IET Biometrics

Research Article

Score-level fusion using generalized extreme value distribution and DSmT, for multi-biometric systems

ISSN 2047-4938
Received on 4th June 2017
Revised 13th January 2018
Accepted on 23rd January 2018
E-First on 16th March 2018
doi: 10.1049/iet-bmt.2017.0076
www.ietdl.org

Renu Sharma[1,2], Sukhendu Das[3] ✉, Padmaja Joshi[1]
[1]*Centre for Development of Advanced Computing, Mumbai, India*
[2]*Computer Science and Engineering, Michigan State University, MI, USA*
[3]*Department of Computer Science & Engineering, Indian Institute of Technology Madras, Chennai, India*
✉ *E-mail: sdas@iitm.ac.in*

Abstract: Human recognition in a multi-biometric system is performed by combining biometric clues from different sources (multiple sensors, units, algorithms, samples and modalities) at different levels (sensor, feature, score, rank and decision level). Low computational complexity and adequate data for fusion make the score-level fusion a preferable option over other levels of fusion. However, incompatibility issue prevails at this level as scores obtained from different uni-biometric systems are disparate in nature. This disparity can be resolved by using score normalisation before fusion. This study first analysed the effect of generalised extreme value distribution-based score normalisation technique on different fusion techniques and then proposes an efficient score fusion technique based on Dezert–Smarandache theory (DSmT). A unique blend of belief assignment and decision-making methods in the DSmT framework is proposed for score-level fusion. For evaluation of the proposed method, experiments are performed on multi-algorithm, multi-unit and multi-modal biometrics systems created from three publicly available datasets: (i) NIST BSSR1 multi-biometric score database, (ii) face recognition grand challenge (FRGC) V2.0 and (iii) LG4000 iris dataset. Comparative studies of performance analysis show the efficiency of the authors' proposed method over other recently published state-of-the-art fusion methods.

1 Introduction

In biometric systems, physiological or behavioural traits are utilised for human identification. Uni-biometric system relies on a single source of biometric information. It faces challenges such as lack of uniqueness, non-universality, non-permanence, low fault tolerance, intra-class variations, low recognition rate, matcher limitation, noisy data, spoof attacks etc. These challenges can be alleviated to some extent by the use of the multi-biometric system for recognition. Multi-biometric system combines cues from different biometric sources for human recognition. On the basis of the source of information, multi-biometric systems can be classified into multi-sensor (multiple sensors used for same biometric modality), multi-unit (multiple units of the same biometric such as left and right eyes of a person), multi-algorithm (multiple algorithms for feature extraction or classification on the same biometric sample), multi-sample (multiple samples of the same biometric), multi-modal (multiple biometric modalities) and hybrid biometric systems. This paper deals with design and experimentations on multi-algorithm, multi-unit, multi-sample and multi-modal biometric systems.

Information fusion from different biometric clues is roughly performed at five levels: sensor, feature, score, rank and decision level. At the sensor and feature-level fusion, incompatibility and high computational complexity issues dominate. A rank and decision level, there is a loss of information required for efficient fusion. Score-level fusion provides a middle path between these two extremes. It provides sufficient information for fusion without increasing the computational complexity of the system. There exists a good trade-off between the information content and ease of use. The major drawback at the score-level fusion is the incompatibility between the scores obtained from the constituent uni-biometric systems. Some matchers give a similarity or dissimilarity score, while another gives a probability measure of the query sample to be genuine or an impostor. In addition, the distributions of scores of different matchers have a disparate location (mean), numerical scale (variance) and shape. Conflicting decisions and failure of the constituent systems further complicate the fusion of information at the score level.

