HAL
open science

Hybrid Template Update System for Unimodal Biometric Systems

Romain Giot, Christophe Rosenberger, Bernadette Dorizzi

To cite this version:

Romain Giot, Christophe Rosenberger, Bernadette Dorizzi. Hybrid Template Update System for Unimodal Biometric Systems. IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2012), Sep 2012, Washington, District of Columbia, USA, France. pp.1–7. hal-00714241

HAL Id: hal-00714241

https://hal.science/hal-00714241v1

Submitted on 3 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hybrid Template Update System for Unimodal Biometric Systems

Romain Giot and Christophe Rosenberger
Université de Caen, UMR 6072 GREYC
ENSICAEN, UMR 6072 GREYC
CNRS, UMR 6072 GREYC
{romain.giot,christophe.rosenberger}@ensicaen.fr

Bernadette Dorizzi
Institut Mines-Telecom sudParis
UMR 5157 SAMOVAR
bernadette.dorizzi@it-sudparis.eu

Abstract

Semi-supervised template update systems allow to automatically take into account the intra-class variability of the biometric data over time. Such systems can be inefficient by including too many impostor's samples or skipping too many genuine's samples. In the first case, the biometric reference drifts from the real biometric data and attracts more often impostors. In the second case, the biometric reference does not evolve quickly enough and also progressively drifts from the real biometric data. We propose a hybrid system using several biometric sub-references in order to increase performance of self-update systems by reducing the previously cited errors. The proposition is validated for a keystroke-dynamics authentication system (this modality suffers of high variability over time) on two consequent datasets from the state of the art.

1. Introduction

Biometric authentication systems allow authenticating individuals by comparing a query provided by the claimant to its biometric reference. Depending on the result of this comparison, the claimant is accepted (the system asserts he/she owns the identity he/she claims) or rejected (the system does not assert he/she owns the identity he/she claims). Usually, the biometric reference is created during the enrollment phase by providing one or several captures. However, most biometric modalities are not permanent and system performance decreases with time. To overcome this drawback, it is possible to re-enroll the user at a fixed time. Sadly, this method has a high cost because it needs time and operators. Enrollment period may also be on a too short timespan to collect enough intraclass variabilities to represent the user as best as possible.

The aim of semi-supervised template update systems is to address these issues by automatically updating the biometric reference of individuals while they use the system. The update system only uses information from the query and from the biometric recognition system. Semi-supervised template update is an active field of research mainly studied for morphological modalities, whereas they are less subject to variabilities than the behavioral ones. As such systems can include impostor's samples in the updated biometric reference, the biometric reference can progressively deviates from the owner's real biometric data and the system attracts more impostors. There are two kinds of systems in the literature, the self-update systems [12, 15, 5] and the co-update systems [13, 1]. Self-update systems allow unimodal system to update the reference automatically after collecting unlabelled data. Their main drawback is the fact that they are not able to attract genuine samples to much dissimilar than the original reference one [10], so they miss several samples during the update procedure. Co-update systems allow using multimodal systems in order to attract these forgotten samples because one biometric reference is updated based on the classification result of the complementary classifier related to the other biometric reference linked to the other modality.

We propose a new hybrid template update system. There are several components in a template update system. Our hybrid system is not clearly based on the optimisation of one particular component. We can see it as both a modification of the way of representing the biometric reference, and the way of updating the user's gallery. A user is represented by several biometric sub-references evolving in parallel by using different template update methods.

The contributions of this work are the following ones: (1) we propose an original hybrid template update system scheme performing better than the classic self-update system from the state of the art. This is an hybrid system because (a) it operates fusion as in co-update systems, whereas it is a self-update system, and (b) user's biometric reference is composed of several biometric sub-references; (2) we propose two metrics in order to evaluate the efficiency of template update systems over several sessions; (3) we evaluate the method with a dataset providing more samples per user than most studies of the state of the art.

The paper is organized as follows. Section 2 gives a quick

overview of the recent works on template update. Section 3 presents the template update architecture we propose, as well as two new evaluation metrics. Section 4 presents the selected protocol to evaluate our contribution. Section 5 presents the experimental results and Section 6 concludes this communication.

