Analysis of unsupervised template update in biometric recognition systems

Luca Didaci *, Gian Luca Marcialis, Fabio Roli
Department of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, 09123 Cagliari, Italy

A R T I C L E   I N F O

Article history:
Available online 1 June 2013

Keywords:
Biometrics recognition
Adaptive systems
Face
Fingerprint
Multi-modal biometrics
Path-based clustering

A B S T R A C T

Performance of mono- and multi-modal biometric systems depends on the representativeness of enrolled templates. Unfortunately, error rate values estimated during the system design are subject to variations due to several aspects: intra-class variations arising on small-medium time-window, and ageing, which is the natural process involving any biometrics. This causes the increase of the False Rejection Rate (genuine users are no more recognized) or the False Acceptance Rate (impostors are misclassified as genuine users), or both. In fact, several vendors strongly suggest to repeat enrolment sessions in order to collect, over time, a set of templates representative enough. As alternative, automatic template update algorithms, which exploit the own-knowledge of the mono- or multi-modal biometric system, on a batch of samples collected during system operations without the human supervision, have been proposed. Preliminary experimental results have shown that these algorithms are promising, but the motivation of their behaviour has not yet been explained. This paper is aimed to fill such gap, by showing that behaviour of self- and co-update may be explained by exploiting the concept of path-based clustering. Therefore, problems as 'intra-class' variations and ageing are dependent on the path-based cluster followed by each algorithm. Moreover, we show that the performance of co-update is superior than that of self-update, by a simulative model. The path-based clustering theory applied to self- and co-update algorithms, as well as the proposed model, are experimentally validated on the large DIEE Multimodal data set, the only one publicly available and explicitly conceived for comparing template update algorithms.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In biometric verification systems, the identity of the user, also called 'client' or 'genuine user', is stored and associated to the related template(s).

Biometric templates are structural or statistical representations of the client's biometric obtained by image or signal processing techniques (Jain et al., 2008). The client's biometric is initially given by images or signals acquired by a certain sensor (face snapshot, voice registration, fingerprint image). The acquisition and computation phase is called 'enrolment'. Templates are stored into the system's data set, and used for comparison with novel input images or signals during system's operations.

Among open problems, updating of 'templates' is crucial (Jain et al., 2008; Infosecurity, 2008). In particular, an appropriate selection of template(s) may increase the system performance, in terms of False Rejection Rate (FRR: percentage of genuine users not recognized) or false acceptance rate (FAR: percentage of unknown users, also called impostors, misclassified as genuine users).

Beside additional enrolment sessions, recent works argued that the genuine/impostor labelling could be done by the matcher itself, thus avoiding human or semi-automatic supervision. Algorithms which perform such labelling, and, consequently, allow to add into the client gallery novel samples as templates, are called 'template update algorithms'. These approaches are inspired from machine learning methods called 'self-training' and 'co-training' (Zhu, 2006; Blum and Mitchell, 1998). In order to keep this link, authors have called them 'self updating' and 'co-updating', respectively (Roli et al., 2008). Therefore, the authors have proposed the application of self- and co-training to biometrics (Roli and Marcialis, 2006; Roli et al., 2007), for which an ad hoc analysis is necessary.

Main assumptions adopted by Blum and Mitchell (1998) are compatibility between classifiers, sufficiency of the views and independence between the views. Sufficiency and compatibility refers to the fact that, when provided with a sufficient number of training patterns, each view is sufficient for a correct classification (sufficiency) and classifiers agree on the label to be assigned (compatibility). Independence will be discussed in Section 2. In general, several authors have theoretically or empirically investigated a number of aspects of co-training, by focusing on how it works under different and less restrictive assumptions than those adopted by Blum and Mitchel. See, for example, Didaci and Roli (2012), Du et al. (2011), Zhou et al. (2007) and Didaci et al. (2012).
The current state-of-the-art on self- and co-update is lacking of an analogous literature in biometric applications, that’s means that there are no papers, to the best of our knowledge, where the self- and co-training algorithms were specifically adapted for the application, or an ad hoc theory has been developed. With regard to biometric applications, previous papers investigated the effect of the typology of enrolled clients on self-update in the sense of the Doddington’s zoo (Doddington et al., 1998; Rattani et al., 2009), and benefits have been pointed out for clients intrinsically prone to low FRR. However, experiments were performed on a face data set made up of very low intra-class variations, thus conclusions were not definitive. Also, an experimental analysis of the self-update performance over time has been done in Marcialis et al. (2008) and Rattani et al. (2011), where this algorithm has shown to maintain the system EER stable over time, with respect to a non-adaptive biometric system. However, no specific theoretical study has gone in depth on the differences between self- and co-update. Although some preliminary papers (Didaci et al., 2008, 2009) have tried to model the co-update behaviour, whilst other ones (Roli et al., 2007; Rattani et al., 2008) have compared self-update and co-update experimentally on small data sets, none of them is able to give general insights about the motivation of their functioning. An ad hoc study is thus necessary to confirm previous achievements.