To mitigate these issues of score-level fusion, we have proposed a unified framework consisting of generalised extreme value (GEV) distribution-based score normalisation and Dezert–Smarandache theory (DSmT)-based score-level fusion. Transforming the scores obtained from different uni-biometric systems using cumulative density function (CDF) of GEV distribution resolves the issue of different interpretation of matching scores. GEV distribution also helps in transforming matching scores into the common domain having similar mean, variance and shape. Its adequacy is proved in earlier published works [1–4]. Other normalisation techniques (other than extreme value theory) ignore the shape parameter while transforming the scores. We have applied the GEV distribution over the genuine scores only, assuming that the genuine scores form extreme/tail values with respect to the entire distribution. The DSmT framework is well known for fusion of uncertain, highly conflicting and imprecise sources of evidence. The framework consists of three steps: assignment of generalised belief function, combination rule and decision-making rule. We have proposed a unique blend of these three steps. Combining GEV-based normalisation and DSmT-based fusion technique as a unified scheme for score fusion along with a unique blend of three stages of DSmT framework forms a robust and efficient multi-biometric system. Extensive experiments on three publicly available databases (i) NIST BSSR1 [5] multi-modal biometric score database, (ii) FRGC V2.0 [6] face database and (iii) LG4000 [7] iris database shows the performance superiority of the proposed method as compared with the other state-of-the-art fusion algorithms.

1.1 Related work

Broadly, score-level fusion techniques can be categorised as: transformation-based, likelihood ratio-based, classifier-based and belief function theory-based techniques. In transformation-based fusion technique, matching scores are first transformed into a

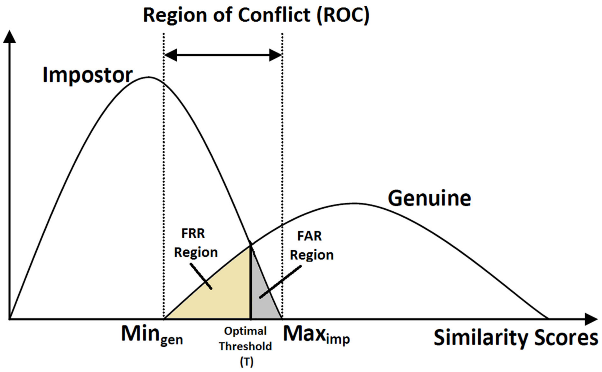


Fig. 1 *Synthetic illustration of the distributions of genuine and impostor scores*

common domain by min–max, *z*-score, double sigmoid or tanh [8] normalisation techniques. Then, the transformed scores are combined with a simple rule of weighted sum, min, max, product [9–12] etc. Eisenbach *et al.* [13] proposed likelihood ratio-based score normalisation and the normalised scores are fused using weighted sum rule, where weights are estimated by pairwise optimisation of projected genuine-impostor overlap Pairwise optimization of projected genuine-impostor overlap (PROPER) method. Various other weighting techniques are also proposed in the recent literature: query-adaptive feature weighting [14], score reliability-based weighting [15] and quality-based weights [16]. Major challenges in transformation-based techniques are the selection of normalisation techniques and its weight parameters which are training dependent and require extensive experiments. These techniques are simple and efficient in time complexity. However, the performance of multi-biometric systems decreases when the conflicting decision is given by the constituent classifiers.

Likelihood ratio or density-based fusion technique [17] requires explicit estimation of genuine and impostor distributions to calculate a likelihood ratio. Likelihood ratio achieves optimal performance at all operating points on receiver operating characteristics (ROCs), provided the estimation of probability density functions for genuine and impostor distributions are accurate. Parametric density estimation methods [18, 19] assume some predefined model for genuine and impostor scores distribution which is not always accurate, whereas non-parametric methods [20, 21] require a large number of training data for estimating the underlying distribution which is an expensive process.

In classifier-based techniques [22, 23], the problem of fusion of the matching scores is transformed into a binary classification (genuine or impostor) problem by concatenating different matching scores to form a feature vector. A fused vector of match scores is treated as a feature vector which is then classified into one of two classes: ‘genuine user’ or ‘impostor’. Various classifiers used for this purpose are support vector machine (SVM) [24, 25] and its variants, Fisher linear discriminant analysis, the Bayesian classifier (beta distribution), multi-layer perceptron and decision tree [25], hidden Markov model (HMM) [26] etc. Artabaz *et al.* [27] combined different primitive score-level fusion rules using a genetic algorithm. The process uses an optimised tree to determine the function structure. The limitation of classifier-based approaches is the unavailability of sufficient data or lack of representative data for training. Over-training and classifier bias [22] are another issue prominent in classifier-based fusion techniques.

