A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study

Ajita Rattani*

Department of Computer Science and Engineering,
Michigan State University,
East Lansing, MI 48824, USA
E-mail: ajita@msu.edu
*Corresponding author

Gian Luca Marcialis and
Fabio Roli

Department of Electrical and Electronic Engineering,
University of Cagliari,
Piazza d'Armi, Cagliari 09123, Italy
E-mail: marcialis@diee.unica.it
E-mail: roli@diee.unica.it

Abstract: Adaptive biometric systems have received a recent spurt in biometric community. These systems have the additional capability to adapt themselves using biometric data available during the system's operation. Although several studies have been proposed in this field, no conclusive evidences can be drawn about the expected performance gain on making the biometric system adaptive. This is due to the adoption of different and inappropriate databases, protocols and tools for evaluating adaptive biometric systems. This paper presents a benchmarking study to facilitate fair comparison and independent replication of the results from different research groups. To this aim, this paper describes DIEE multi-modal database consisting of face and fingerprint biometrics and a protocol tailored for adapting as well as evaluating adaptive biometric systems. In addition, several tools for evaluating and visualising the performance gain on making the biometric system adaptive are provided as well. To the best of our knowledge, this is the first attempt to benchmark database, protocol and tools for evaluating adaptive biometric systems operating in verification mode.

Keywords: adaptive biometrics; biometric template update; self-update; co-update; benchmarking; dataset; protocol; tools.

Reference to this paper should be made as follows: Rattani, A., Marcialis, G.L. and Roli, F. (2013) 'A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study', *Int. J. Biometrics*, Vol. 5, Nos. 3/4, pp.266–287.

Biographical notes: Ajita Rattani is a Post-Doctoral Fellow at the Department of Computer Science and Engineering, Michigan State University, USA. She received her PhD from the Department of Electrical and Electronic Engineering, University of Cagliari, Italy in 2010. Her research interests include pattern recognition, classifier fusion, machine learning, computer vision and biometrics. She has bagged more than 30 research papers in

international conferences and journals. Further, she has served as a reviewer/technical programme committee member for high impact conferences and journals in her field of interest.

Gian Luca Marcialis received his MS and PhD in Electronic and Computer Science Engineering from the University of Cagliari, Italy, in 2000 and 2004, respectively. He is currently an Assistant Professor and a member of the Pattern Recognition and Applications group at the University of Cagliari. His research interests are in the fields of biometrics. In particular, these interests are in fingerprint vitality detection, biometric template update by semi-supervised approaches, and fusion of multiple biometric matchers for person recognition. He co-authored more than 60 papers. He is co-organiser of International Fingerprint Liveness Detection Competition (first, second, third edition held in 2009, 2011, 2013, respectively).

Fabio Roli received his MS degree, with honours, and PhD in Electronic Engineering from the University of Genoa, Italy. He is a Professor of Computer Engineering and Director of the Pattern Recognition and Applications Laboratory. His research activity is focused on the design of pattern recognition systems and their applications to biometric recognition, multimedia text categorisation, and computer security. He is a member of the governing boards of the International Association for Pattern Recognition and of the IEEE Systems, Man and Cybernetics Society. He is a Fellow of the IEEE, and of the International Association for Pattern Recognition.

1 Introduction

Increasing security concerns ranging from individual identity theft to corporate and national security are driving the biometric market (Jain et al., 2007). Whilst this technology continues to expand, an intrinsic characteristic of the technology is that system error rate simply cannot attain absolute zero. The major cause of these errors is the presence of substantial variations in the available operational data. The cause of these variations can be due to vulnerable acquisition conditions (i.e., uncontrolled environment) like pose changes, illumination variations, human sensor interactions, occlusions, etc. (Jain et al., 2007). As a consequence, significant miss-match errors result due to temporary variations (such as expression, pose variations) in the input data. Apart from these being biological tissues in nature, biometric traits change over time, causing temporal and permanent changes in the operational biometric data.

In contrary to this, enrolment samples (templates) are often acquired in a controlled environment and are usually very few, mostly single image (Jain et al., 2007). As an important consequence, the enrolled samples become un-representative of the intra-class variation of the input operational data. This is due to the compound effect of limited enrolment samples and the intra-class variations (both temporary and temporal) in the input biometric data. Evidences of performance degradation as a result of limited enrolment samples together with the presence of variations in the operational data abound (Tan et al., 2006; Flynn et al., 2003).

Standard solutions to deal with the issue of template representativeness are:

a re-enrolment sessions at fixed time interval

b multi-biometrics (Ross et al., 2006)

c feature invariance/signal restoration schemes (for instance, illumination normalisation scheme for face biometric) (Zhou et al., 2007)

d virtual biometric synthesis (for instance, pose correction, age transformation techniques) (Patterson et al., 2006).

Solution (a) may cause user inconvenience, (b) may fail to adapt to temporal changes in the biometric traits and (c) or (d) may be prone to estimation or fitting errors.

A promising way is to auto-update the templates to the variation of query samples available during the operational phase (Rattani et al., 2009, 2012, 2013; Poh et al., 2012). These systems are termed as adaptive biometric systems. The adopted learning mechanism is based on semi-supervised learning (Zhu, 2008). Semi-supervised learning is a machine learning technique based on the joint use of labelled and unlabeled data in boosting the classifier's performance. However, one drawback of this learning technique can be that misclassification errors cause impostor intrusion into the updated template set. Accumulation of impostor samples in the user's template gallery could be counter-productive to the performance of the biometric system. As a consequence, most of the adaptive biometric systems use only the highly confidently classified genuine samples for template adaptation.

The type of data used for adaptation can also make a difference. In particular, in a video-based biometrics, one can exploit the fact that the person whose biometric trait being sampled remains same for the entire video sequence (Connolly et al., 2010). This is since each consecutive pair of images in the sequence are a fraction of second apart from each other. This property cannot be exploited for still image-based biometric systems where the identity of the user has to be inferred for each input image. Consequently, this is a harder problem. We are interested in the problem of adaptation for a biometric system based on still input (query) images. Template adaptation using video sequence is out of the scope of this manuscript.

Recently, adaptive biometrics have received significant attention from the research community (Rattani et al., 2008a, 2008b, 2009, 2013; Uludag et al., 2004; Jiang and Ser, 2002; Ryu et al., 2006; Liu et al., 2003; Roli and Marcialis, 2006; Roli et al., 2007; Pavani et al., 2009; Amayeh et al., 2009; Poh et al., 2010, 2012; Garca and Perales, 2008; Gayar et al., 2006; Martinez and Fuentes, 2003; Giot et al., 2011; Freni et al., 2008; Marcialis et al., 2008). This research direction is expected to gain further momentum because of its key promulgated advantages: one no longer needs to collect large number of input samples during the enrolment session, re-enrolment sessions are no longer necessary instead, input data available during the system operation is used for adaptation. These conveniences can significantly reduce the cost of maintaining a biometric system.

Despite these existing studies, it is not straight forward to compare two adaptive biometric systems. This is because current papers adopt different databases with different types of variations for instance, either illumination or expression changes for face biometrics (Rattani et al., 2009). Moreover, most of the adopted datasets are assembled over a short time span and contain limited number of samples per user. As an important consequence, performance of an adaptive biometric system cannot be evaluated over a long time span.

