A Novel Approach to Combining Client-Dependent and Confidence Information in Multimodal Biometrics

Norman Poh and Samy Bengio

IDIAP Research Institute, Rue du Simplon 4, CH-1920 Martigny, Switzerland
norman@idiap.ch, bengio@idiap.ch

Abstract. The issues of fusion with client-dependent and confidence information have been well studied separately in biometric authentication. In this study, we propose to take advantage of both sources of information in a discriminative framework. Initially, each source of information is processed on a per expert basis (plus on a per client basis for the first information and on a per example basis for the second information). Then, both sources of information are combined using a second-level classifier, across different experts. Although the formulation of such two-step solution is not new, the novelty lies in the way the sources of prior knowledge are incorporated prior to fusion using the second-level classifier. Because these two sources of information are of very different nature, one often needs to devise special algorithms to combine both information sources. Our framework that we call "Prior Knowledge Incorporation" has the advantage of using the standard machine learning algorithms. Based on $10 \times 32 = 320$ intramodal and multimodal fusion experiments carried out on the publicly available XM2VTS score-level fusion benchmark database, it is found that the generalisation performance of combining both information sources improves over using either or none of them, thus achieving a new state-of-the-art performance on this database.

1 Introduction

Previous studies have shown that combining several biometric authentication systems is a potential way to improve the overall system accuracy [1]. It has also been shown that fusion with client-dependent and confidence information can *further* improve the system performance. Studies using *client-dependent information* include client-dependent threshold [2], model-dependent score normalisation [3] or different weighing of expert opinions using linear [4] or non-linear combination [5] on a per client model basis. Some of the existing approaches to incorporate the *confidence or quality information* are a multivariate polynomial regression function [6], a statistical model (that reconciles expert opinions) [7] and a modified Support Vector Machine algorithm [8]. Specific to speaker authentication, in [9], the first formant of speech was used as an indicator of quality to weigh the Log-Likelihood Ratio (LLR) of each speech frame. Thus, instead of taking the average LLR as commonly done, a weighted average LLR was used. These studies have shown that incorporation of client-dependent and confidence information are important means to improve multimodal biometric systems.

In this study, we would like to verify whether fusion using both of these sources of information is more beneficial than using either one or none at all. To the best of our knowledge, this issue has not been examined before. This is perhaps because these two sources of information are very different, and strategies employed to integrate one source of information is completely different from or incompatible with the other. We propose a novel way to fuse these two sources of information in two steps: first incorporate the prior knowledge on a per expert basis and then combine them using a second classifier. The idea of using a second classifier is not new. This strategy is called post-classification in [10]. However, deriving ways to incorporate the prior knowledge into the scores, on a per expert basis, prior to fusion is new. This framework is called "Prior Knowledge Incorporation" (PKI). It should be noted that the prior knowledge incorporated scores, on their own, may not necessarily be very useful if not further combined with other scores. The advantage of this technique is that, due to PKI scores, (the first step), information sources can be combined independently. In terms of implementation, this means modular integration is possible. Secondly, the second-level classifier can be implemented using standard off-the-shelf machine-learning algorithms, thus eliminating the need to create a specific fusion algorithm for this purpose. In principle, any sources of prior knowledge can be combined this way. In practice, the amount of prior knowledge possibly employed is limited by the information given by the baseline expert systems.

In order to verify this hypothesis, three sets of fusion control experiments were carried out, i.e., fusion using the original expert scores, fusion using client-dependent normalised scores and fusion using confidence. These baseline experiments are then compared to fusion using all the available information sources. Based on 32 fusion data sets taken from the publicly available XM2VTS score fusion benchmark database [11], it is concluded that fusion with both sources of information is more beneficial than using either one or none of them.

This paper is organised as follows: Sections 2 and 3 discuss briefly how the client-dependent information and confidence information can be computed, on a per expert basis. Section 4 discusses how these seemingly different sources of information can be fused together using the PKI framework. The database and results are presented in Sections 5 and 6, respectively. They are followed by conclusions in Section 7.