Belief function theory recently introduced another category of methodologies in score-level fusion techniques. This class of techniques is quite favourable when biometric classifiers produce conflicting, uncertain and imprecise results. Combination of scores by Dempster–Shafer theory (DST) [28] was first introduced in [29]. Later, DST was widely used for combining various biometric modalities [30–32]. Mezai and Hachouf [33] proposed particle swarm optimisation technique for weighting the belief assignments in DST framework and proportional conflict redistribution (PCR5) rule for the combination. Nguyen *et al.* [34] incorporated an error factor in the calculation of belief assignment. DST-based score-level fusion techniques failed when conflicting scores are produced by the sources. Thereafter, an extension of the classical DST, called DSmT [35], was introduced for biometric score fusion by Singh *et al.* [36]. They proposed a robust face recognition system using an integrated framework of fusion at image level and score level of visible and infrared face images. Vatsa *et al.* [37, 38] and Abbas *et al.* [39] have also used DSmT-based fusion in various multi-biometric systems. Vatsa *et al.* [37] incorporated quality factor in the DSmT framework for fingerprint verification using level-2 and level-3 features. In another work, Vatsa *et al.* [38] proposed a dynamic reconciliation scheme for biometric fusion to handle various scenarios occurring in the probe dataset. Abbas *et al.* [39] proposed DSmT-based framework of fusion for handwritten signature verification. In [40, 41], subjective logic fusion operator is introduced in biometric fusion via belief fusion. Fakhar *et al.* [42] proposed score fusion based on two fuzzy sets (genuine and impostor) which are modelled using an automatic membership function generation method. Then, the new fuzzy sets are fused with a fuzzy aggregation operator to get a final score.

To summarise, score-level fusion appears to be more versatile and powerful among different combination schemes, as it provides a scope of relatively comparing the different score values from multiple cues rather than a simple weighted method of combinations. Among different score-level fusion techniques, belief function theory and DSmT appear as attractive alternative strategies, because of their power to deal with imprecise, uncertain and conflicting decisions, using specific probabilistic criteria.

In our work, we have focused on belief function-based score-level fusion technique. The work presented in this paper can be divided into two parts. The first part is an extension of our previous work [43] (GEV distribution-based score normalisation technique), where we have analysed the effect of GEV-based score normalisation on various score-level fusion techniques. Earlier works assumed that score normalisation is helpful only in transformation-based techniques, but in this work, we have demonstrated that our normalisation technique improves the accuracy of all categories of score-level fusion techniques. Although the work by He *et al.* [44] for reduction of high-scores effect normalisation and our earlier work (Sharma *et al.* [43]) for GEV-based normalisation techniques show that normalisation techniques improve the SVM and sum rule-based score-level fusion, but results were not shown using likelihood ratio-based and belief fusion-based methods. In the second part, we have proposed an efficient DSmT-based score fusion technique. Previous DSmT-based techniques differ from our proposed method in the way assignment of belief function and decision making are performed. Our method focused on the region of conflict by modelling it into two parts: False acceptance rate (FAR) and False rejection rate (FRR) (Fig. 1) at the time of belief assignment and redistributing it proportionally to the belief assignment of the impostor and genuine at the time of decision making, which makes it more efficient than other methods.

Various belief assignments and their combination rules have been developed during past several decades. Overview of belief function and the leading combination rules are discussed in Section 2. Algorithmic details of our proposed method are discussed in Section 3. Performance evaluation along with experimental results is provided in Section 4. Concluding remarks are summarised in Section 5.

2 Overview of belief function theory and its combination rules

Theory of belief functions or evidence is a general framework for dealing with uncertain, imprecise and conflicting data. It allows combining the evidence obtained from different experts and arriving at a degree of belief function for the final decision. The fusion algorithms in belief function theory consist of three steps: assignment of belief functions, the rule for combining these belief functions and the final decision based on the fused information. The three steps are briefly described below.