This is the aim of the present paper. First, contributions to the state-of-the-art are given in terms of a conceptual explanation of the behaviour of both algorithms according to the path-based clustering method. Second, an analytical model for the co-update algorithm, inspired from the above concepts, is proposed. Third, the model is used to analytically show that co-update algorithm is much more efficient than the self-update algorithm in terms of performance and number of acquired representative templates.

The conceptual explanation and the proposed model can be adapted to any self-update and co-update methods where classification systems are based on template matching, where each pattern admits two different representations (feature spaces), and the independence among matchers can be assumed. The experimental validation has been done on the case-study of bi-modal biometric verification systems, but without loss of generality.

Experimental evidences supporting our claims are carried out the data set collected at DIEE Laboratories, which, to the best of our knowledge, is the only one non-chimerical multi-modal data set publicly available explicitly conceived for adaptive biometric systems.

Rest of the paper is organized as follows. Section 2 describes self-update and co-update algorithms. Section 3 performs a theoretical analysis of self and co-update performance by exploiting the path-based clustering approach. Experiments are presented in Section 4. Conclusions are drawn in Section 5.

2. Algorithms summary

In order to make better understandable the aims of this paper, we briefly describe self-update and co-update algorithms. They must be performed for each client stored in the system database.

For sake of clarity, we refer to face and fingerprint biometrics. In all cases, biometric verification systems are the focus of this work. When a person submit her/his biometric (e.g. face) and claims a certain identity, the system acquires the biometric signal and extracts a statistical or structured set of features. This set is compared to the template, an analogous set of features stored in the system memory after the registration process. These two feature sets are compared and a match score is computed. Match score is the similarity level among compared feature sets. It can be derived by evaluating the distance from two statistical feature sets (Turk and Pentland, 1991), or defining an appropriate metric from two structural representation (NIST, 2012; Jain et al., 1997; Wiskott et al., 1997). Good surveys on methods for evaluating the match score of fingerprint and face templates can be found in Maltoni et al. (2003) and Li and Jain (2005).

Self-update algorithm.

For each client stored in the system memory:

1. A batch of biometric samples B whose size is $|B| = N_{TOT}$ is collected during a certain time period.
2. A subset of k samples ($k \leq N_{TOT}$) is extracted with re-insertion from B and submitted to the system.
3. The system tries to verify the claimed identity for each of the k samples. The verified samples are added to the client’s gallery.
4. Steps 2–3 are repeated until a certain stop criterion is met.\(^4\)

Co-update algorithm.

The co-update algorithm works under the following hypothesis: two biometric traits are conditionally independent given the identity (Blum and Mitchell, 1998). For each couple of biometric samples $(x^{(1)}, x^{(2)})$ from the same individual:

$$p(x^{(1)} = \hat{x}^{(1)}|x^{(2)} = \hat{x}^{(2)}) = p(x^{(1)} = \hat{x}^{(1)})$$

(1)

$$p(x^{(2)} = \hat{x}^{(2)}|x^{(1)} = \hat{x}^{(1)}) = p(x^{(2)} = \hat{x}^{(2)})$$

(2)

In other words, the probability that $x^{(1)}$ assumes a particular value $\hat{x}^{(1)}$ is independent on the value of $x^{(2)}$ (and vice versa). Eqs. (1) and (2) require that, for biometrics from the same individual, the appearance of the first and the second biometrics are independent each others. This assumption is met if there is no correlation among biometrics at hand.