Although a number of independent European projects have contributed several multi-biometric databases, e.g., the BANCA (Bailly-Bailliere et al., 2003), XM2VTS (Luttin, 1998) etc., none of these databases fulfil the requirements to study the problem of adaptive biometric systems.

Specifically, a database appropriate for studying adaptive biometric systems should exhibit the following characteristics:

- large number of temporary (occurring due to uncontrolled environment), and temporal variations (occurring due to the lapse of time)
- longitudinal database consisting of large number of samples per user, collected in different sessions over a period of time
- maintenance of the chronological order of time for the collected samples.

Furthermore, most of the papers adopt different protocols for the evaluation of adaptive biometric systems (Rattani et al., 2009). In other words, existing studies differ in experimental attributes such as number of initial enrolled templates used, threshold parameter set for input sample selection and adaptation, assumption of impostor attack, adopted mode of adaptation, i.e., online or offline etc. These mentioned parameters (attributes) have direct influence on the obtained performance gain of the adaptive over the baseline biometric system. As a consequence, existing adaptive biometric systems cannot be compared and the expected performance gain on making the biometric system adaptive cannot be gauged (Rattani et al., 2009).

The aim of this manuscript is to present a benchmarking study for adaptive biometric systems. Accordingly, the contributions of this manuscript are as follows:

- Benchmarking DIEE multi-biometric database explicitly collected for studying the problem of adaptive biometric systems. This database can be obtained by contacting the authors.
- Benchmarking protocol that simulates the process of adaptation in real biometric systems.
- Standardising tools for evaluating and visualising the performance gain of adaptive biometric systems.

For this benchmarking study, we consider biometric systems operating in verification mode (1:1) and the system only adapts to operational data samples available as still images. The paper is organised as follows: Section 2 explains adaptive biometric systems, its different types of learning algorithms and experimental protocol attributes that make existing adaptive systems incomparable. Section 3, details the multi-biometric test-bed and the proposed protocol. Section 4, describes various tools for quantising and visualising the performance gain on making the biometric system adaptive. Example experimental results are presented in Section 5.

2 Adaptive biometric systems

While a non-adaptive biometric system has two processes, namely, enrolment and matching, an adaptive biometric system has an additional process called *adaptation* or *updating* process. The aim of the adaptation process is to continuously adapt the biometric system to the intra-class variation (temporary as well as temporal) of the input operational data (Rattani et al., 2009). This can be done by updating the matching parameters, templates stored in the system database, or both. This paper focus on template adaptation by augmenting the user's template gallery with the genuinely classified input samples for the biometric systems operating in verification mode (1:1).

Note that temporary variations in the biometric data, for instance, changes in pose, expression, illumination, occlusions due to user accessories, are not the genuine changes. However, they cause disproportionate contribution to false reject rate (FRR) due to the limited representation capability of the enrolled templates. In fact, the literature has shown that complete invariance to temporary changes (for instance pose variations) cannot be attained and the process even introduce artefacts to the data (transformed and pose corrected image) due to the estimation error (Blanz and Vetter, 2003). However, until a perfect pose, expression correction algorithms are found, our conventional wisdom suggests to handle these variations through learning and adaptation.

Self- and co-update are the commonly adopted algorithms from semi-supervised learning for adapting uni- and multi-biometric systems (Rattani et al., 2009), respectively. Next, we briefly describe the basic self- and co-update algorithms for biometric system adaptation.

2.1 Self-updating

In self-updating procedure (Rattani et al., 2008a, 2009, 2013), a biometric system is trained with an initial set of enrolled templates named *T*. A batch of samples, *U*, is available during the system's operation (Roli and Marcialis, 2006) over a period of time. Among all the samples in *U*, only those samples whose matching score on comparison with the enrolled templates exceed a given *updating threshold* (thr^*) (i.e., highly confidently classified samples) are used to update the enrolled templates. This is done to reduce the probability of impostor introduction (false acceptance) into the updated template set (Ryu et al., 2006; Liu et al., 2003; Roli and Marcialis, 2006). This procedure is presented in Algorithm 1.

Algorithm 1 Self-update algorithm

1 Given:
- $T = \{t_1,\ldots,t_M\}$ is the template set.
- $U = \{u_1,\ldots,u_N\}$ is the adaptation set.
- $U^* = \phi$ is an empty set.

2 Estimate thr^* on T

3 For $h = 1,\ldots,N$
- $s_h = matchscore(u_h, T)$
- If ($s_h > thr^*$), then $U^* = U^* \cup \{u_h\}$

4 End For

5 $T = T \cup U^*$

However, recent works argue that usage of only highly confidently classified samples for adaptation may result in capture of samples that are quite similar to the enrolled templates. At the same time, discarding samples representing substantial variations (Rattani et al., 2009). On the other hand, on relaxing the updating threshold, the system may become prone to classification errors and update the enrolled templates using impostor samples (Rattani et al., 2008a, 2009).

2.2 Co-updating

Co-updating process uses mutual and complementary help of two biometric systems to adapt themselves. Each system, trained on initial enrolled templates, classifies the input data. For each highly confidently classified sample, the respective system teaches the other one by adding the corresponding sample to its template set (Rattani et al., 2008b, 2009).

A unique advantage of co-update over self-update is that templates can be updated using samples with very drastic changes. As the two biometric systems are conditionally independent, one system's highly confident data are i.i.d samples for the other system (Roli et al., 2007; Rattani et al., 2008b).

For instance, in an audio-visual system, face-based system can confidently update its template even if the query face sample shows significant pose changes, provided the complementary speech modality has a very high confidence in determining the authenticity. However, co-updating may fail to recognise a real change if all the modalities show variation, for e.g., the seasonal cough can cause a weary face simultaneously with a temporary changed vocal tract. However, this is a very rare situation.

Co-updating procedure is presented in Algorithm 2, where T^{b0} and T^{b1} are the template sets and U^{b0} and U^{b1} are the input batches of samples for $b0$ and $b1$ modalities (for example, fingerprint and face). Similar to self-update, a sample is highly confidently classified if its matching score on comparison with the enrolled templates is above the set threshold $thr^{b0}*$ and thr^{b1*}.

Algorithm 2 Co-update algorithm

1 Given modalities $b0$ and $b1$:

- $T^{b0} = \{t_1^{b0},\dots,t_M^{b0}\}$ and $T^{b1} = \{t_1^{b1},\dots,t_M^{b1}\}$ as the template sets.
- $U^{b0} = \{u_1^{b0},\dots,u_N^{b0}\}$ and $U^{b1} = \{u_1^{b1},\dots,u_N^{b1}\}$ as available batches of samples where sample u_h^{b0} is coupled with u_h^{b1}.
- $U^{b0}* = \phi$ and $U^{b1}* = \phi$ as empty sets.

2 Estimate $thr^{b0}*$ on T^{b0} and $thr^{b1}*$ on T^{b1}.

3 For $i = 0, 1$

(a) For $h = 1,\dots,N$

i $s_h^{bi} = matchscore(u_h^{bi}, T^{bi})$

ii If $(s_h^{bi} > thr^{bi}*)$, then $U^{bi}* = U^{bi}* \cup \{u_h^{bi}\}$

4 $T^{b0} = T^{b0} \cup U^{b0}*$

5 $T^{b1} = T^{b1} \cup U^{b1}*$

2.3 Attributes related to experimental protocol: existing literature

For the evaluation of the adaptive biometric systems, usually three subset of the image database is used: training, adaptation and evaluation (Rattani et al., 2009; Poh et al., 2012). Training set acts as an initial enrolled templates and adaptation set is used for updating the templates. Performance of the adaptive biometric system is evaluated on the test set. Comparative assessment of the adaptive and baseline system (without adaptation) is done by evaluating their performances on the same test set.