2. Similar and Recent Works

In this section, we present the most recent works in template update. Bhatt *et al.* present a co-update method allowing to update two related classifiers in an online way [1]. The SVM boundary decision is updated in a semi-supervised way. Namely, if a classifier returns with a high probability that a sample corresponds to a particular label, whereas the second classifier disagrees, this new sample is used for an online update of the second classifier. The authors show, on a face recognition problem, that their system improves performance both in accuracy and in computational time. The validation is done on an aggregated database of 1833 subjects providing 20150 images. There is an average of 11 images per individual, which can be considered as small for a template update study.

Rattani *et al.* present self-update and co-update for biometric modalities where a biometric sample can be used as a biometric reference [12]. Such kind of information can be irrelevant for some biometric modalities, like keystroke dynamics, which are not consistent enough to work with only one sample. They analyse the behavior of the updating method by representing the samples as nodes in a graph and similarities as edges between nodes. They show that the graph can contain independent sub-graphs. Samples from a sub-graph cannot attract samples from other sub-graphs as they are too much dissimilar. The samples present in other sub-graphs contain more variabilities but will not be used in the template update system. Co-update allows attracting these samples. The study is done with 40 users providing each 50 samples (of face and fingerprint) on 5 sessions captured on 1.5 years.

Seeger and Bours list various factors used to specify an evaluation scenario of a template update system for keystroke dynamics [15]. Note that most of the results of this paper are also relevant for other modalities, and therefore this paper is worth reading. The authors show that different evaluation scenarios give different interpretation of the template update system performance. This is a problem, because almost no template update study uses the same kind of scenario and because most studies do not explain which scenario configuration has been chosen.

Giot *et al.* raise some questions, without answering them, about the evaluation of template update systems [5]. They show that, in addition to the scenario parameters presented in [15], most studies also present a great variability in the way of computing the performance of the template update system. They use three different ways encountered in the template update literature to evaluate the performance of a keystroke dynamics template update system using exactly the same set of scores. They show that different interpretations can be proposed whereas the scores are identical. These recent works assert the fact it is necessary to clearly specify the way of computing the performances, and the need of standardised evaluation procedures.

3. Proposed Semi-supervised Template Update Method and Evaluation Metrics

This section presents the proposed template update component and associated evaluation metrics.

3.1. Template update based on multiple galleries evolution

Here are some definitions for the paper. A *user's gallery* is a set of biometric samples used to represent a user, while a *biometric reference* is a model representing a user and has been computed with the samples of its gallery. These two different terms are both named model, biometric reference, or template in the literature.

Our contribution is inspired by the co-update systems [13, 1], although we use a mono-modal system, and the various works on gallery update [14, 6]. In all previous works, the biometric reference of the user is unique because a user is represented by only one gallery or one sample or one model. But in our work, the biometric reference is composite: this *biometric meta-reference* contains several *biometric sub-references* evolving with various template update methods, but authentication will be done with a unique biometric authentication method (whereas in multibiometrics, people use a multi-algorithm scheme when only one modality is used). Fig. 1 summarizes the proposed system (green area) for a system using two biometric sub-references per user (*i.e.*, two different biometric template update systems evolve in parallel), and tab. 1 presents the difference between self-update, co-update and hybrid-update.

The defined system is independent of the other components of a template update system (pink and blue areas in fig. 1). When a query is compared to the biometric meta-reference of the claimant, it is in fact compared to each biometric sub-reference. The scores are fused in order to obtain one aggregated score. We assume nothing on the update decision method; it can be based on a double thresholding method, a quality index, or anything else. With a double thresholding method, the decision is taken on the aggregated score, so we do not know which of the biometric sub-reference is responsible of the update decision. This is not a problem because we are in a monomodal system and the comparison scores produced by the comparison to the various biometric sub-references must be highly correlated;

Table 1. Self-update, co-update and our hybrid-update behaves all differently

	Self-update	**Co-update**	**Hybrid-update**
One modality	yes	no	yes
Several types of classifiers	no	yes (one per modality)	implementation choice (one per sub-reference or the same for all)
Update decision source	classifier score	disagrement between the two classifiers scores	aggregated score