The co-update algorithm description is given in the following.

Let us call “master” the biometric that assumes the supervisors role, and “slave” the biometric whose gallery is augmented thanks to the master biometric, as suggested in Didaci et al. (2008, 2009, 2011)\(^3\):

1. A batch of biometric samples pairs $B$ whose size is $|B| = N_{TOT}$ is collected during a certain time period.
2. A subset of $k$ samples ($k \leq N_{TOT}$) is extracted with re-insertion from $B$ and submitted to the system.
3. The system tries to verify the claimed identity for each of the $k$ samples, using the master matcher. If the claimed identity is verified, the ‘Slave’ sample is added to the client’s slave gallery.
4. Master and slave matchers roles are inverted.
5. Steps 2–4 are repeated until a certain stop criterion is met.\(^4\)

\(^4\) Usually, the adopted stop criterion refers to a certain number of iterations.

\(^3\) In the following, the terms Master and Slave will be used for indicating: the biometric roles, and also related galleries, matchers and performance. For example: master gallery means the gallery of the master biometrics).

\(^4\) Usually, the adopted stop criterion refers to a certain number of iterations.
The main difference between co-training and co-update is that the latter require the addition of a novel sample only from the side of the ‘slave’ matcher, whilst the role of the ‘master’ one is limited to the pattern ‘selection’. Thus, only the slave pattern is added by co-updating. This avoids the problem of self-updating, usually prone to add patterns very similar to the existing ones in the gallery without a significant contribution in decreasing the error rate (Rattani et al., 2008).

The concept of self–training appeared in Nagy and Shelton (1966) for general pattern recognition problems. Roli and Marcialis (2006) applied it to biometric applications. Terms “self-update” and “co-update”, appeared first in Roli et al. (2007, 2008), and further recalled in Didaci et al. (2011, 2008, 2009) and Rattani et al. (2008). Roli et al. (2007) have proposed the first experimental comparison between self-update and co-update on a small data set: co-update allows an increase of the gallery size faster than self-update. In Didaci et al. (2008, 2009), a model is proposed in order to simulate co-update behaviour. Model is tested under ideal conditions, namely, impostors absence and experiments are performed on a very small initial gallery. The use of a very small gallery allows to highlight the effectiveness of the semi-supervised method.

Finally, some clarifications about the sampling with replacement adopted for self- and co-update are necessary. In the real world, we have no control over samples that are submitted to the system. We have to deal with biometric samples distributed in a continunum. However, two biometrics close enough can be considered as “identical” in terms of expressive power of the client identity. Thus, it seems appropriate to use a “sampling with replacement” strategy to extract, at each iteration, samples, or couples of samples, from the batch.

3. Self-update and co-update modelling by path-based clustering

3.1. Conceptual representation

Self-update process described so far can be associated to the so-called path-based clustering (Fischer and Buhmann, 2003), as shown in Fig. 1.

A path-based cluster is defined introducing a pairwise similarity function $f(x_i, x_j)$ on the samples (graph nodes) $x_i, x_j$ and a threshold value $\tau$. Samples $x_i, x_j$ are connected by an edge if $f(x_i, x_j) > \tau$. Thus, two samples $x_i, x_j$ which are assigned to the same cluster, are either similar ($f(x_i, x_j) > \tau$), or exists a “chain” of intermediate samples $x_{i_1}, \ldots, x_{i_n}$ such that: $f(x_{i_1}, x_{i_2}) > \tau, \ldots, f(x_{i_{n-1}}, x_{i_n}) > \tau$.

The starting template can be viewed as the seed of the clustering process. Only nodes (i.e. input samples) that follow the rules of path-based clustering can be reached from those seeds. From the biometric system point of view this means that there are samples, representing variations, or changes, of the subject appearance, that cannot be exploited by self updating, since they are intrinsically out of the path-based cluster. These samples may represent abrupt changes$^5$ of the input data, including intrinsically isolated samples. On the other hand, there are data which do not form paths in the graph, because the intermediate sample(s) connecting them with existing templates (samples already added into the client’s gallery) is (are) not yet submitted to the system. Therefore, that isolated samples may be neglected by self-update until the required samples, or other samples alternatively connected, appear.