Next, the attributes related to experimental protocol that make different adaptive biometric systems incomparable are mentioned as follows:

- *Number of initial enrolled templates:* a biometric system trained with different number of enrolled templates exhibit different baseline performance. As a result, performance of the biometric system trained on different number of enrolled templates will not be comparable on adaptation.

- *Assumption of impostor attack or not:* impostor attacks are assumed by introducing impostor samples in the adaptation set (U and $U^{a,b}$ in Algorithms 1 and 2). This assumption enables the study of counter-productive effect on adaptation using impostor samples.

- *Online or offline adaptation:* Online adaptive systems adapt themselves as soon as the input sample is available. On the other hand, offline methods adapt themselves after the batch of input samples have been collected over a fixed time period. In contrary to offline methods, online methods adhere to chronological ordering of samples as a function of time.

- *Sequence of presentation of genuine and impostor samples:* The sequence of presentation of samples in U and $U^{a,b}$ (in the Algorithms 1 and 2) plays an important role for online adaptive biometric systems. A system updated using genuine samples followed by the impostors will report more and biased performance improvement over the system updated using the random sequence of genuine and impostor samples. This is due to the fact that the probability of impostor intrusion decreases on adaptation using genuine sequence first (Rattani et al., 2009).

- *Update threshold criterion:* This criterion determine how aggressive or conservative a learning strategy should be. As commonly adopted algorithms (like self- and co-update) update the templates only if the matching score of input sample exceed the updating threshold (i.e., thr^* and $thr^{b0,b1}$ for Algorithms 1 and 2). Performance of these adaptive systems differ on varying the updating threshold (Marcialis et al., 2008). Further, these thresholds can be computed on the global or user-specific basis. On the global basis, threshold is computed for the system as a whole, i.e., the same threshold is applied for each user. For user-specific basis, threshold is computed independently for each user. In Marcialis et al. (2008), authors further reported that adaptive biometric system based on user-specific threshold results in higher performance than global threshold. For the sake of completeness, Rattani et al. (2008a, 2012) propose graph-based adaptation schemes, based on min-cut and harmonic function, that bypass the need of updating threshold. However, majority of the existing systems are based on self- and co-update or their variant.

Next, we describe the parameters of the experimental protocol that are directly related to the size of the adopted biometric database as follows:

- *Number of samples used for adaptation:* Size of the database is described as number of subjects (m) and samples per subject (n) (in $m \times n$ format). Size of the database directly determines the number of samples used for adaptation.
- *Type of intra-class variations:* Eventual capture of the temporal variance is possible only if the large number of samples are collected at different time slots over a long time span (Rattani et al., 2008a, 2008b, 2009; Pavani et al., 2009; Liu et al., 2003; Martinez and Fuentes, 2003). Databases collected over short time span can capture only temporary variations (as a result of changes in pose, illumination conditions, etc.) (Amayeh et al., 2009; Roli and Marcialis, 2006).
- *Number of update cycles:* Each adaptation process on the availability of the input batch of samples is referred to as an update cycle. Existing adaptive biometric systems are evaluated over few update cycles (usually one or two) (Amayeh et al., 2009; Roli and Marcialis, 2006; Rattani et al., 2008a, 2008b). This is mainly due to the adopted small-sized database. For biometric recognition systems involving longer terms, adaptation and performance assessment involving multiple update cycles is more realistic.
- *Performance gain assessment:* The methods present in the literature have adopted the following technique for assessing the performance of adaptive biometric systems. These techniques are as follows:
  1. Joint test and adapt: with templates updated with U_t batch of samples available at time t. On the availability of U_{t+1} sample (or batch), the same data is first used for the evaluation of the progressive performance of the templates updated till batch U_t and then updating the templates (Ryu et al., 2006; Liu et al., 2003; Roli and Marcialis, 2006).
  2. Separate test and adapt: with templates updated with U_t samples. Next available sample U_{t+1} (or batch) is used only to assess the performance of the system updated till U_t. Updating is performed using the following next batch of samples, i.e., U_{t+2} (Roli et al., 2007; Rattani et al., 2008a; Martinez and Fuentes, 2003).

  The followed performance gain assessment approach is dependent on the size of the adopted database. Joint test and adapt strategy is convenient for the systems with limited memory and database size.

In Table 1, we give a representative list of the existing adaptive biometric systems based on the above mentioned attributes related to the experimental protocol. Rattani et al. (2008a), Giot et al. (2011) and Marcialis et al. (2008) followed the random sequence of genuine and impostor samples for adaptation. Ryu et al. (2006) analysed all the three techniques for input sample presentation, i.e., genuine sequence first, impostor sequence first and random sequence. Ryu et al. (2006) mentioned that performance of the adaptive biometric system differs for different sample presentation techniques.

Jiang and Ser (2002), Ryu et al. (2006), Liu et al. (2003) and Roli et al. (2007) follow joint test and adapt strategy and references (Uludag et al., 2004; Rattani et al., 2008a; Pavani et al., 2009; Gayar et al., 2006; Roli and Marcialis, 2006) follow separate test and adapt strategy. In fact, most of the existing systems did not explicitly mention the evaluation methodology followed, i.e., whether separate test and adapt (STA) or joint test and adapt (JTA) performance assessment techniques. These studies have been bifurcated into STA and JTA by the authors.

Table 1 A representative list of existing adaptive biometric systems based on different protocol attributes (related to experimental attributes as well as database properties)

Reference	*Dataset*	*Size ($m \times n$)*	*No. of initial templates*	*Impostor attack*	*Performance assessment*
Uludag et al. (2004)	Home-made	50 × 100	5	No	STA
Jiang and Ser (2002)	FVC2000 DB1-4	100 × 8	1	No	JTA
	Home made	12 × 200	1		
Ryu et al. (2006)	Home-made	41× 10	1	Yes	JTA
Liu et al. (2003)	Home-made	20 × 220	10	No	JTA
	PIE	30 × 75	5	No	
	Facial expression	7 × 24	3	No	
Roli and Marcialis (2006)	AR	100 × 14	1	No	STA
Roli et al. (2007)	AR/FVC-2002 DB2	100 × 8	1	No	JTA
Rattani et al. (2008a)	Equinox	57 × 129	1	Yes	STA
Pavani et al. (2009)	GEFA	35 × 340	11 ± 5	No	STA
Amayeh et al. (2009)	Notre Dame	132 × 10	2	No	JTA
Poh et al. (2010)	BANCA	52 × 12	1	No	JTA
Garca and Perales (2008)	Home made	60 × 250	3	No	JTA
Gayar et al. (2006)	UMIST	20 × 25	3	No	STA
Martinez and Fuentes (2003)	UMIST	15 × 20	1	No	JTA
Giot et al. (2011)	Keystroke-DB-CMU	100 × 40	1	Yes	JTA
Freni et al. (2008)	FVC-DB	100 × 8	1	Yes	STA

Note: Acronym STA and JTA stands for separate test and adapt and joint test and adapt performance assessment techniques.