2 Deriving Client-Dependent Information

There exists a vast literature in this direction. A survey can be found in [12, Sec. 2]. There are two families of approaches, namely, score normalisation and threshold normalisation. The former aims at normalising the score such that a global decision threshold can be found easily. The latter manipulates the decision threshold directly. It has been shown that [12] both families are dual forms of each other. The disadvantage of the latter category is that it is dependent on a specific cost of false acceptance and false rejection while the former does not have to be. Hence, client-dependent score normalisation methods are considered here.

Examples of existing methods are Z-, D- (for Distance), T- (for Test) and more recently, F-Norms (for F-ratio). In the terms used in [3, 13], Z-Norm [13] is impostor-centric (i.e, normalisation is carried out with respect to the impostor distributions calculated "offline" by using additional data), T-Norm [13] is also impostor-centric (but with respect to a given utterance calculated "online" by using additional cohort impostor models). D-

Norm [14] is neither client- nor impostor-centric; it is specific to the Gaussian Mixture Model (GMM) architecture and is based on Kullback-Leibler distance between two GMM models. In [2], a client-centric version of Z-Norm was proposed. However, this technique requires as many as five client accesses. Due to user-friendliness aspect, one often does not have many client-specific biometric samples. To overcome this problem, F-Norm was proposed [12]. It is client-impostor centric. Based on the experiments reported, as few as two client scores are needed to perform this normalisation. It was shown that F-Norm is superior over Z-Norm because F-Norm uses the client-specific impostor information in addition to the client-specific information.

In this study, as an extension of [12], F-Norm is used. Suppose that the score of a system is y. It indicates how likely that a given biometric sample belongs to a client. Let $\mu^k(j)$ be the mean score of client with the unique identity j given that the true class-label $k = \{C, I\}$ (either a client or an impostor) is known (from a development set). Let the (class-dependent but) client-independent mean be μ^k, for $k = \{C, I\}$. The resultant F-ratio transformed normalisation is:

$$y^F = A(j)(y - B(j)), \tag{1}$$

where,

$$A(j) = \frac{2a}{\beta(\mu^C(j) - \mu^I(j)) + (1-\beta)(\mu^C - \mu^I)}, \tag{2}$$

and

$$B(j) = \gamma\mu^I(j) + (1-\gamma)\mu^I \tag{3}$$

The terms $A(j)$ and $B(j)$ are associated to client j (client-dependent) and are derived from F-ratio. They are each controlled by the parameters $\beta \in [0, 1]$ and $\gamma \in [0, 1]$ on a per fusion experiment basis. The term $2a$ determines the "desired" distance between the client-specific mean and the client-specific impostor mean. a is a constant and is fixed to 1. β and γ adjust between the client-dependent and client-independent information. When $\beta = 0$ and $\gamma = 0$, it can be shown mathematically that F-ratio normalisation is equivalent to no normalisation at all. In biometric authentication, one often has abundant client-specific (simulated) impostor information. Preliminary experiments in [12] show that $\gamma = 1$ is always optimal. The experimental results confirm that due to abundant client-specific impostor information, the shift in $B(j)$ can always be estimated reliably. As a consequence, the only parameter needs to be optimised, on a per experiment and per expert basis, is the β parameter. It can be optimised using different approaches, among which the direct approach is to use the line search procedure [15, Sec. 7.2].

3 Deriving Confidence Information

It has been shown in [16] that confidence can be derived from a "margin". The margin can be defined from False Acceptance (FA) Rate (FAR) and False Rejection (FR) Rate (FRR) with respect to a threshold Δ. FAR and FRR are defined as follows:

$$\text{FAR}(\Delta) = \frac{\text{number of FAs}(\Delta)}{\text{number of impostor accesses}}, \tag{4}$$

$$\text{FRR}(\Delta) = \frac{\text{number of FRs}(\Delta)}{\text{number of client accesses}} . \tag{5}$$