On the basis of the observations above this representation allows to draw the following claims, that will be supported by experiments reported in Section 4.3:

1. Initial template selection is crucial for the self-update efficiency: if only one template is available, it should belong to a path-based cluster such that many samples may be potentially reached. If more templates are available, they should be located in different path-based clusters;
2. Samples temporary neglected may anyway belong to one of existing path-based clusters, but cannot be reached by self-update due to the temporary absence of intermediate samples or alternative paths.

Let us consider now the co-update algorithm, and consider the case of one template per client and per biometric. As an input sample (made up of a pair of biometrics) is submitted, the match score of the master biometric is computed (Fig. 2, at $t_1$). If this score is higher than the updating threshold, the sample corresponding to the slave biometric is added to the client’s gallery (Fig. 2, at $t_2$). We may represent this “connection” as a “virtual” edge cross-linking the template of the master biometric and the sample of the slave biometric. Master and slave biometrics invert their role and the process is repeated (Fig. 2, at $t_3$ and $t_4$). Even in this case, we have a sort of path-based clustering, but these virtual edges connect samples that can be ‘non-similar’. Due to the hypothesis of conditional independence among biometrics, the ‘virtual path’ (Fig. 2, at $t_4$) can connect samples that represent significant intra-class variations, since no constraint due to the updating threshold is followed for the slave biometric. We may refer to this process as a virtual path-based clustering.

This means that samples representing variations, or (abrupt) changes, of the subject appearance, as well as intrinsically isolated samples can be exploited by co-updating thanks to the virtual

$^5$ This term is often used in the pattern recognition field named change detection, often applied to video-surveillance problems (Ziliani and Cavallaro, 2001).
edges mechanism. Moreover, samples of one biometric are virtually connected by all samples of the complementary biometric.

This representation allows to draw the following claims, that will be supported by experiments reported in Section 4.3:

1. for the algorithm efficiency, initial template selection for the co-update is less crucial than for the self-update, even in the case of only one template available per biometric;
2. all samples can potentially belong to one of existing virtual path-based clusters.

To sum up, co-update has an expressive power much superior than that of self-update, as we will show in the next Sections.

### 3.2. An analytical view of path-based clustering for self- and co-update

Aim of this Section is connecting the conceptual representation above with the main parameter of interest for biometric applications, namely, FRR (given that FAR = 0%).

Let us consider the representation given in the previous section. Two synthetic parameters represents the batch B: the number of connections of a node (sample) and the relative fraction of isolated samples over B.

The number of connections $\mathcal{m}(x)$ of a node $x$ in the path-based graph is defined as the number of incident edges to $x$. A node $x$ with $\mathcal{m}(x) = 0$ is an ‘isolated node’. Thus, according to the discussion done in Section 3.1, $m$ is a random variable falling in the range $[0, N_{\text{max}}]$, where $N_{\text{max}}$ is the maximum size achievable by the gallery.

In the case of self-update, $N_{\text{max}}$ is the size of the path-based cluster(s) drawn from the initial template(s), so $N_{\text{max}} < N_{\text{TOT}}$, whilst in the case of co-update, $N_{\text{max}} = N_{\text{TOT}}$.

Let $f_m$ be the fraction of all nodes such that $\mathcal{m}(x) = m$. This is a random variable, whose expected value is:

$$\bar{m} = \sum_{m=0}^{N_{\text{max}}} m \cdot f_m = \sum_{m=1}^{N_{\text{max}}} m \cdot f_m$$

Equation (3) shows the probability that a novel path is drawn from an existing template: first term depends on intrinsically isolated samples, and second one depends on the absence of an intermediate “chain” of samples able to connect template to an input biometric. In the self-update case, $N_{\text{max}} < N_{\text{TOT}}$ and $f_0 = f_0^0$. In the co-update case, $N_{\text{max}} = N_{\text{TOT}}$ and $f_0 < f_0^0$.

Let $n$ be the current size of the gallery (master gallery, in the co-update case). According to our assumptions and hypotheses, a certain sample $x$ is rejected if:

1. $x$ is an isolated sample, in the co-update case, or is a sample not belonging to the path-based cluster(s) drawn from the initial template(s), in the self-update case. This event occurs with probability $f_0$.
2. $x$ does not belong in the case 1 – event that occurs with probability $(1 - f_0)$ – but no samples “similar” to $x$ are present in the gallery. Let us call $r$ the number of samples similar to $x$ that are present in the gallery. The probability that $x$ will be rejected is $p(r = 0)$.