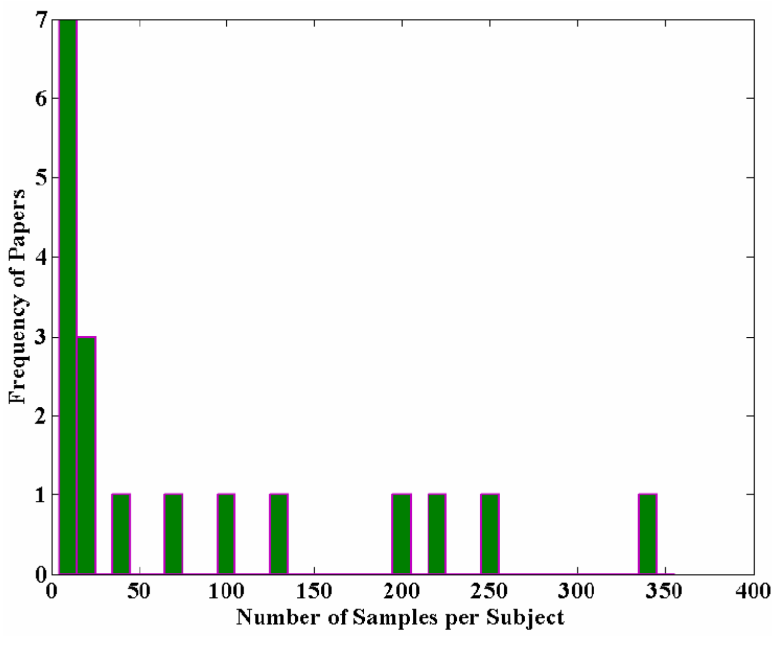
Most of the studies did not assume impostor attacks for the evaluation of the adaptive systems, for instance, Poh et al. (2010), Gayar et al. (2006), Uludag et al. (2004), Roli and Marcialis (2006) and Jiang and Ser (2002). As a consequence, the reported performance gain has been over-estimated due to non-inclusion of the impostor samples into the adaptation set used for template update. Please note that some of the existing papers even did not explicitly and clearly indicated the training, evaluation and the test set (Rattani et al., 2009).

Further, enlisted adaptive biometric systems (in Table 1) have set the update threshold criterion in the range [0, 1]% FAR, i.e., from stringent to moderate. As a result, these systems differ in their capability in capturing samples and their vulnerability to impostor intrusion.

Furthermore, it can be noticed that databases like *UMIST*, *FVC-Db*, *BANCA*, *AR*, as adopted in the existing adaptive biometric systems, contain limited number of samples per person. Figure 1 shows the histogram of number of samples per subject from the databases adopted in existing literature. This figure shows that most of the existing studies adopted databases with samples per person in the range [10, 20]. As a result, these databases have captured only temporary variations (i.e., no temporal variations) making them inappropriate for studying the problem of adaptation. In fact, certain adopted databases also contain very limited type of temporary variations. For instance, UMIST face database contain only pose variations (i.e., no illumination or expression variations, etc). Some of the authors have self-collected databases (indicated as home-made database in the table (Ryu et al., 2006; Garca and Perales, 2008) over a medium time span to study the problem of adaptive biometric systems. However, these databases are not released in the public domain. Other recently available databases like MOBIO and NIST either contain video sequence or insufficient number of samples per person, making them inappropriate for evaluating static image-based adaptive biometric systems.

To sum up, it can be seen from Table 1 that attributes such as database size, number of initial enrolled samples, assumption of impostor attack and the adopted performance assessment approach are not consistent across different studies (Jiang and Ser, 2002; Ryu et al., 2006; Liu et al., 2003; Roli and Marcialis, 2006; Roli et al., 2007). In fact, most of the existing studies have adopted small-sized databases containing only limited type of temporary variations. Thus, there exist a need for benchmarking database and protocol for the comparative assessment of the existing adaptive biometric systems in the literature.

Figure 1 Histogram of the number of samples per subject from the databases adopted in the existing literature (see online version for colours)



Notes: It can be seen that most of the studies adopted databases with samples per person in the range [10; 20]. Thus the adopted databases have not captured temporal variations over medium/ long time span and are inappropriate for the study of adaptive biometric systems.

3 DIEE multi-biometric dataset and protocol

This section describes DIEE multi-biometric dataset and the experimental protocol for adaptive biometric systems.

3.1 DIEE multi-biometric database: face and fingerprints

The DIEE multi-biometric database is collected with an aim to standardise the platform and to integrate multi-disciplinary research efforts for adaptive biometric systems that uses still images for template adaptation. This multi-biometric dataset consists of face, index and thumbprint modalities from 49 subjects (users). Most of the subjects are the university undergraduate and graduate students. Users are not trained before the data collection and usually were non-cooperative. The database is collected over a span of 1.5 years in ten sessions. Face images are captured using CANON camera and fingerprint images are captured using biometrika fingerprint sensor. The details regarding database characteristics are given in Table 2.

Table 2 Characteristics of the DIEE multi-biometric database

Size of the database	49 × 100
No. of subjects	49
No. of samples per subject	100
Time span (years)	1.5
No. of sessions	10
Traits	Face and fingerprint (index and thumb)
Sensors	CANON camera (face)
	Biometrika sensor (fingerprint)

Table 3 The sequence of covariates captured for the ten images acquired in each session for face biometrics

Image no.	*Co-variates*
1	Neutral
2	Semi-left pose (45 degrees)
3	Semi-right pose (45 degrees)
4	Illuminated-Neutral face
5	Illuminated-semi-left pose (45 degrees)
6	Illuminated-semi-right (45 degrees)
7	Angry facial expression
8	Happy facial expression
9	Surprised facial expression
10	Eyes semi-closed

Note: The chronological ordering in captured samples are maintained for each session as indicated with the image number.

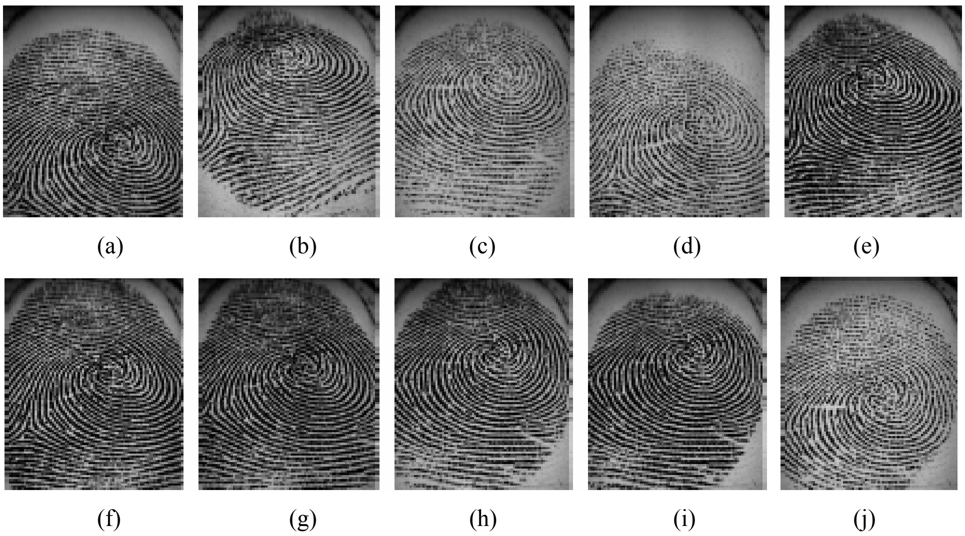
Figure 2 The sequence of variations captured for the ten images acquired in each session for face biometrics, (a) 1 (b) 2 (c) 3 (d) 4 (e) 5 (f) 6 (g) 7 (h) 8 (i) 9 (j) 10 (see online version for colours)

Note: The chronological ordering of samples are maintained for each session as indicated with the image number.