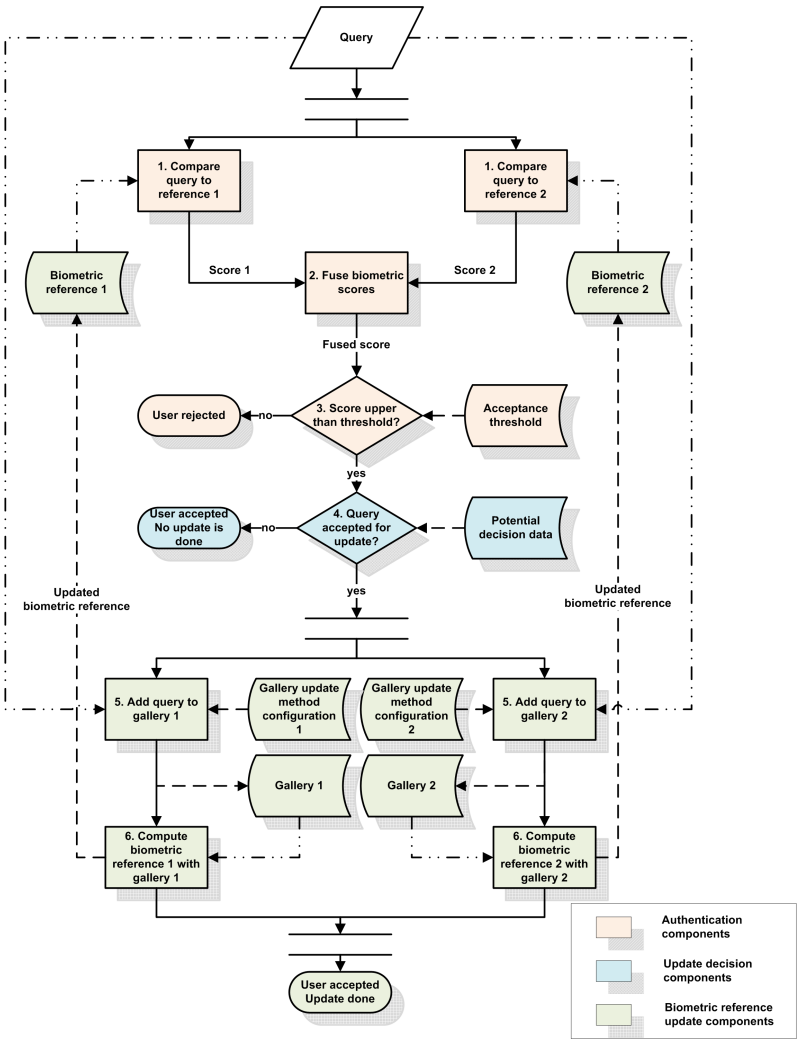

Figure 1. Workflow of the hybrid template update system (when two template update systems are used, in an online scenario).

we expect the opposite in co-update systems. When the biometric meta-reference must be updated, we update each of its biometric sub-reference (and not the less influent one as in co-update) using the accepted query and the template update method specific to each biometric sub-reference. We expect this way to decrease the updating errors. Each couple of gallery update and method to compute the biometric reference can be replaced by an online classifier update [1]. In this case, each online classifier must be different in order not to evolve identical biometric sub-references, as the aggregated score would be the same than the score of each online classifier. Different fusion rules and biometric sub-references updating methods can be used in this new update procedure.

3.2. Proposed Evaluation Metrics

There is a lack of evaluation metrics for template update in the literature [5]. Rattani *et al.* use the ratio of impostors'

samples present in the gallery after the update [10]. It is used in an offline update procedure, whereas we want to evaluate the system in an online way (*i.e.* the ratio of impostors can evolve after each query presentation to the biometric meta-reference). Poh *et al.* explain how to estimate the authentication performance over time [8]. This procedure requires a dataset where samples are very well spread on a large time span, which is never the case when samples are acquired among various sessions (a lot of samples on a short time span and no samples at all on a long time span).