To sum up:

$$\text{FRR} = f_0 \cdot (1 - f_0) \cdot p(r = 0)$$

The value $p(r = 0)$ can be computed by observing that if the gallery is large enough, that is, $n > N_{\text{max}} - \bar{m}$, at least one sample similar to $x$ will be in the gallery:

$$p(r = 0) = 0$$

otherwise, the random variable $r$ can be modelled by a hypergeometric distribution:

$$p(r = 0) = \frac{N_{\text{max}} - 1 - \bar{m}}{\frac{n - 1}{n - 1}}$$

Eq. (4) shows the probability that a novel path is drawn from an existing template: first term depends on intrinsically isolated samples, and second one depends on the absence of an intermediate “chain” of samples able to connect template to an input biometric. In the self-update case, $N_{\text{max}} < N_{\text{TOT}}$ and $f_0 = f_0^0$. In the co-update case, $N_{\text{max}} = N_{\text{TOT}}$ and $f_0 < f_0^0$.

Estimating FRR is possible only for co-update, whilst it is an open problem for self-update, since $N_{\text{max}}$ is unknown. On the other hand, thanks to the relationships $f_0^0 < f_0^0$, it can be proven that $\text{FRR}^0 < \text{FRR}^0$. Eq. (6) also points out that FRR is a function of the size of the gallery, $n$, thus it is necessary to propose a theory able
to predict this number at each iteration of the self- and co-update processes.

Discussion above clearly supports claims 1–2 we did for self- and co-update in the previous Section. Eq. (4) clearly indicates the components which determines the FRR of the system, that is, the probability that a path is not drawn from an existing template: the presence of an isolated sample, or the temporary absence of an intermediate “chain” of samples.

3.3. Path-based and virtual path-based clustering: a stochastic view

Let \( n \) be the current size of the gallery, for a given identity, as the self- or co-update process starts. Let us assume self- and co-update working at zeroFAR operational point, that is, at a threshold for which FAR = 0.

In the updating stage, a set of \( k \) biometric samples from a batch set \( B \) is presented to the system. On \( k \), only \( k_{\text{VER}} \) samples will be over the threshold, that is, the system is affected by an intrinsic False Rejection Rate given by the correspondent zeroFAR value:

\[
k_{\text{VER}} = k \cdot (1 - \text{zeroFAR})
\]

where zeroFAR is the FRR value at FAR = 0. In the case of self-update, zeroFAR is that of the matcher itself, whilst, in the case of co-update, it is the one of the master matcher. Eq. 7 is an average value which is obtained by modelling the number of verified samples as a Binomial random variable with probability equal to \( (1 - \text{zeroFAR}) \) (Didaci et al., 2008). The next step is to compute, among \( k_{\text{VER}} \) accepted samples, how many of them are not present into the gallery. In other words, we have to compute the number \( \Delta n \) of samples which will be added to the gallery. This can be reached by modelling \( \Delta n \) as a random variable conditionally dependent on the number \( k_{\text{OFF}} \) of different samples on \( k_{\text{VER}} \) and not present into the gallery. The random variable \( k_{\text{OFF}} \) is modelled as a multinomial distribution:

\[
p(k_{\text{OFF}}) = \binom{N}{k_{\text{OFF}}} \sum_{i=0}^{k_{\text{OFF}}} (-1)^{k_{\text{OFF}}-i} \binom{k_{\text{OFF}}-i}{i} \left( \frac{k_{\text{OFF}}}{N} \right)^{1+N_{\text{OFF}}}
\]

Whose expected value is

\[
E[k_{\text{OFF}}] = \sum_{k_{\text{OFF}}} k_{\text{OFF}} \cdot p(k_{\text{OFF}}) = N \cdot \left[ 1 - \left( \frac{1}{N} \right)^{k_{\text{OFF}}} \right]
\]

At each iteration, \( \Delta n \) allows predicting the size of the gallery \( n \) \((n = n + \Delta n)\) and, recalling that FRR is a function of \( n \) (see (4), (6)), it allows predicting the system performance. Thus, evaluating the size of the gallery is crucial.