In each session, different types of temporary variations are captured for each modality. The sequence of covariates (temporary variations) captured in a single session for face images are listed in Table 3. Chronological ordering of the samples are maintained for each session. Figure 2 demonstrate the sequence of facial images captured in a single session for a randomly selected subject.

For fingerprint, no specific sequence is followed but temporary variations like rotation towards left, rotation towards right, partial print and possible non-linear skin deformations are introduced in each session as shown in Figure 3.

Figure 3 The sequence of variations captured for the ten images acquired in each session for fingerprint biometrics, (a) 1 (b) 2 (c) 3 (d) 4 (e) 5 (f) 6 (g) 7 (h) 8 (i) 9 (j) 10

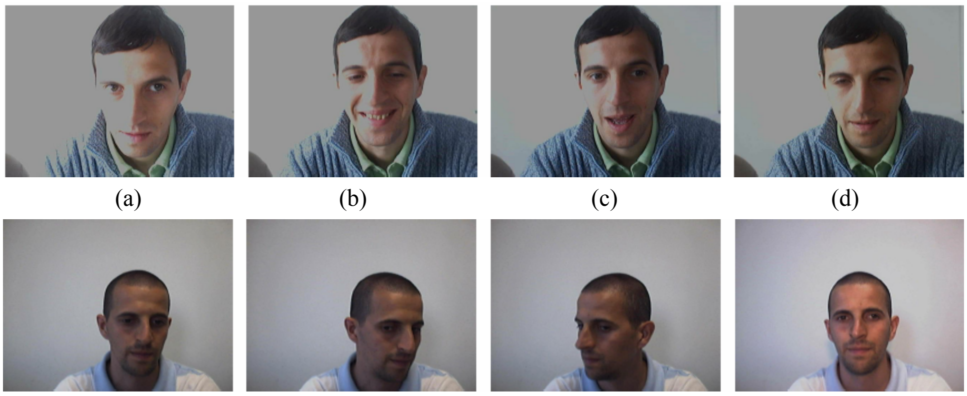


Note: The chronological ordering of samples are maintained for each session as indicated with the image number.

Two consecutive sessions are captured in about three weeks time, thus capturing temporal variations as well. Figure 4 shows the example face images taken from the different sessions of a randomly selected subject. Visual inspection confirms the presence of temporal variation (ageing, see Figure 4).

In Table 4, we summarise the performance of baseline face and fingerprint system. The scores for each modality are computed using Verifinger and Verilook SDK. This baseline performance is computed by comparing the enrolled samples with rest of the remaining samples from different batches. The number of enrolled samples is two, taken from the first session and matched to 98 test samples from the remaining different sessions. Accordingly, genuine and impostor scores account to be 180 and 8,640 respectively, for each subject in the database. The baseline error [equal error rate (EER)] for face, index finger and thumbprint modalities are 27.47%, 13.72% and 9.49%, respectively. *Though rarely mentioned, performance of the baseline system (without adaptation) is an important indicative of the appropriateness of the database for the study of adaptive biometric systems*. Specifically, very low baseline error indicates that initial enrolled templates are well representative of the variations of the remaining samples of the users. Accordingly, the adaptation process may not be required. On the contrary, high baseline error is an ideal condition for studying the process of adaptation. High baseline error for each of the modalities in the dataset (Table 4) confirms the presence of substantial variations in the samples that are unrepresentative of the enrolled templates.

Figure 4 Example face images taken from different sessions for a randomly chosen user from the dataset, (a) image 1 (b) image 2 (c) image 3 (d) image 4 (e) image 1 (f) image 2 (g) image 3 (h) image 4 (see online version for colours)



(a) (b) (c) (d)

(e) (f) (g) (h)

Note: On the visual inspection, the presence of temporal variations (ageing effect) is clearly visible across sessions.

Table 4 Error rate associated with baseline systems based on face and fingerprints modalities from the DIEE dataset

Modality	*Reference systems*	*EER (%)*
Face	VeriLook-SDK	27.47
Finger-thumb	VeriFinger-SDK	13.72
Finger-index	Veri_nger-SDK	9.49

Apart from the high baseline error, attributes of this proposed DIEE multi-biometric database such as:

1 large number of samples per subject

2 presence of both temporary as well as temporal variations

3 maintenance of the chronological ordering of sessions (the sessions are maintained in the order of time) and samples within each session makes this database appropriate for the study of adaptive biometric systems.

Note that due to the maintenance of chronological ordering of samples, both the online and offline versions of the adaptive biometric systems may be evaluated on this database.

3.2 A unified protocol for adaptive biometric systems

Section 2.3 presented several protocol attributes due to which existing adaptive biometric systems cannot be compared. In this light, we next describe a protocol which fixes the experimental attributes causing different experiments to be comparable. Further, this protocol simulates template adaptation for real biometric systems.

Let U_i represent an input batch of samples captured at session i. For instance, the batch U_1 consist of samples from the first session.

- Training:

  1 The system is trained with two enrolled samples per user from the first session, i.e., batch U_1. In other words $T \subset U_1$ and U_1 is re-defined as $U_1 = U_1 \setminus T$. Two templates are used in order to represent 'moderate' initial conditions for the baseline classifier.

  2 Updating threshold is set by estimating genuine and impostor score distributions from T.

- Adaptation:

  1 At each updating cycle i, batch U_i $i \in [1, 8]$ when available at slot i is used for updating the template set of the respective user using Algorithms 1 and 2.

  2 In order to simulate a realistic scenario where the probability of impostor attack is quite possible, five random impostor samples are introduced into each batch U_i.

  3 Score for each input sample with the template T are always computed using the mean rule (Ross et al. (2006)).

  4 Templates are updated by augmenting the existing template set with the genuine classified input samples (i.e., $T = T \cup U^*$, Algorithm 1) (Rattani et al., 2009). Further, offline template (selection) management strategies may be applied over the updated template set in order to obtain the reduced number of templates (Freni et al., 2008). The use of offline template selection schemes are beyond the scope of this manuscript.

  5 Updating threshold is re-estimated on the updated template set using the method similar to that adopted during training.

- Performance assessment:
  1. *Joint test and adapt strategy* is adopted for performance assessment. This strategy makes the best use of available samples per user basis (Rattani et al., 2009).
  2. Accordingly, after updating using the batch U_i, i.e., after each updating cycle i, batch U_{i+1} where $i + 1 \in [2, 9]$ is first used for testing the system performance. Batch U_{i+1} of the same user is used for evaluating the genuine score distribution. Batch U_{i+1} of all the other users is used for evaluating the impostor score distribution.
  3. The performance of the system is gauged using EER after each update cycle i, i.e., EER_i.
  4. After the evaluation of the system, the same batch U_{i+1} is used for the process of adaptation.