Let $\Theta = \{\theta_1, \theta_2\}$ be the frame of discernment consisting of a finite set of mutually exclusive and exhaustive propositions. Belief

Table 1 Various combination rules of belief function theory, where $\psi(\theta_j)$ is the characteristic non-emptiness function of a set θ_j such that $\psi(\theta_j) = 1$ if $\theta_j \notin \Phi$ and $\psi(\theta_j) = 0$; otherwise, where $\Phi \triangleq \{\Phi, \phi\}$, Φ is the set of all elements of D^Θ which are empty because of the constraints defined by particular application. I_t is the total ignorance and the union of all singletons and $\upsilon = \nu(X) \cup \nu(Y)$, $\nu(X)$ is the union of all singletons that comprise *X* and *Y*

Combination rule	Formula
mean combination rule	$m_{\text{mean}}(\theta_j) = \dfrac{m_1(\theta_j) + m_2(\theta_j)}{2}$
Dempster and Shafer rule [45]	$m_{\text{DST}}(\theta_j) = \dfrac{\Sigma_{(X,\, Y \in 2^\Theta),\, (X \cap Y = \theta_j)} m_1(X)m_2(Y)}{1 - \Sigma_{(X,\, Y \in 2^\Theta),\, (X \cap Y = \phi)} m_1(X)m_2(Y)}$
Smets' rule [49]	$m_{\text{con}}(\theta_j) = \Sigma_{(X,\, Y \in 2^\Theta),\, (X \cap Y = \theta_j)} m_1(X)m_2(Y)$
hybrid DSmT rule [35]	$m_{\text{DSmH}}(\theta_j) = \psi(\theta_j)[S_1(\theta_j) + S_2(\theta_j) + S_3(\theta_j)]$
	$S_1(\theta_j) = \Sigma_{X,\, Y \in D^\Theta,\, X \cap Y = \theta_j} m_1(X)m_2(Y)$
	$S_2(\theta_j) = \Sigma_{X,\, Y \in \Phi,\, [\upsilon = \theta_j] \vee [(\upsilon \in \Phi) \wedge (\theta_j = I_t)]} m_1(X)m_2(Y)$
	$S_3(\theta_j) = \Sigma_{X,\, Y \in D^\Theta,\, X \cup Y = \theta_j,\, X \cap Y \in \Phi} m_1(X)m_2(Y)$
PCR (PCR5) [35]	$m_{\text{PCR5}}(\theta_j) = m_{\text{con}}(\theta_j) + \Sigma_{Y \in 2^\Theta,\, \theta_j \cap Y = \phi} \left(\dfrac{m_1(\theta_j)^2 m_2(Y)}{m_1(\theta_j) + m_2(Y)} + \dfrac{m_2(\theta_j)^2 m_1(Y)}{m_2(\theta_j) + m_1(Y)} \right)$

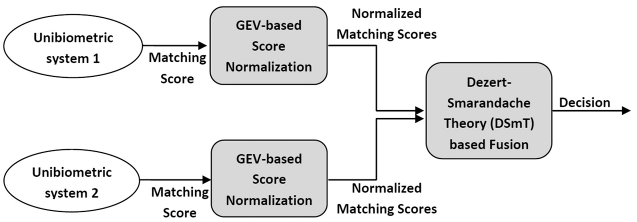


Fig. 2 *Architecture of the proposed method of multi-biometric fusion*

functions or basic belief assignments $m(.)$ map frame of discernment elements onto [0, 1]. In probability theory, belief functions are defined such that $m(\theta_1) + m(\theta_2) = 1$, whereas in a DST framework [28] belief functions which operate on the power set $2^\Theta = \{\phi, \theta_1, \theta_2, \theta_1 \cup \theta_2\}$, are defined as

$$\begin{gathered} m(\phi) = 0 \\ m(\phi) + m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) = 1. \end{gathered} \tag{1}$$

DSmT framework proposed by Dezert and Smarandache [35] operates on hyper-power set, $D^\Theta = \{\phi, \theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_1 \cap \theta_2\}$ and the generalised basic belief assignment (gbba) is defined, such that

$$\begin{gathered} m(\phi) = 0 \\ m(\phi) + m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) + m(\theta_1 \cap \theta_2) = 1. \end{gathered} \tag{2}$$

The first step involves defining the formulation of belief functions confirming the constraints.