To overcome these issues, we propose two evaluation metrics: (i) the *Impostor Update Selection Rate* (IUSR) which corresponds to the ratio of impostor's samples involved in the update process among all the tested impostor's samples; (ii) the *Genuine Update Miss Rate* (GUMR) which corresponds to the ratio of genuine's samples not involved in the update process among all the tested genuine's samples. Say, we have N_t, N_i, N_g respectively the total number of tested samples, the number of tested impostor samples and the number of tested genuine's samples ($N_t = N_i + N_g$). Say, we have U_i, U_g respectively the total number of impostor's samples selected in the updating process and the number of genuine's samples selected in the updating process. The error rates can be estimated as follows:

$$\widehat{IUSR} = \frac{U_i}{N_i} \tag{1}$$

$$\widehat{GUMR} = \frac{N_g - U_g}{N_g} \tag{2}$$

In a system without template update mechanism, $IUSR = 0$ and $GUMR = 1$. The best template update systems tend to have a $IUSR$ as close as possible to 0 because the inclusion of impostor's samples is problematic as the biometric reference will attract more easily impostors. It also have the lower $GUMR$ as possible, but not equal to zero, because missing genuine samples can be a good thing when these samples are too noisy. Next section presents the configuration we have chosen to evaluate our new update procedure.

4. Protocol

This section presents the precise configuration we have defined to evaluate our new template update procedure.

4.1. Parametrization

Various parameters must be configured in order to evaluate a template update system and reproduce the study [15, 5]. Tab. 2 summarises the various parameters used in the experiment. We have chosen to evaluate the proposed system on a behavioral modality which presents more important temporal variations than a morphological modality. Among the available behavioral modalities, keystroke dynamics is the one having the biggest datasets in term of number of sessions. The template update system is evaluated for each session using only the scores computed during this session. In order not to give over-optimistic results [5], we do not average the performance of each session with the performance of the previous sessions. As the set of queries tested against a biometric reference is randomly built, results can vary among the runs. To cope with this variability, we launch the experiment 100 times and present the averaged results. Two different keystroke dynamics datasets are used in order to validate the proposed work. We use the DSL2009 [7] (51 users, 400 samples per user, 8 sessions) and the GREYC2009 [4] (100 users, 60 samples per user, 5 sessions) databases. Due to the lack of space, only results on DSL2009 are presented, but conclusions are similar for GREYC2009. Session 1 is used for the enrolment stage, *i.e.* to generate the initial biometric sub-references (user's gallery size is the number of samples per session per user). The other sessions serve to test and update the biometric meta-reference.

Table 2. Experiment parameters.

Parameter	**Value**
Modality	Keystroke dynamics
Authentication method	Distance computing [2]
Update decision	Online double threshold semi-supervised
Update threshold	Empirically fixed
Update mechanism (of sub-references)	None, sliding window, growing window
Number of sub-references	A biometric meta-reference is composed of 2 biometric sub-references
Fusion of references comparison distances	Mean value, minimum value (as we work with distance)
Aggregation combinations	(None, Sliding), (None, Growing), (Sliding, Growing)
Number of sessions	8 on DSL2009, 5 on GREYC2009
Respect to chronology	Yes
Presentation orders	Random
Input size	30% of impostors
Evaluation computing	Online (*i.e.* joint adapt-and-test strategy per session [9])
Evaluation metrics	EER, FNMR, FMR, IUSR, GUMR (scores of current session, no error average with previous sessions)

4.2. Configuration

The template update methods are based on simple gallery update methods as in [6]. Each time a gallery is modified, the associated biometric sub-reference is re-computed from

scratch. Three gallery update methods are used: (i) *none*, the gallery is not modified, there is no update; (ii) *sliding window*, the selected query replaces the oldest sample of the gallery; (iii) *growing window*, the selected query is added to the gallery.

Three gallery aggregations methods are used: (a) *parallel sliding*, where one biometric reference is never updated, and the other one is updated with the sliding window; (b) *parallel growing*, where one biometric reference is never updated, and the other one is updated with the growing window; (c) *double parallel*, where one biometric reference is updated using the sliding window, and the other one is updated using the growing window. Aggregation methods (a) and (b) produce two biometric sub-references following this rule: in the best case, one biometric sub-reference represents the behavior of the user at the initial enrolment, while the other represents its very last way of typing.

Two score fusion methods are used: (1) the mean of the scores, and (2) the minimum value of the scores[1], and each gallery aggregation method is used using each score fusion method.