At the current state of our work, estimation of \( N \) is unsolved for self-update. Since the zeroFAR value is correlated with \( N \), no analytical expression of zeroFAR can be given for self-update, but only for co-update. Despite this gap, thanks to the fact that \( N < N_{\text{TOT}} \) for self-update, it is easy to proof that the value \( E[\Delta n] \) for self-update is less than \( E[\Delta n] \) for co-update. In other words self-update is 'slower' than co-update in adding novel templates to the gallery. At the same time, thanks to the concepts of virtual edge and path, the number of connections of a sample, when using co-update, is equal to \( N = N_{\text{TOT}} \). This guarantees that co-update may potentially add to the clients gallery a number of intra-class variations much larger than that of self-update.

This is obviously not possible when using self-update, where the number of connections of a sample is \( N < N_{\text{TOT}} \). In this case, \( N \) is strongly reduced due to the presence of intrinsically isolated samples and the eventual absence of intermediate chains of samples, a problem which can be overcome by co-update.

4. Experimental results

Experiments are carried out with the following aims:

1. Showing the effectiveness of the conceptual representation based on the path-based clustering by pointing out the characteristics of clusters for some selected clients.

2. Evaluating the simulation ability of the presented model when working hypothesis (zeroFAR operational point and exact knowledge of \( f_o \) and \( m \) parameters) are fully or partially respected.

4.1. Data set

Data set adopted for experiments is the DIIIEE Multi-modal Data Set. This data set has been collected at Computer Science Laboratory of Department of Electrical and Electronic Engineering at University of Cagliari. Selected biometrics are the fingerprint of right thumb and index, and the subject face.

The DIIIEE Multimodal Data set is publicly available by contacting the authors.

4.2. Experimental protocol

First of all, well-known face and fingerprint matchers have been selected for experiments:

- **Face matching.** PCA (Turk and Pentland, 1991) has been applied to all images of available data set. All images are then converted according to the matrix transformation derived. Matching score is obtained by converting the Euclidean distance between template and input samples according to the min–max normalization rule.

- **Fingerprint matching.** Minutiae have been extracted by NIST MinDTCT algorithm and compared by NIST Bozorth3 algorithm (NIST, 2012) for obtaining match scores.

In both cases, one template per client has been extracted. Table 1 shows the performance of individual matchers in terms of EER and zeroFAR operational points, computed on the whole data set. Some performance differences can be noticed, especially between fingerprint and face matchers. Table 1 allows us to correlate the proposed model performance with individual matchers characteristics.
In the following, experiments described are:

- **Experiment (1)**. Path-based and virtual path-based clusters are shown for two representative clients, in order to point out the advantages of our modelling.

- **Experiment (2.1)**. Evaluation of the proposed stochastic model. Model parameters and zeroFAR threshold have been estimated on the whole data set, in order to avoid estimation errors and fully respect the model’s working hypothesis. Thresholds are user-specific, that is, an appropriate zeroFAR threshold is estimated for each client.

- **Experiment (2.2)**. Evaluation of the proposed stochastic model when the zeroFAR operational point is bad estimated. This is a more realistic condition, where impostors could be potentially inserted into the clients gallery. In this and previous experiment, results refer only to the co-update algorithm, since no exact evaluation of model parameters can be done for self-update.

In all cases, $N_{TOT}$ is the size of $B$.

Multimodal verification systems investigated are based on:

- **M1**: right thumb and index fingerprint matchers;
- **M2**: face and right index matchers;
- **M3**: face and right thumb matchers.

### 4.3. Experiment 1. Path-based and virtual path-based clusters

_Figs. 3 and 4_ show a subset of the path-based clusters when performing self updating on all images available. _Fig. 3_ points out that client 1 is characterized by a path-based cluster whose size is much wider than that of other clusters, that is, a dominant cluster.