In comparison to the existing studies, the proposed protocol has the following properties:

- *Fixes the initial number of enrolled templates:* this is important as the baseline system trained on the fixed number of samples will exhibit the same initial performance. Training the classifier with few samples (two) from the first session will simulate the realistic condition where limited enrolment samples are available for training the classifier.
- *Reserves session implications:* i.e., first session is used for training the classifier and rest others for adaptation and testing.
- *Maintain the chronological ordering* of the samples within session and between the sessions acquired at different time slots.
- *Considers the probability of impostor attacks:* As five random impostor samples are included into the adaptation set used for template adaptation. Efficacy of the adaptive biometric system to the misclassification errors and adaptation using impostor samples can be taken into account after the update process is complete. Impact of misclassification error is assessed on examining EER value and visual inspection of the updated template set after each update cycle.
- *Allows evaluation over multiple update cycles:* Thanks to the adopted database and mentioned protocol, efficacy of the adaptive biometric systems can be evaluated over nine update cycles. In other words, performance of the adaptive biometric systems can be evaluated over *nine* independent test batch of samples available at different time stamps.

Many of the previous studies have adopted different values for the above mentioned attributes in their experimental analysis. As a consequence, either efficacy of the adaptation procedure cannot be simulated in the real operational environment or existing adaptive systems cannot be compared. To this front, we propose to benchmark the mentioned protocol as it has the potential to simulate the real biometric system. Possible adoption of the mentioned protocol will enable straight forward comparison of the adaptive biometric systems from different research groups. Importantly, cross-validation is not applicable for the evaluation of the adaptive biometric system (Rattani et al., 2013). Because samples (images) used for adaptation should exhibit temporary as well as temporal variations by maintaining the chronological ordering of the time. Reservation of the session implication and maintenance of the chronological ordering of time is not possible through cross-validation. Therefore, performing cross-validation will negate the merits of the adopted database and will fail to simulate the process of adaptation in the real operational environment. Further, note that the aim of the authors is not to propose a new protocol, but to highlight the merit of the existing protocol with the aim of benchmarking it.

4 Tools for performance evaluation

Supposing that a biometric system adopts an updating strategy, the following two pressing questions need to be addressed for evaluating the efficacy of adaptation:

- How much gain in the performance is expected? This in turn is dependent on the ability to capture operational data with drastic changes.
- Does this justify the risk of adopting the wrongly labelled samples (impostors)?

Accordingly,

- performance assessment metrics
- analysis of the updated template sets

should be adopted for the evaluation of the efficacy of the adaptive biometric systems. Next, we mention these measures in detail as follows:

4.1 Performance assessment metric

Majority of the state-of-the-art adaptive biometric methods have adopted the metrics of the traditional biometric system for the performance evaluation of the adaptive biometric systems.

In particular, EER and rank-one recognition performance are quoted in the existing studies. EER quantifies the probability of error at an operating threshold where the rate of false acceptance is equal to that of false rejection. EER is often quoted in a biometric verification scenario (Jain et al., 2007). The rank-one recognition performance, on the other hand, is quoted in a biometric identification scenario (Jain et al., 2007). It is defined as the probability of the target user indeed ranked top from the gallery of registered users.

However, only few of the research papers have explicitly quoted the obtained (%) performance gain of the adaptive biometric system over the baseline. In this paper, we define the performance gain as the amount of relative improvement with respect to the baseline system. We recommend the use of this secondary metrics, i.e., performance gain metric, for the fair comparison of the adaptive over baseline biometric system.

In order to handle differences in the basic adaptive systems, we propose a modification in the equation of the performance gain metric. For example, for the systems quoting EER, performance gain can be evaluated as:

$$Perf.gain(\%) = \frac{EER_b - EER_a}{EER_b} * 100 \tag{1}$$

where EER_a denotes the performance of the adaptive biometric system and EER_b is the performance of the baseline system.

Using this metric, the obtained performance gain of the adaptive biometric system over baseline can be explicitly reported in the research papers. This metric will allow fair comparison of the adaptive biometric systems following different mode of operation like verification or identification.

4.2 Analysis of the updated template set

The expected performance gain is related to the capability of the adaptive biometric system in capturing large number of samples with substantial intra-class variations. Thus, analysis of the updated template set will provide information/statistics indicating the efficacy of the adaptive biometric system in capturing variations. However, only few of the existing papers report statistics from the analysis of the updated template set (Rattani et al., 2008a, 2008b, 2012; Roli and Marcialis, 2006) for instance (%) of samples captured, (%) of impostor introduced, etc. On the other hand, these statistics should be accompanied with the experimental evaluation for in-depth insight into the efficacy of the proposed adaptive biometric systems.

Therefore, we propose to include the following statistics to facilitate the analysis of the behaviour and effectiveness of the updating algorithm:

- (%) of samples captured by the adaptive biometric system over time for each update cycle (Rattani et al., 2008a, 2008b, 2012)
- (%) of subjects affected by impostor intrusion (Rattani et al., 2008a, 2008b; Marcialis et al., 2008).
- (%) of subjects not updated (Marcialis et al., 2008).
- variance of the score statistics on the pair-wise comparison of the enrolled templates (Rattani et al., 2008a, 2008b).

Thus, we recommend the analysis of the updated template set to provide more insight into the functioning of the adaptation techniques.

4.3 Plots for performance visualisation

Similarly, the graphical plots illustrating the performance of the adaptive over baseline system should show the capability of the adaptive system in capturing samples and analysing the performance gain as a function of updated templates (i.e., number of captured samples) over time. However, majority of the papers do not report these plots. In fact, majority of the papers did not perform evaluation of the adaptive biometric systems over time. This is due to the unavailability of the large multimodal databases collected over a long time span. Thanks to the availability of the DIEE database, performance of the adaptive biometric system can be evaluated over time and the following statistics can be reported as follows:

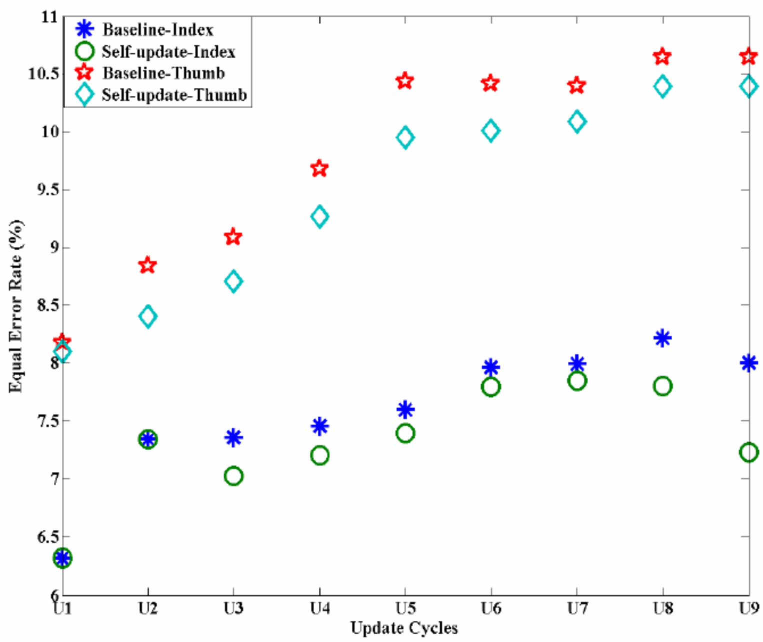
- *EER vs. number of samples used for adaptation:* This plot can be used to illustrate the performance of adaptive system as a function of number of samples appended to the template set over time. This plot can be used to infer the stage at which augmenting the training set increases, stabilises or reduces the performance of the system.
- *EER vs. update cycles:* This plot can be used to analyse EER as a function of update cycles. The performance of the adaptive system can be analysed over time for each update cycle. In addition, performance of the baseline can also be compared with the performance of adaptive biometric system (Rattani et al., 2012).