Replacing Δ by the associated expert score y, the margin of the score y is defined as:

$$q = |\text{FAR}(y) - \text{FRR}(y)| \tag{6}$$

Hence, when incorporated into an existing discriminant function, q modifies the discriminant function *dynamically*, i.e., *a per example basis*. Suppose that y_i is the score of expert $i = 1, \ldots, N$. Linear combination of $\{y_i, q_i y_i\}$ from different expert systems, with weight $w_{1,i}$ associated to y_i and $w_{2,i}$ associated to $q_i y_i$, is equivalent to computing $y_i \times (w_{1,i} + q_i w_{2,i})$, for all i [16]. Note that from the term $(w_{1,i} + q_i w_{2,i})$, it is obvious that q_i has a direct influence on the gradient of the resultant discriminative function on a *per example basis*. Hence, $\{y_i, q_i y_i\}$, can be seen as a form of Prior Knowledge Incorporation (PKI). Using equal weight in linear combination, in [16], it was shown that fusion with $\{q_i y_i | \forall_i\}$ has a better generalisation performance than fusion without the margin information (the classical way), i.e., $\{y_i | \forall_i\}$. Furthermore, fusion with $\{y_i, q_i y_i | \forall_i\}$ *consistently* outperforms $\{q_i y_i | \forall_i\}$, even though the generalisation performance is not always significant based on the HTER significance test [17].

4 Combing Both Sources of Information: A Prior Knowledge Incorporation (PKI) Framework

In the previous sections, the client-dependent and confidence information are employed on a per expert basis, independently of the other expert scores. The concept of PKI was introduced when discussing how confidence (based on margin) can be combined. In this section, we extend this concept to incorporate the client-dependent information as well, i.e., using $\{y_i, q_i y_i, y_i^F | \forall_i\}$. In principle, we could combine any other sources of information or prior knowledge this way. The only limit is the amount of prior knowledge captured by the available data (scores in this case).

Suppose that a linear combination is used to fuse $\{y_i, q_i y_i, y_i^F | \forall_i\}$. Let $w_{1,i}$, $w_{2,i}$ and $w_{3,i}$ be weights associated to y_i, $q_i y_i$ and y_i^F, respectively, for all i. Let the bias term be $-\Delta$, where Δ is the final decision threshold. Note that in this study, a separate training procedure of the Δ parameter is employed to minimise Weighted Error Rate (WER) *on the development set*. WER is defined as:

$$\text{WER}_\alpha(\Delta) = \alpha\text{FAR}(\Delta) + (1-\alpha)\text{FRR}(\Delta), \tag{7}$$

where $\alpha \in [0, 1]$ balances between FAR and FRR. This procedure requires the computation of fused scores on both the development and evaluation sets. In this way, during testing , based on a specified WER, the obtained threshold from the development set can be applied to the evaluation set. A separate threshold estimation procedure is necessary because algorithms that optimise the parameters of the fusion classifiers (weights in the linear combination case) *do not* necessarily optimise WER. For instance, SVM maximises the margin; Fisher discriminant maximises the Fisher-ratio criterion, etc.

The fused score can be written as:

$$
\begin{aligned}
y_{COM} &= \sum_i \left[y_i w_{1,i} + q_i y_i w_{2,i} + y_i^F w_{3,i} \right] - \Delta \\
&= \sum_i \left[y_i w_{1,i} + q_i y_i w_{2,i} + B(j)(y_i - A(j)) w_{3,i} \right] - \Delta \\
&= \sum_i \left[y_i \left(\underbrace{w_{1,i}} + \underbrace{q_i w_{2,i}} + \underbrace{B(j) w_{3,i}} \right) \right] - \sum_i \left[\underbrace{B(j) A(j) w_{3,i}} \right] - \underbrace{\Delta}, \qquad (8)
\end{aligned}
$$

where Eqn. (1) was used to replace the term y_i^F. The first underbraced term is the *global weight* on a *per expert basis*; the second is the weight contribution due to the confidence information on a *per example basis*; and the third is the weight contribution due to the client-dependent information source on a *per client basis*. These three weights are *linearly* combined to weight the score y_i. Then the fourth underbraced term introduces the client-dependent shift on a *per expert and per client basis*. Finally, the last underbraced term introduces the *global shift* to the final discriminative function. This term (Δ) is optimised by minimising WER for a given α value. From fusion point of view, the first three underbraced terms introduce tilt and while the last two underbraced term introduces shift to the decision hyperplane.

Although the PKI scores are simple to obtain, their linear combination can be a very complex function as shown here. It should be noted that even though non-linear combination can also be used (using the SVM algorithm with non-linear kernels , polynomial expansion of the terms $\{y_i, q_i y_i, y_i^F | \forall_i\}$, etc), simple linear solution is preferred to avoid overfitting. Furthermore, most of the non-linear part of the problem should have been solved by the base experts, thus eliminating the need for a complex second-level classifier.