4.3. Evaluation

The question is how to qualify if an updating system performs well ? We will see that, for keystroke dynamics systems, using no update results in a FNMR reduction with time (keystroke dynamics must be one of the rare biometrics having such behavior, because of the typing habituation, but we think that the FNMR is expected to increase after several additional sessions) and an increasing FMR (impostors are also expected to type better the password). A good update system is a system where FNMR and FMR both decrease (or remain stable) over time. In addition to these measures, we also present the IUSR and GUMR which provide information on the updating errors. The update decision is based on the similarity score, so FNMR/IUSR and FMR/FGMUR errors can be correlated. The EER is also used because of its ease of reading.

5. Experimental Results

The baseline scenario without template update is "None", and the baseline scenarios with template update are the self-updates with the "Sliding" and "Growing" gallery management ; they correspond to previous works published in [3]. Our new contributions in this paper are the other ones ("Parallel sliding", "Parallel growing", "Parallel both", "Parallel min sliding", "Parallel min growing", "Parallel min both").

It is well known that decreasing the FNMR of a biometric system correspond to increasing the FMR (and vice versa). We can observe a similar behavior, linked to the time, on fig. 3. Methods allowing decreasing the FNMR over time

[1]The recognition method produces dissimilarity scores

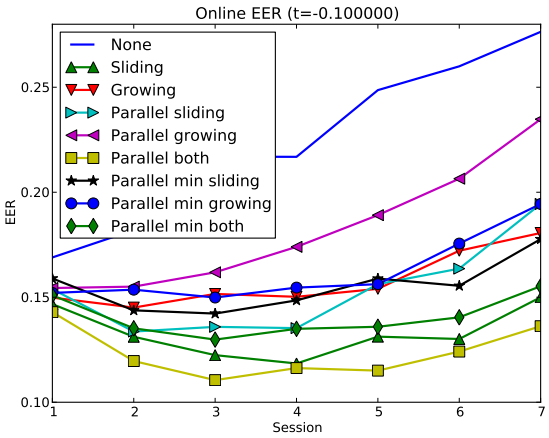

Figure 2. EER over sessions for each of the template update systems.

tend to increase the FMR over time. As the EER evolution cannot give us such kind of information (see fig. 2), we think that, in opposition to previous papers [11, 5], providing the EER of template update systems may be not be a good idea, and it would be better to provide the FNMR and FMR in order to see their difference of evolution. In addition, when using a double threshold mechanism, the EER threshold can be incompatible with the update threshold.

Looking on the fig. 3 (and not taking into account fig. 2, even if it could assert that too), we see that the "Parallel both" method is the most appropriate. It is not the best method in term of FMR nor FNMR, but it is the sole method presents in the best methods each time. As it seems to be a good compromise, we can say that using several sub-references improves performances against using only one ("growing" and "sliding"). To assert this conclusion, we manually ranked each update method (sorted by global performance) on the following rates: FMR, FNMR, EER, FISUR, GUMR. For each update method, we sumed all the ranks of the various criteria and sort them. Tab. 3 presents the ranking results. We have also computed the ranks without using the EER. We think we cannot trust the EER values, because (i) it may be hard to configure the system with the thresholds allowing to obtain the EER; (ii) the EER threshold may be incompatible with the one used for the update decision. Although ranks are different with the two ways of computing, the two bests and two worsts methods are the same. The two best methods are *parallel both* and *parallel min both*, which is the proposed method when we evolve in parallel two biometric sub-references using the growing window and the sliding window. It shows the benefit of the proposed method when evolving different biometric sub-references. The two worst methods are *Parallel min growing* and *None*. It is easy to understand. In the first case, there are two biometric sub-references: the initial reference which quickly becomes not