5 Example experimental results on the DIEE multi-biometric dataset

This section presents example experimental results giving a uniform view of the performance of the state-of-the-art template update algorithms like self- and co-update by using mentioned protocol and tools in Sections 3 and 4.

Using the mentioned dataset and the protocol, performance of the existing self- and co-update-based adaptive biometric system can be evaluated over time (Rattani et al., 2009) using EER vs. update cycle as well as EER vs. number of samples appended at each update cycle.

Figure 5 Performance of the self-updating thumb and index fingerprint systems under the assumption of impostor attack and threshold set at 0.001% FAR (see online version for colours)



Note: It can be seen that performance gain owing to adaptation is not much however the biometric classifier attains stability over time.

Figure 5 shows the performance of self-updating procedure for index and thumbprint biometrics using EER vs. update cycle plot. Performance of the self-adaptive system is also compared with the baseline matcher evaluated on the same test bed. The threshold for adaptation is set at 0.001% FAR. X axis shows the update cycles and Y axis shows the obtained EER (EER %).

Apart from performance assessment, analysis of updated template set provides important information for evaluating the efficacy of the adaptive systems. The obtained statistics reflect on the ability of template update scheme in capturing large number of samples for adaptation and its vulnerability to impostor intrusion. However, most of the previous studies did not report this statistics (Rattani et al., 2009; Uludag et al., 2004; Jiang and Ser, 2002; Ryu et al., 2006), also most of the papers did not mention the percentage of the users affected by impostors intrusion due to non-inclusion of impostor attack during the process of adaptation.

Table 5 present statistics regarding the percentage of users affected by impostor intrusion and percentage of users not updated at all for self-updating algorithm. These statistical measures are computed on the analysis of the updated template set for each user in the database.

Table 5 Statistics, i.e., percentage of users affected by impostor intrusion and percentage of users not updated at all are computed for self-updating algorithm

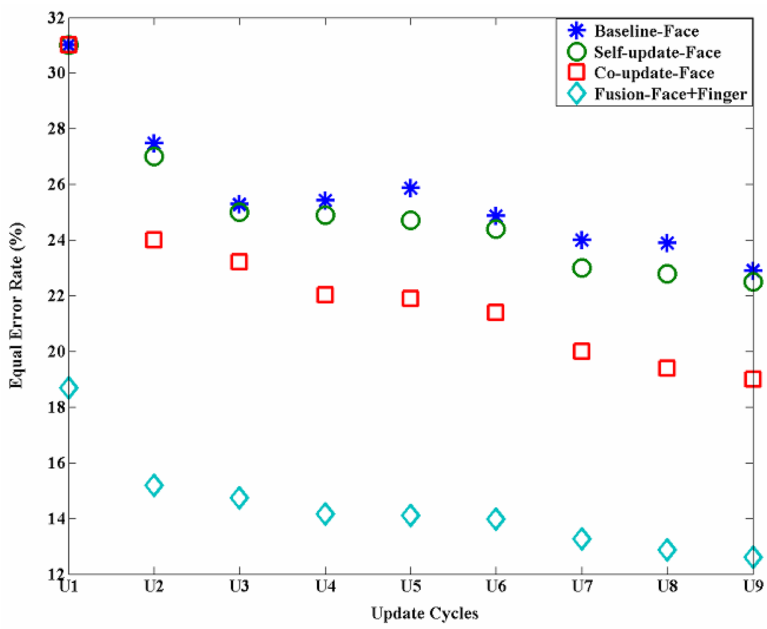
Modality	*(%) updating threshold*	*(%) of users affected by impostor intrusion*	*(%) of users not updated*
Self-updating index	0.001	4.08	12.02
Self-updating thumb	0.001	4.08	14.48
Self-updating face	0.001	0.2	9.2

Figure 6 shows the performance assessment of co-updating for face and fingerprint modalities over time using EER vs. update cycle curve. The performance of co-updating is also compared with the performance of self-updating-based technique. Legend 'fusion' in the subfigure [Figure 6(a)] correspond to performance of co-update face and fingerprint (multi-biometric) system fused at score level using the simple sum rule.

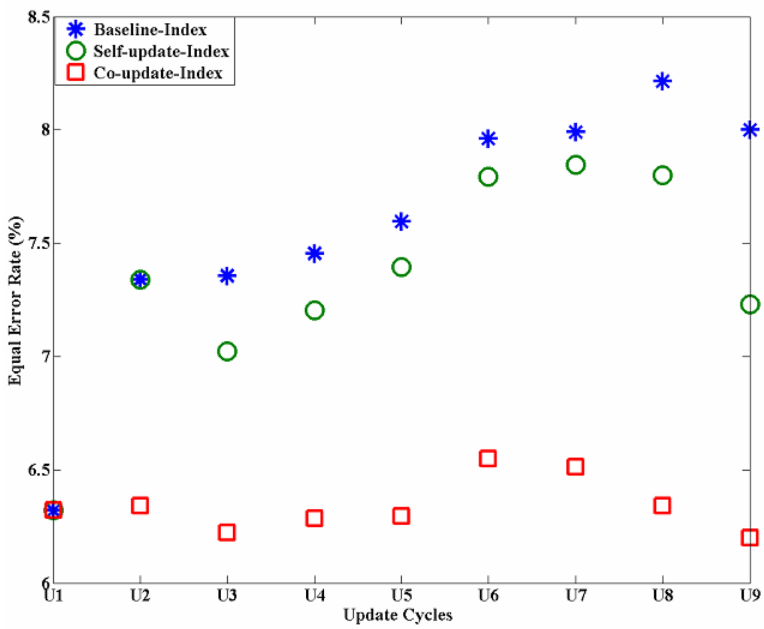
Most of the existing studies, could not report evaluation of the adaptive biometric system over time due to the adopted database and the protocol (Rattani et al., 2009; Uludag et al., 2004; Ryu et al., 2006; Liu et al., 2003; Roli and Marcialis, 2006). Evaluation over time indicates the capability of life long learning for adaptive biometric systems. Further, variance in the performance of adaptive system on evaluation over different test set (available over time) can be analysed using the adopted tools.

These experimental results indicate that operation at the stringent threshold, in order to avoid the possibility of adaptation using impostor samples, strongly impacts on the benefits achieved by self-training algorithm. Hence, self-updating may capture limited amount of samples and may result in limited performance improvement over time. Co-updating, on the other hand, using mutual and complimentary help of bi-modal systems can capture drastic changes in the input samples. Thus, irrespective of the stringent threshold settings and representativeness of the initial enrolled samples, co-updating guarantees enhanced and stable performance over time.

Figure 6 A EER vs. update cycle curve for the performance evaluation of co-updating in comparison to self-updating and baseline classifier over time for (a) face and (b) index fingerprint modalities (see online version for colours)



(a)



(b)

6 Conclusions

In order to establish a common platform for adaptive biometric systems, this paper proposed an appropriate DIEE multi-biometric dataset and a protocol with the aim of benchmarking them. The described DIEE multi-biometric database contains different type of temporary as well as temporal variations collected in different sessions, reserves session implication, maintains the chronological ordering of samples within each session and has the high baseline error for each modality. The time span of the database collection is 1.5 years. All these attributes mark this database appropriate for the study of adaptive biometric systems. Besides this dataset, the proposed protocol provides the fair evaluation and comparison of template update techniques over time, and the related tools allow quantitative evaluation of the performance gain obtained on making the biometric system. This benchmarking study will facilitate fair comparison and independent replication of the results from different research groups.