5 Database and Evaluation

The publicly available[1] XM2VTS benchmark database for score-level fusion [11] is used. There are altogether 32 fusion data sets and each data set contains a fusion task of two experts. These fusion tasks contain multimodal and intramodal fusion based on face and speaker authentication tasks. For each data set, there are two sets of scores, from the *development* and the *evaluation* sets. The development set is used *uniquely* to train the fusion classifier parameters, including the threshold (bias) parameter, whereas the evaluation set is used uniquely to evaluate the generalisation performance. They are in accordance to the two originally defined Lausanne Protocols [18]. The 32 fusion experiments have 400 (client accesses) $\times$ 32 (data sets)= 12,800 client accesses and 111,800 (impostor accesses) $\times$ 32 (data sets) = 3,577,600 impostor accesses.

The most commonly used performance visualising tool in the literature is the Decision Error Trade-off (DET) curve [19]. It has been pointed out [20] that two DET curves resulting from two systems are not comparable because such comparison does not take into account how the thresholds are selected. It was argued [20] that such threshold should

[1] Accessible at http://www.idiap.ch/$\sim$norman/fusion

be chosen *a priori* as well, based on a given criterion. This is because when a biometric system is operational, the threshold parameter has to be fixed *a priori*. As a result, the Expected Performance Curve (EPC) [20] was proposed. This curve is constructed as follows: for various values of α in Eqn. (7) between 0 and 1, select the optimal threshold Δ on a development (training) set, apply it on the evaluation (test) set and compute the HTER on the evaluation set. This HTER is then plotted with respect to α. The EPC curve can be interpreted similarly to the DET curve, i.e., the lower the curve, the better the generalisation performance. In this study, the *pooled* version of EPC is used to visualise the performance. The idea is to plot a single EPC curve instead of 32 EPC curves for each of the 32 fusion experiments. This is done by calculating the *global* false acceptance and false rejection errors over the 32 experiments for *each* of the α values. The pooled EPC curve and its implementation can be found in [11].

6 Experimental Results

The client-*dependent* setting is used to derive F-Norm transformed scores. On the other hand, the client-*independent* setting is used to derive the margin scores. Three sets of control experiments are performed, namely with original scores $\{y_i|\forall_i\}$, F-Norm transformed scores $\{y_i^F|\forall_i\}$ and margin-derived confidence scores $\{y_i q_i|\forall_i\}$. For each set of experiments, three types of fusion classifiers are used, namely, a Gaussian Mixture Model (GMM), a Support Vector Machine (SVM) with a linear kernel and the mean operator. Both GMM and SVM employed are using standard algorithms, without any particular modification. The hyper-parameters are selected automatically via cross-validation. Figures 1(a)–(c) show the generalisation performance of these three sets of control experiments. Each curve is a pooled EPC curve over 32 fusion multimodal and intramodal datasets. Figure 2 complements Figure 1 by showing the corresponding ROC curves.

To compare these three control experiments with the ones fusing all sources of information, i.e., $\{y_i, y_i q_i, y_i^F|\forall_i\}$, we plotted the best of each pooled EPC curves in (a)–(c) on (d). As can be seen in (d), fusion with all sources of information using SVM has the best generalisation performance, bringing a new state-of-the-art overall performance on this benchmark data set. Considering significant performance improvement with respect to the 3×3 sets of control experiments, for large range of α values (> 0.6 for the best pooled EPC curve of the 9 control experiments over 32 fusion data sets), one can conclude that fusion using client dependent and confidence information sources via PKI is a feasible approach.