Obtained clusterization is “true” only for this particular data set, without the aim to generalize to any pattern recognition problem, but it is in agreement with the user population proposed by Doddington et al. which pointed out the existence of clients easy to be recognized (Doddington et al., 1998). In this case, selection of the template allows the best performance if a sample of the dominant cluster is selected. Due to the presence of the dominant cluster, it is also very likely to capture a pattern from this cluster during enrolment, so the template selection phase is not crucial. On the other hand, _Fig. 4_ shows that other clients can be characterized by many path-based clusters. Even in this case we are in agreement with Doddington et al. (1998), where clients difficult to be recognized are pointed out. This means that the choice of the initial template can be crucial depending on the particular working environment of the system. Alternatively, more than one templates are necessary for this subject.

Moreover, we may notice, in both cases, the presence of several isolated samples. It is possible that some of them have not been inserted because of: (1) the lack of the intermediate samples (not present in the batch $B$), (2) the too stringent threshold.

Item (2) allows us to point out the independence of the co-update algorithm from the updating threshold. In fact, _Fig. 5_ shows the virtual path-based graphs of the same clients, where it is evident that samples “isolated” in the self-update view, are connected in the co-update view. Considered samples have been added into the client gallery thanks to the contribution of the Master matcher, and can be exploited as well during the recognition process, when the system is operating.

This is a clear advantage of co-updating over self-updating, and points out that co-update is able to insert more samples to the clients gallery than self-update.

### 4.4. Experiment 2.1. Validation of the stochastic model based on the path-based clustering view

_Fig. 6_ shows both the gallery size and the related False Rejection Rate when considering the bi-modal system M2. Similar results
have been obtained on M1 and M3, and have been omitted for sake of space. On x-axis, we put the amount of unlabelled data submitted to the system. Dotted curves refer to values estimated by using the proposed framework, whilst solid ones to the values obtained by experiments according to the protocol described in Section 4.2. Table 2 shows the maximum difference between the estimated values of FRR and size of the gallery, and the corresponding empirical values.

Since all model parameters have been set on the basis of all available data, these results refer to ideal working hypothesis. However, they show that the model is very precise and coherent with our assumptions. Reported results also show the adherence between the proposed model and the real gallery size and expected FRR with respect to the zeroFAR operational point. For example, it can be noticed from Fig. 6 that, in order to reach a FRR near to 10%, it is enough to collect less than 80 ‘genuine users’ samples during
system operations (see plot related to right index). Correspondingly, designer can predict the size of the gallery corresponding the required FRR, under the condition that used $m$ and $f_0$ are representative of the collected batch. In our example, each gallery size will be, on average, less than 30 samples.

4.5. Experiment 2.2. Validation of the model in case of bad zeroFAR estimation

In this experiment, the model parameters are estimated on the whole $B$, which is fully available to the designer, but zeroFAR thresholds are computed on a subset of data, for simulating the impossibility in estimating this operational point exactly. The aim of this experiment is to see if model predictions are subjected to drift, and, secondly, the impact of impostors presence on such predictions. In other words, to investigate if the model holds his usefulness in this real case for the designer, and at which extent. Results are shown in Fig. 6(a–d) and refer to bi-modal system M2. Similar results have been obtained for M1 and M3, thus have not been reported for sake of space. We plot the average number of samples in the client gallery as computed by proposed model, and the 'true' value of genuine samples. Since thresholds are not well estimated, as usual in a real verification scenario, some impostors trials are misclassified and wrongly inserted into client gallery (dashed lines). It can be also seen that the availability of more samples improve the zeroFAR threshold estimation, thus predicted curves are more adherent to the empirical ones. From Fig. 7(a)–(d), the percentage amount of samples used for thresholds evaluation is increasing. Four percentages are reported here: 20%, 25%, 50%, 80%. As a consequence, it is possible to see that the number of impostors decreases.

Results shown in Fig. 7 point out that, with regard to designer expectations related to the gallery size, the model is still reliable but it cannot be used over all iterations. For example, we may consider Fig. 7(a) and (b) where the impostors effect is more evident. We can notice that, for the first 50 iterations model predictions are still reliable, since the amount of inserted impostors is too small. But, when iterations are more than 50, these predictions become gradually unreliable. This means that, if used in real scenarios, co-update needs, anyway, human intervention.

