

[14] J. Kannala and E. Rahtu. “BSIF: Binarized statistical image features”. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. Nov. 2012, pp. 1363–1366.

[15] F. Schroff, D. Kalenichenko, and J. Philbin. “FaceNet: A unified embedding for face recognition and clustering”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition*. June 2015, pp. 815–823. DOI: 10.1109/CVPR.2015.7298682.

[16] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 770–778.

[17] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-Excitation Networks”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2017), pp. 7132–7141.

[18] G. Orrù, G. L. Marcialis, and F. Roli. “An experimental investigation on self adaptive facial recognition algorithms using a long time span data set”. In: *IEEE 8th Int. Conf. Image Processing, Theory, Tools and Applications*. Nov. 2018.

[19] G. Orrù et al. “Are Adaptive Face Recognition Systems still Necessary? Experiments on the APE Dataset”. In: *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*. 2020, pp. 77–82. DOI: 10.1109/IPAS50080.2020.9334946.

[20] A.K. Jain, P. Flynn, and A.A. Ross. *Handbook of Biometrics*. Springer US, 2007. ISBN: 9780387710419. DOI: 10.1007/978-0-387-71041-9.

[21] Ruud Bolle and Sharath Pankanti. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Ed. by Anil K. Jain. Norwell, MA, USA: Kluwer Academic Publishers, 1998. ISBN: 0792383451.

[22] A. K. Jain, A. Ross, and S. Prabhakar. "An Introduction to Biometric Recognition". In: *IEEE Trans. Cir. and Sys. for Video Technol.* 14.1 (Jan. 2004), pp. 4–20. ISSN: 1051-8215. DOI: 10.1109/TCSVT.2003.818349. URL: http://dx.doi.org/10.1109/TCSVT.2003.818349.

[23] Proyecto Fin de Carrera. *Facial Recognition Algorithms*. 2010.

[24] Anil K. Jain and Stan Z. Li. *Handbook of Face Recognition*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005. ISBN: 038740595X.

[25] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. 2001, pp. I–I. DOI: 10.1109/CVPR.2001.990517.

[26] Jens Rettkowski, Andrew Boutros, and Diana Göhringer. "HW/SW Co-Design of the HOG algorithm on a Xilinx Zynq SoC". In: *Journal of Parallel and Distributed Computing* 109 (2017), pp. 50–62. ISSN: 0743-7315. DOI: https://doi.org/10.1016/j.jpdc.2017.05.005. URL: http://www.sciencedirect.com/science/article/pii/S0743731517301569.

[27] K. Zhang et al. "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks". In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503. DOI: 10.1109/LSP.2016.2603342.

[28] Mei Wang and Weihong Deng. "Deep Face Recognition: A Survey". In: *CoRR* abs/1804.06655 (2018).

[29] Mejda Chihaoui et al. "A Survey of 2D Face Recognition Techniques". In: *Computers* 5.4 (2016). ISSN: 2073-431X. DOI: 10.3390/computers5040021. URL: https://www.mdpi.com/2073-431X/5/4/21.

[30] M. A. Turk and A. P. Pentland. "Face recognition using eigenfaces". In: *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 1991, pp. 586–591. DOI: 10.1109/CVPR.1991.139758.

[31] J. Lu et al. "Transform-Invariant PCA: A Unified Approach to Fully Automatic Face Alignment, Representation, and Recognition". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* (Nov. 2013). ISSN: 0162-8828. DOI: 10.1109/TPAMI.2013.194. URL: doi.ieeecomputersociety.org/10.1109/TPAMI.2013.194.

[32] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995. ISBN: 0387945598.

[33] Heiko Hoffmann. "Kernel PCA for Novelty Detection". In: *Pattern Recogn.* 40.3 (Mar. 2007), pp. 863–874. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2006.07.009. URL: https://doi.org/10.1016/j.patcog.2006.07.009.

[34] T. Ojala, M. Pietikainen, and T. Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.7 (July 2002), pp. 971–987. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2002.1017623.

[35] T. Ahonen, A. Hadid, and M. Pietikainen. “Face Description with Local Binary Patterns: Application to Face Recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.12 (Dec. 2006), pp. 2037–2041. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2006.244.

[36] Chengjun Liu and H. Wechsler. “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition”. In: *IEEE Transactions on Image Processing* 11.4 (Apr. 2002), pp. 467–476. ISSN: 1057-7149. DOI: 10.1109/TIP.2002.999679.

[37] Wenchao Zhang et al. “Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition”. In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Oct. 2005, pp. 786–791. DOI: 10.1109/ICCV.2005.147.

[38] Z. Cao et al. “Face recognition with learning-based descriptor”. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 2010, pp. 2707–2714. DOI: 10.1109/CVPR.2010.5539992.

[39] Changxing Ding and Dacheng Tao. “A Comprehensive Survey on Pose-Invariant Face Recognition”. In: *ACM Trans. Intell. Syst. Technol.* 7.3 (Feb. 2016), 37:1–37:42. ISSN: 2157-6904. DOI: 10.1145/2845089. URL: http://doi.acm.org/10.1145/2845089.