7 Conclusions

In this study, we proposed to fuse two seemingly different sources of information using the Prior Knowledge Incorporation (PKI) framework. These sources of information are client-dependent and confidence information. Although fusion with both sources of information has been studied separately in biometric authentication, to the best of our knowledge, fusing both information sources has not been well investigated before. Because these information sources are of different nature, intuitively, a *new* combination algorithm would be necessary. However, using the proposed PKI framework, we show that these information sources can be combined at the score level by a linear transformation, for each source of prior knowledge. The advantage is modularity: prior knowledge can be incorporated on a per expert basis (the first step) and the resultant PKI scores can be fused by a second-level classifier using standard machine learning algorihtms (the second step). Thus, this eliminates the need to devise specific fusion algorithms for this purpose. Based on the experiments carried out on 32 intramodal and multimodal fusion data sets taken from the publicly available XM2VTS benchmark database, over 10 fusion classifiers (3 fusion baselines on the original scores; 3 with client-dependent fusion baselines; 3 with margin-enhanced confidence baselines; and a final fusion with all information sources), fusion with both information sources using the PKI framework has the best generalisation performance and its performance is significant over large values of operating (false acceptance/false rejection) costs as compared to the most competing technique, i.e., fusion with client-dependent information.

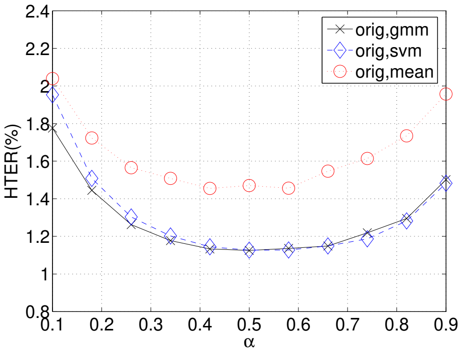
8 Acknowledgment

This work was supported in part by the IST Program of the European Community, under the PASCAL Network of Excellence, IST-2002-506778, funded in part by the Swiss Federal Office for Education and Science (OFES) and the Swiss NSF through the NCCR on IM2. This publication only reflects the authors' view.

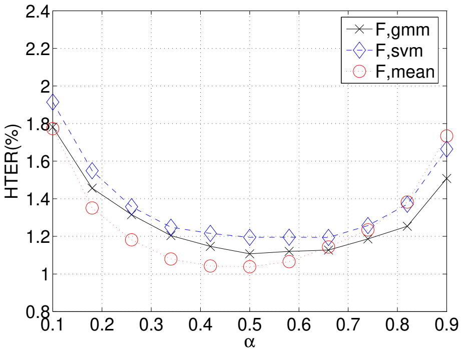
References

1. J. Kittler, G. Matas, K. Jonsson, and M. Sanchez, "Combining Evidence in Personal Identity Verification Systems," *Pattern Recognition Letters*, vol. 18, no. 9, pp. 845–852, 1997.
2. J.R. Saeta and J. Hernando, "On the Use of Score Pruning in Speaker Verification for Speaker Dependent Threshold Estimation," in *The Speaker and Language Recognition Workshop (Odyssey)*, Toledo, 2004, pp. 215–218.
3. J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target Dependent Score Normalisation Techniques and Their Application to Signature Verification," in *Springer LNCS-3072, Int'l Conf. on Biometric Authentication (ICBA)*, Hong Kong, 2004, pp. 498–504.
4. A. Jain and A. Ross, "Learning User-Specific Parameters in Multibiometric System," in *Proc. Int'l Conf. of Image Processing (ICIP 2002)*, New York, 2002, pp. 57–70.
5. A. Kumar and D. Zhang, "Integrating Palmprint with Face for User Authentication," in *Workshop on Multimodal User Authentication (MMUA 2003)*, Santa Barbara, 2003, pp. 107–112.
6. K-A. Toh, W-Y. Yau, E. Lim, L. Chen, and C-H. Ng., "Fusion of Auxiliary Information for Multimodal Biometric Authentication," in *Springer LNCS-3072, Int'l Conf. on Biometric Authentication (ICBA)*, Hong Kong, 2004, pp. 678–685.
7. J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Multimodal Biometric Authentication using Quality Signals in Mobile Communnications," in *12th Int'l Conf. on Image Analysis and Processing*, Mantova, 2003, pp. 2–11.
8. J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Kernel-Based Multimodal Biometric Verification Using Quality Signals," in *Defense and Security Symposium, Workshop on Biometric Technology for Human Identification, Proc. of SPIE*, 2004, vol. 5404, pp. 544–554.
9. D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia, "On the Use of Quality Measures for Text Independent Speaker Recognition," in *The Speaker and Language Recognition Workshop (Odyssey)*, Toledo, 2004, pp. 105–110.