Table 2

<table>
<thead>
<tr>
<th>System</th>
<th>FRR</th>
<th>SIZE GALLERY</th>
<th>Biometric 1</th>
<th>Biometric 2</th>
<th>Biometric 1</th>
<th>Biometric 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.0198</td>
<td>0.50</td>
<td>0.0327</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0170</td>
<td>2.26</td>
<td>0.0567</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0357</td>
<td>2.38</td>
<td>0.0615</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results have been obtained for M1 and M3, thus have not been reported for sake of space. We plot the average number of samples in the client gallery as computed by proposed model, and the 'true' value of genuine samples. Since thresholds are not well estimated, as usual in a real verification scenario, some impostors trials are misclassified and wrongly inserted into client gallery (dashed lines). It can be also seen that the availability of more samples improve the zeroFAR threshold estimation, thus predicted curves are more adherent to the empirical ones. From Fig. 7(a)–(d), the percentage amount of samples used for thresholds evaluation is increasing. Four percentages are reported here: 20%, 25%, 50%, 80%. As a consequence, it is possible to see that the number of impostors decreases.

Results shown in Fig. 7 point out that, with regard to designer expectations related to the gallery size, the model is still reliable but it cannot be used over all iterations. For example, we may consider Fig. 7(a) and (b) where the impostors effect is more evident. We can notice that, for the first 50 iterations model predictions are still reliable, since the amount of inserted impostors is too small. But, when iterations are more than 50, these predictions become gradually unreliable. This means that, if used in real scenarios, co-update needs, anyway, human intervention.

Fig. 8 shows the Area under ROC (AuR) curves for each iteration. They are referred to multi-modal system M2. For lack of space, and as the data referred to other biometrics do not add anything to the reported results, we only show the values for fingerprint index. Four different sizes of training data for estimating zeroFAR threshold have been taken into account, as for results reported in Fig. 7. For the first 50 iterations, system performance increases (AuR decreases), but after about 100 iterations the AuR increases due to impostors wrongly inserted into client gallery. This effect is more evident using a small amount of data for estimating zeroFAR threshold (i.e. 20% or 25% of the data).
Figs. 7 and 8 show that, due to presence of impostors, the final ROC curve is worse than initial one. This points out the need of a periodical human supervision, aimed to remove eventual impostors wrongly inserted due to estimation errors on thresholds.

5. Discussions and conclusions

The conceptual and theoretical model, based on the path-based clustering, proposed in this paper, is aimed to explain the functioning of self- and co-update and the promising results obtained in previous works. The user population is characterized according to the paths defined by templates and the batch of samples for updating. From this characterization, we formulated general inferences about self and co-update potentialities, their advantages and drawbacks.

Our model showed that self-update is not effective as co-update in capturing intra-class variations. The amount of genuine samples introduced into the clients gallery are intrinsically less than that added by co-updating, thanks to the fact that co-update overcame self-update limitations by drawing virtual paths across biometrics.

Our theoretical findings have been confirmed by experiments on a large multi-modal, non-chimerical data set, publicly available by contacting the authors.

We also investigated the robustness of the model where zero-FAR operational point is bad estimated. Reported results clearly show that, although model predictions are still reliable, however, the human intervention is necessary to remove all impostors from clients gallery.

According to our experiments the amount of wrongly inserted samples is less than 10\%, over the whole samples added, so filtering them could be relatively easy: this could be done, for example, by plotting the path-based clusters in order to see that, eventually, those samples could be isolated and, thus, removed. This is
obviously much simpler than selecting significant intra-class variations by supervision.

Moreover, the model predictions are reliable even if few samples are used for thresholds estimation, and a few of impostor samples are wrongly inserted.

Future works will include the refinement of self-update modeling, in order to give a quantitative estimation of its performance, and also the modelling of FAR, as done for FRR. This model improvement could bring a significant contribution in order to assess the applicability of co-update algorithm in real scenarios.

Acknowledgements

This work has been partially supported by Regione Autonoma della Sardegna through the Regional Law no. 7 for Fundamental and Applied Research, in the context of the funded project ‘Adaptive biometric systems: models, methods and algorithms’, grant no. CP4_442. Luca Didaci and Gian Luca Marcialis also thank University of Cagliari for the special grant received in the context of their proposed “Joung Researchers” Project on the topic ‘Adaptive biometric systems for personal recognition’.

References