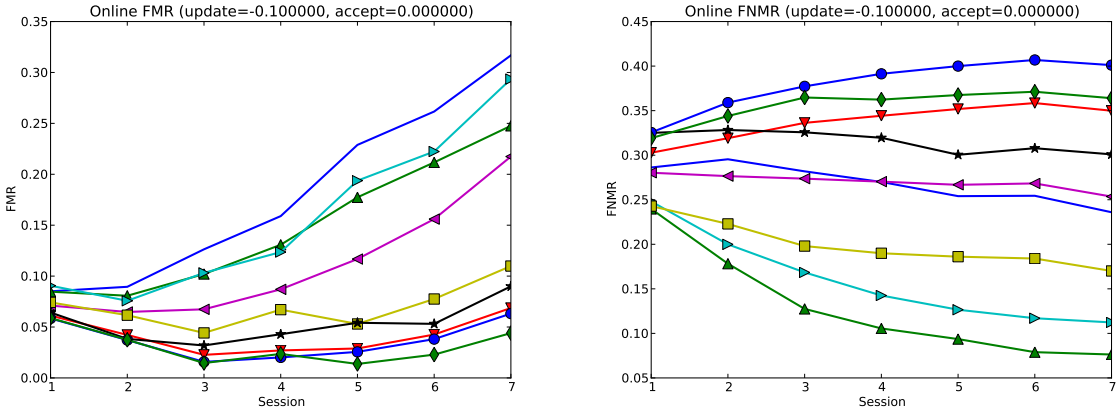

Figure 3. FNMR and FMR over sessions for each of the template update systems (accept threshold of 0.0, update threshold of -0.1).

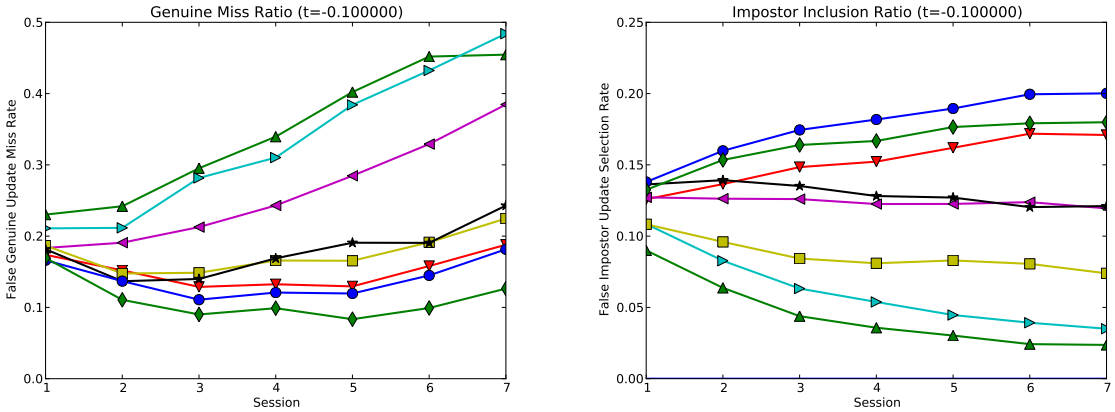

Figure 4. Update Error over sessions (update threshold of -0.1).

Table 3. Manual ranking of each method among various criteria. Top three methods using the EER in the ranks sum are in bold. Top three methods without EER are in italics.

Method	**FMR**	**FNMR**	**EER**	**FISU**	**GMN**	**Score**	**Rank**	**Score without EER**	**Rank**
Parallel both*	5	3	1	4	4	17	**1**	16	*2*
Parallel min both*	1	8	3	8	1	18	**2**	15	*1*
Sliding	7	1	2	2	8	23	**3**	21	4
Parallel sliding*	8	2	4	3	7	24	4	21	4
Growing	3	7	6	7	3	26	5	19	*3*
Parallel min sliding*	4	6	5	6	5	26	5	21	4
Parallel growing*	6	4	8	5	6	29	7	21	4
Parallel min growing*	2	9	7	9	2	29	7	22	8
None	9	5	9	1	9	33	9	24	9

representative and results in rejecting genuine samples and the growing window which can contains and keep a lot of impostor samples. This behavior is explained in fig. 4 where we see that this method is the one attracting the highest number of impostors.

Fig. 4 shows that, for most methods, having a higher IUSR implies a lower GUMR (and vice-versa) except for *parallel both*, which is never the best method, but is always in the top methods. It is the only method attracting not too many impostors and rejecting not too many genuine samples. The same behavior is observed on fig. 3. Thus, being the method having not too much FMR and not much FNMR, the EER is low in comparison to other methods.

Fig. 4 and fig. 3 show that there is a strong relationship

between FNMR and GUMR, and FMR and IUSR. This proves that to reduce the FNMR (respectively FMR), it is necessary to reduce the GUMR (respectively IUSR). There is one limit of the present evaluation procedure which is linked to the chosen update selection procedure. As we use a double threshold scheme, it is necessary to specify the two thresholds. Results would be different with thresholds performing badly. This may be an issue in an operational scenario (the optimum thresholds may be hard to obtain). A good practice would be to compute the threshold of a selected operational point using enrolment samples of all users, and compute the update threshold using it (using the EER threshold computed with first session divided by 2 gives us similar results).