References

Amayeh, G., Bebis, G. and Nicolescu, M. (2009) 'Improving hand-based verification through online finger template update based on fused confidences', *Proc. of the 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems*, Vol. 13, Washington, DC, USA, pp.352–357.

Bailly-Bailliere, E., Bengio, S., Bimbot, S., Hamouz, M., Kittler, J., Mariethoz, J., Matas, J., Messer, J., Popovici, V., Poree, F., Ruiz, B. and Thiran, J.P. (2003) 'The BANCA database and evaluation protocol', *Proc. of 4th International Conf. Audio- and Video-based Biometric Person Authentication, AVBPA'03*, Vol. 2688, pp.625–638.

Blanz, V. and Vetter, T. (2003) 'Face recognition based on fitting a 3d morphable model', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp.1063–1074.

Connolly, J., Granger, E. and Sabourin, R. (2010) 'An adaptive classification system for video-based face recognition', *Information Sciences*, 1 June, Vol. 192, pp.50–70.

Flynn, P.J., Bowyer, K.W. and Phillips, P.J. (2003) 'Assessment of time dependency in face recognition: an initial study', *Proc. 4th Int. Conf. on Audio and Video based Biometric Person Authentication*, Guilford, UK, Vol. 2688, pp.44–51.

Freni, B., Marcialis, G.L. and Roli, F. (2008) 'Replacement algorithms for fingerprint template update', *Proc. of ICIAR08*, Springer, LNCS 5112, pp.884–893.

Garca, R. and Perales, J. (2008) 'Adaptive templates in biometrical authentication', *Proc. of 16th Int'l. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision*, Czech Republic.

Gayar, N., El Shaban, S.A. and Hamdy, S. (2006) 'Face recognition with semi-supervised learning and multiple classifiers', *Proc. of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, Venice, Italy, pp.296–301.

Giot, R., Dorizzi, B. and Rosenberger, C. (2011) 'Analysis of template update strategies for keystroke dynamics', *IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, CIBIM, Paris, France, pp.21–28.

Jain, A.K., Flynn, P. and Ross, A. (Eds.) (2007) *Handbook of Biometrics*, Springer Publishers, ISBN: 978-0-387-71040-2.

Jiang, X. and Ser, W. (2002) 'Online fingerprint template improvement', *PAMI*, Vol. 24, No. 8, pp.1121–1126.

Liu, X., Chen, T. and Thornton, S.M. (2003) 'Eigenspace updating for non-stationary process and its application to face recognition', *Pattern Recognition*, Vol. 36, No. 9, pp.1945–1959.

Luttin, J. (1998) *Evaluation Protocol for the XM2FDB Database (Lausanne Protocol)*, Communication 98-05, Tech. Rep., IDIAP, Switzerland.

Marcialis, G.L., Rattani, A. and Roli, F. (2008) 'Biometric template update: an experimental investigation on the relationship between update errors and performance degradation in face verification', *Joint IAPR Int'l. Workshop on SSPR+SPR08*, Springer, LNCS 5342, Orlando, Florida, USA, pp.684–693.

Martinez, C. and Fuentes, O. (2003) 'Face recognition using unlabeled data', *Computacion y Sistems, Iberoamerican Journal of Computer Science Research*, Vol. 7, No. 2, pp.123–129.

Patterson, E., Ricanek, K., Albert, M. and Boone, E. (2006) 'Automatic representation of adult aging in facial images', *Proc. of the 6th IASTED Int'l. Conf. on Visualization, Imaging, and Image Processing*, Palma de Mallorca, Spain, p.612.

Pavani, S., Sukno, F., Butako, C., Planes, X. and Frangi, A. (2009) 'A confidence based update rule for self-updating human face recognition systems', *Proc. of Int'l. Conf. on Biometrics*, LNCS 5558, Alghero, Italy, pp.151–160.

Poh, N., Kittler, J., Marcel, S., Matrouf, D. and Bonastre, J.F. (2010) 'Model and score adaptation for biometric systems: coping with device interoperability and changing acquisition conditions', *Proc. of Int'l. Conf. on Pattern Recognition*, Istambul, Turkey, pp.1229–1232.

Poh, N., Rattani, A. and Roli, F. (2012) 'Critical analysis of adaptive biometric systems', *IET Biometrics*, Vol. 1, No. 4, pp.179–187.

Rattani, A., Freni, B., Marcialis, G.L. and Roli, F. (2009) 'Template update methods in adaptive biometric systems: a critical review', *Proc. of 3rd Int'l. Conf. on Biometrics*, Sardinia, Alghero, Spain, Vol. 5558, pp.847–856.

Rattani, A., Marcialis, G.L. and Roli, F. (2008a) 'Biometric template update using the graph mincut: a case study in face verification', *Proc. of 6th IEEE Biometric Symposium*, Tampa, USA, pp.23–28.

Rattani, A., Marcialis, G.L. and Roli, F. (2008b) 'Capturing large intra-class variations of biometric data by template coupdate', *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, Alaska, USA, pp.1–6.

Rattani, A., Marcialis, G.L. and Roli, F. (2013) 'Biometric system adaptation by self-update and graph-based techniques', *Journal of Visual Languages and Computing*, Vol. 24, No. 1, p.19.

Rattani, A., Marcialis, G.L., Granger, E. and Roli, F. (2012) 'A dual-staged classification-selection approach for automated update of biometric templates', *Proc. of 21st International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, pp.2972–2975.

Roli, F. and Marcialis, G.L. (2006) 'Semi-supervised pca-based face recognition using self training', *Proc. of Joint IAPR Int. Workshop on S+SSPR06*, Springer, LNCS 4109, HongKong, China, pp.560–568.

Roli, F., Didaci, L. and Marcialis, G.L. (2007) 'Template co-update in multimodal biometric systems', *Proc. of IEEE/IAPR Int'l. Conf. on Biometrics*, Springer, LNCS 4642, Seoul, Korea, pp.1194–1202.

Ross, A., Nandakumar, K. and Jain, A.K. (2006) *Handbook of Multibiometrics*, Springer Publishers, 1st ed., ISBN: 0-3872-2296-0.

Ryu, C., Hakil, K. and Jain, A. (2006) 'Template adaptation based fingerprint verification', *Proc. of the 18th Int'l. Conf. on Pattern Recognition*, Hong Kong, Vol. 4, pp.582–585.

Tan, X., Chen, S., Zhou, Z. and Zhang, F. (2006) 'Face recognition from a single image per person: a survey', *Pattern Recognition*, Vol. 39, No. 9, pp.1725–1745.

Uludag, U., Ross, A. and Jain, A.K. (2004) 'Biometric template selection and update: a case study in fingerprints', *Pattern Recognition*, Vol. 37, No. 7, pp.1533–1542.

Zhou, X., Kittler, J. and Messer, K. (2007) 'Illumination invariant face recognition: a survey', *Proc. of Int'l. Conf. on Biometric, Theory, Applications and Systems*, Crystal City, VA, pp.1–8.

Zhu, X. (2008) *Semi-supervised Learning Literature Survey*, Tech. Rep., Computer Science TR 1530.