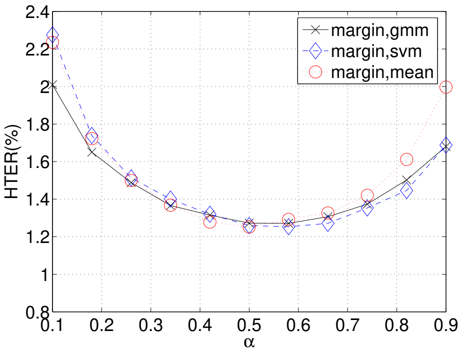
10. C. Sanderson and K. K. Paliwal, "Information Fusion and Person Verification using Speech and Face Information," IDIAP-RR 22, IDIAP, 2002.
11. N. Poh and S. Bengio, "Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication," Research Report 04-44, IDIAP, Martigny, Switzerland, 2004, Accepted for publication in *AVBPA 2005*.
12. N. Poh and S. Bengio, "Improving Single Modal and Multimodal Biometric Authentication Using F-ratio Client Dependent Normalisation," Research Report 04-52, IDIAP, Martigny, Switzerland, 2004.
13. R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score Normalization for Text-Independant Speaker Verification Systems," *Digital Signal Processing (DSP) Journal*, vol. 10, pp. 42–54, 2000.
14. M. Ben, R. Blouet, and F. Bimbot, "A Monte-Carlo Method For Score Normalization in Automatic Speaker Verification Using Kullback-Leibler Distances," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, 2002, vol. 1, pp. 689–692.
15. C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1999.
16. N. Poh and S. Bengio, "Improving Fusion with Margin-Derived Confidence in Biometric Authentication Tasks," Research Report 04-63, IDIAP, Martigny, Switzerland, 2004, Accepted for publication in *AVBPA 2005*.
17. S. Bengio and J. Mariéthoz, "A Statistical Significance Test for Person Authentication," in *The Speaker and Language Recognition Workshop (Odyssey)*, Toledo, 2004, pp. 237–244.
18. J. Matas, M. Hamouz, K. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Begun, N. Capdevielle, W. Gerstner, S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Comparison of Face Verification Results on the XM2VTS Database," in *Proc. 15th Int'l Conf. Pattern Recognition*, Barcelona, 2000, vol. 4, pp. 858–863.
19. A. Martin, G. Doddington, T. Kamm, M. Ordowsk, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance," in *Proc. Eurospeech'97*, Rhodes, 1997, pp. 1895–1898.
20. S. Bengio and J. Mariéthoz, "The Expected Performance Curve: a New Assessment Measure for Person Authentication," in *The Speaker and Language Recognition Workshop (Odyssey)*, Toledo, 2004, pp. 279–284.