6. Conclusion

We have presented a hybrid template update method allowing to update several biometric references in parallel. The parallel evolution of biometric sub-references allows reducing the update error rates and the performance decreases over time in comparison to classic methods using one reference. The method has been validated on a template update system for keystroke dynamics on two datasets. One of the datasets contains 400 samples per users which is larger than most studies from the state of the art for template update of morphological modalities. We have shown that our scheme gives better performance than the classical ones (self-update with sliding or growing windows). Although the method has been evaluated in an online semi-supervised scenario, it could be used in offline scenarios or supervised scenarios too. The implementation uses two sub-references, but it would be useful to analyse if using more sub-references would improve the performances. It would be interesting to validate the proposition in other contexts and other modalities (signature for example), as well as with online classifiers instead of methods using a gallery and update decision methods.

References

[1] H. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Noore, and A. Ross. On co-training online biometric classifiers. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7, 2011.

[2] T. de Magalhaes, K. Revett, and H. Santos. Password secured sites: stepping forward with keystroke dynamics. In *International Conference on Next Generation Web Services Practices*, 2005.

[3] R. Giot, B. Dorizzi, and C. Rosenberger. Analysis of template update strategies for keystroke dynamics. In *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). Special Session on Adaptive Classification Systems for Biometric Recognition.*, pages 21–28, 2011.

[4] R. Giot, M. El-Abed, and C. Rosenberger. Greyc keystroke: a benchmark for keystroke dynamics biometric systems. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, 2009.

[5] R. Giot, C. Rosenberger, and B. Dorizzi. Performance evaluation of biometric template update. In *International Biometric Performance Testing Conference 2012*, 2012.

[6] P. Kang, S.-s. Hwang, and S. Cho. Continual retraining of keystroke dynamics based authenticator. In S.-W. Lee and S. Li, editors, *Proceedings of ICB 2007*, volume 4642 of *Lecture Notes in Computer Science*, pages 1203–1211, 2007.

[7] K. Killourhy and R. Maxion. Comparing anomaly-detection algorithms for keystroke dynamics. In *IEEE/IFIP International Conference on Dependable Systems & Networks, 2009. DSN'09*, pages 125–134, 2009.

[8] N. Poh, J. Kittler, R. Smith, and J. Tena. A method for estimating authentication performance over time, with applications to face biometrics. In *12th Iberoamerican Congress on Pattern Recognition CIARP*, 2007.

[9] N. Poh, R. Wrong, J. Kittler, and F. Roli. Challenges and research directions for adaptive biometric recognition systems. In *Advances in Biometrics*, pages 753–764, 2009.

[10] A. Rattani, G. Marcialis, and F. Roli. Biometric template update using the graph mincut algorithm : A case study in face verification. In *Biometrics Symposium, 2008. BSYM '08*, pages 23 –28, Sept. 2008.

[11] A. Rattani, G. Marcialis, and F. Roli. Self adaptive systems: An experimental analysis of the performance over time. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on*, pages 36 –43, 2011.

[12] A. Rattani, G. Marcialis, and F. Roli. Temporal analysis of biometric template update procedures in uncontrolled environment. In G. Maino and G. Foresti, editors, *Image Analysis and Processing – ICIAP 2011*, volume 6978 of *Lecture Notes in Computer Science*, pages 595–604. 2011.

[13] F. Roli, L. Didaci, and G. Marcialis. Template co-update in multimodal biometric systems. In *International Conference on Biometrics (ICB 2007)*, pages 1194–1202, 2007.

[14] T. Scheidat, A. Makrushin, and C. Vielhauer. Automatic template update strategies for biometrics. Technical report, Otto-von-Guericke University of Magdeburg, Magdeburg, Germany, 2007.

[15] M. M. Seeger and P. Bours. How to comprehensively describe a biometric update mechanisms for keystroke authentication. In *3rd International Workshop on Security and Communication Networks (IWSCN 2011)*, 2011.