[40] Athanasios Voulodimos et al. “Deep Learning for Computer Vision: A Brief Review”. In: *Computational intelligence and neuroscience* 2018 (2018), p. 7068349. ISSN: 1687-5265. DOI: 10.1155/2018/7068349. URL: https://europepmc.org/articles/PMC5816885.

[41] Y. Taigman et al. “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 1701–1708. DOI: 10.1109/CVPR.2014.220.

[42] O. M. Parkhi, A. Vedaldi, and A. Zisserman. “Deep Face Recognition”. In: *British Machine Vision Conference*. 2015.

[43] Qiong Cao et al. “VGGFace2: A dataset for recognising faces across pose and age”. In: *CoRR* abs/1710.08092 (2017).

[44] Gary B. Huang et al. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Tech. rep. 07-49. University of Massachusetts, Amherst, Oct. 2007.

[45] B. F. Klare et al. “Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A”. In: *IEEE Conf. Computer Vision and Pattern Recognition*. June 2015, pp. 1931–1939. DOI: 10.1109/CVPR.2015.7298803.

[46] C. Whitelam et al. “IARPA Janus Benchmark-B Face Dataset”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. July 2017, pp. 592–600. DOI: 10.1109/CVPRW.2017.87.

[47] Shuai Ruan et al. "Multi-Pose Face Recognition Based on Deep Learning in Unconstrained Scene". In: *Applied Sciences* 10 (July 2020), p. 4669. DOI: 10.3390/app10134669.

[48] Guodong Guo and Na Zhang. "A survey on deep learning based face recognition". In: *Computer Vision and Image Understanding* 189 (2019), p. 102805. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2019.102805. URL: http://www.sciencedirect.com/science/article/pii/S1077314219301183.

[49] Ignacio Serna et al. "Algorithmic discrimination: Formulation and exploration in deep learning-based face biometrics". In: *arXiv preprint arXiv:1912.01842* (2019).

[50] A. George and S. Marcel. "Deep Pixel-wise Binary Supervision for Face Presentation Attack Detection". In: *2019 International Conference on Biometrics (ICB)*. 2019, pp. 1–8. DOI: 10.1109/ICB45273.2019.8987370.

[51] Brandon RichardWebster et al. "Visual psychophysics for making face recognition algorithms more explainable". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 252–270.

[52] C. Pagano et al. "Context-Sensitive Self-Updating for Adaptive Face Recognition". In: *Adaptive Biometric Systems: Recent Advances and Challenges*. Ed. by Ajita Rattani, Fabio Roli, and Eric Granger. Cham: Springer International Publishing, 2015, pp. 9–34. ISBN: 978-3-319-24865-3. DOI: 10.1007/978-3-319-24865-3_2.

[53] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. "A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study". In: *IJBM* 5 (2013), pp. 266–287.

[54] Xiaoming Liu, Tsuhan Chen, and Susan M. Thornton. "Eigenspace updating for non-stationary process and its application to face recognition". In: *Pattern Recognition* 36.9 (2003). Kernel and Subspace Methods for Computer Vision, pp. 1945–1959. ISSN: 0031-3203.

[55] Haitao Zhao et al. "Incremental PCA based face recognition". In: *ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, 2004.* Vol. 1. Dec. 2004, 687–691 Vol. 1. DOI: 10.1109/ICARCV.2004.1468910.

[56] P. Jonathon Phillips et al. "Overview of the Face Recognition Grand Challenge". In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. CVPR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 947–954. ISBN: 0-7695-2372-2. DOI: 10.1109/CVPR.2005.268.

[57] Yandong Guo et al. "MS-Celeb-1M: A Dataset and Benchmark for Large Scale Face Recognition". In: *European Conference on Computer Vision*. Springer. 2016.

[58] Qiong Cao et al. "VGGFace2: A Dataset for Recognising Faces across Pose and Age". In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (2017), pp. 67–74.

[59] Ajita Rattani et al. "Template Update Methods in Adaptive Biometric Systems: A Critical Review". In: *Advances in Biometrics*. Ed. by Massimo Tistarelli and Mark S. Nixon. Springer Berlin Heidelberg, 2009, pp. 847–856. ISBN: 978-3-642-01793-3.

[60] Pierluigi Tuveri et al. "A classification-selection approach for self updating of face verification systems under stringent storage and computational requirements". In: *18th IAPR Int. Conf. on Image Analysis and Processing*. Vol. 9280. 2015, pp. 540–550. DOI: 10.1007/978-3-319-23234-850.

[61] Rodney Goh et al. "The CMU Face In Action (FIA) Database". In: *Analysis and Modelling of Faces and Gestures*. Ed. by Wenyi Zhao, Shaogang Gong, and Xiaoou Tang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 255–263. ISBN: 978-3-540-32074-6.

[62] A. Martinez and Robert Benavente. "The AR face database". In: *Tech. Rep. 24 CVC Technical Report* (Jan. 1998).

[63] B. F. Klare et al. "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1931–1939. DOI: 10.1109/CVPR.2015.7298803.

[64] A. Rattani, G. Luca Marcialis, and F. Roli. "Self adaptive systems: An experimental analysis of the performance over time". In: *2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*. Apr. 2011, pp. 36–43. DOI: 10.1109/CIBIM.2011.5949222.