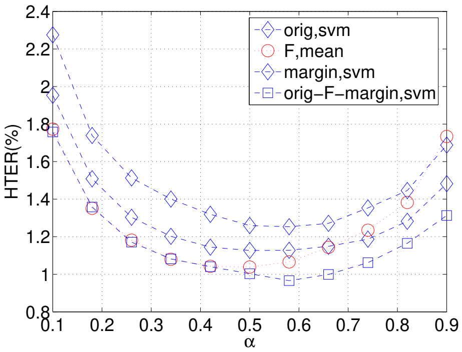
(a) original



(b) client dependent

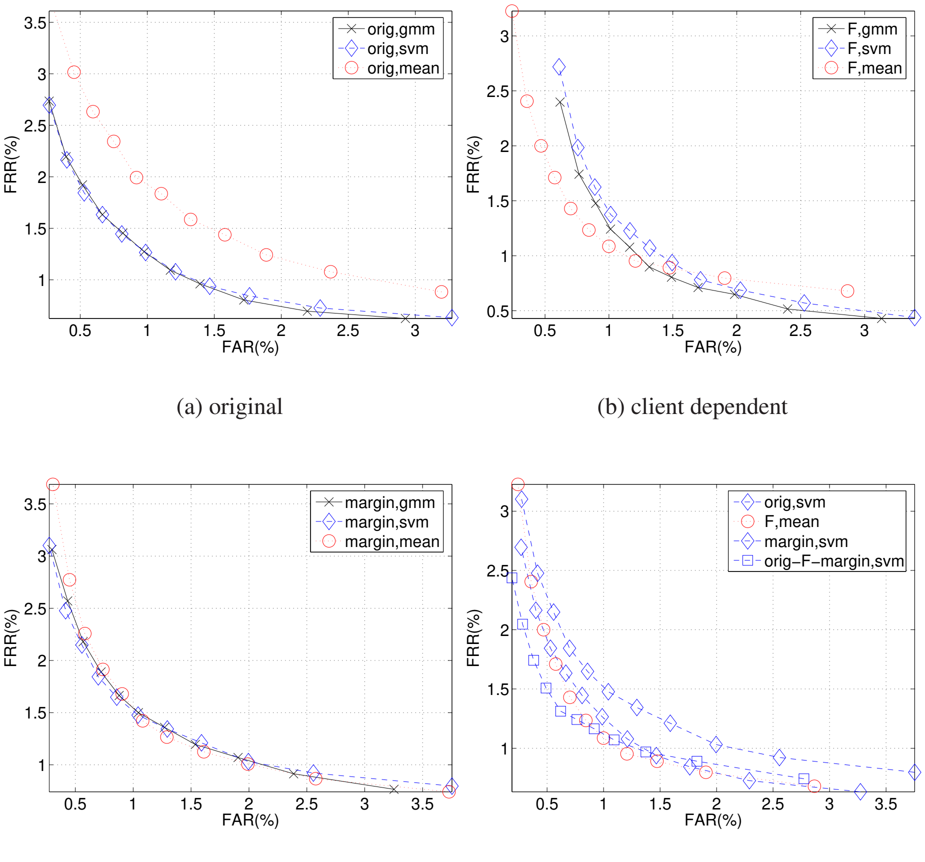


(c) confidence



(d) both information

Fig. 1. Pooled EPC curves from 32 XM2VTS benchmark fusion data sets of three baseline experiments (a)–(c) and fusion with all information sources (d). (a) is fusion with the original scores, $\{y_i|\forall_i\}$, (b) is fusion with F-ratio transformed scores, $\{y_i^F|\forall_i\}$,and (c) is fusion with margin-derived confidence, $\{y_i q_i|\forall_i\}$, each using a GMM, an SVM with linear kernel and the mean operator. The best three pooled EPC curves in (a)–(c) are plotted in (d) (the top three in the legend), together with fusion with all sources of information, i.e., $\{y_i, y_i q_i, y_i^F|\forall_i\}$ using an SVM with linear kernel, denoted as "orig-F-margin,SVM". The pooled EPC of this curve is compared to the "best overall fusion" (lowest HTER in the EPC curve across different α) in each of (a)–(c). "orig-F-margin,SVM" is better than "F-mean" for $\alpha > 0.6$ according to the HTER significance test at 90% of confidence. Below $\alpha > 0.6$, both EPC curves are not *significantly different*.



(a) original

(b) client dependent

(c) confidence

(d) both information

Fig. 2. Pooled ROC curves from 32 XM2VTS benchmark fusion data sets of three baseline experiments (a)–(c) and fusion with all information sources (d). (a) is fusion with the original scores, $\{y_i|\forall_i\}$, (b) is fusion with F-ratio transformed scores, $\{y_i^F|\forall_i\}$,and (c) is fusion with margin-derived confidence, $\{y_i q_i|\forall_i\}$, each using a GMM, an SVM with linear kernel and the mean operator. The "best" three pooled ROC curves (i.e., the EPC curve with the *lowest* HTER value across different α values) in (a)–(c) are plotted in (d), together with the one that fuses all sources of information, i.e., $\{y_i, y_i q_i, y_i^F|\forall_i\}$ using an SVM with linear kernel, denoted as "orig-F-margin,SVM". This figure complements Figure 1. As confirmed by the HTER significance test, for FRR above 1.2%, "orig-F-margin,SVM" is significantly different (and better) than "F-mean" but below 1.2%, their difference is *insignificant*. This phenomenon is due to few client accesses as compared to impostor accesses. As a result, low FRR values cannot be interpreted reliably compared to low FAR values.