In the second step, belief assignments of different sources are combined using specific rules. Dempster [45] proposed normalised conjunctive combination rule, which produces useful results with imprecise and uncertain data, but failed when conflicting scores are produced by the sources. Other drawbacks of DST framework are explained in the work presented in [46–48]. To circumvent the shortcomings of the DST framework, Smets' [49] proposed the transferable belief model which is a un-normalised conjunctive rule. Recently, Dezert and Smarandache [35] proposed a paradoxical and plausible reasoning-based fusion algorithm which is a generalisation of DST framework and Bayes theory. They have also proposed different versions of PCR rules (PCR1–PCR6) in the order of increasing computational complexity and efficiency. A brief list of combination rules based on belief function theory, considering two sources of combination (m_1 and m_2) are given in Table 1.

The third step in the belief function theory involves decision making from the combined basic belief assignments by transforming them into probability measures. In the DST framework, classical pignistic transformation is used to build a probability measure, as

$$\forall \theta_j \in 2^\Theta \quad \text{Bet}P\{\theta_j\} = \sum_{X \in 2^\Theta} \frac{|X \cap \theta_j| m_{\text{fus}}(X)}{|X|} \tag{3}$$

where $|\cdot|$ denotes the cardinality of the set and $m_{\text{fus}}(X)$ is the fused basic belief function. Usually, the maximum of pignistic probability measure is used for making a final decision. In the DSmT framework, generalised pignistic transformation is used for calculating probability measure, as

$$\forall \theta_j \in D^\Theta \quad \text{Bet}P\{\theta_j\} = \sum_{X \in D^\Theta} \frac{C(X \cap \theta_j) m_{\text{fus}}(X)}{C(X)} \tag{4}$$

where $C(X)$ is the DSm cardinal of proposition X. There exists another variant of probability measure called DSm probabilistic Dezert-Smarandache Probabilistic (DSmP) transformation, which is defined as

$$\text{DSmP}_\epsilon(\theta_j) = m_{\text{fus}}(\theta_j) + (m_{\text{fus}}(\theta_j) + \epsilon)\omega \tag{5}$$

where ω is the weighting factor defined as

$$\omega = \sum_{\substack{j \\ X_j \in D^\Theta \\ \theta_j \subset X_j \\ C(X_j) \geqslant 2}} \left(\frac{m_{\text{fus}}(X_j)}{\sum\limits_{\substack{k \\ X_k \in D^\Theta \\ X_k \subset X_j \\ C(X_k) = 1}} m_{\text{fus}}(X_k) + \epsilon C(X_j)} \right) \tag{6}$$

where $\epsilon \geq 0$ is the tuning parameter and $C(X_j)$ is the DSm cardinal of X_j. Likelihood ratio test can also be used for the final decision.

3 Description of the proposed method

The architecture of the proposed method for score-level fusion is given in Fig. 2. Our flowchart diagram of Fig. 2 is a special form of analytic hierarchy process [50, 51]. For a multi-biometric system, matching scores obtained from different uni-biometric systems are first normalised using GEV a distribution-based score normalisation technique [43]. A brief description of GEV-based score normalisation is given in Section 3.1. Post-normalisation, the DSmT-based fusion technique has been used, whose detailed description is given in Section 3.2.

3.1 Score normalisation technique

Disparate nature of the scores obtained from the different uni-biometric systems obligate the introduction of normalisation step in the score-level fusion process. In our earlier work [43], we have proposed the normalisation technique based on EV theory, exploiting the GEV distribution for the tail or EVs. GEV distribution is applied over the genuine data, assuming that the genuine scores are extreme/tail values with respect to the entire set (genuine and impostor scores). The parameter set (location, scale and shape) is computed from the training genuine data using maximum-likelihood estimation method. The CDF of GEV distribution is given as

$$G(x, \mu, \sigma, k) = \begin{cases} \exp(-\left(1 + k\left(\frac{x-\mu}{\sigma}\right)^{-(1/k)}\right) & \text{if} \quad k \neq 0 \\ \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right) & \text{if} \quad k = 0 \end{cases} \tag{7}$$

such that $1 + k(x-\mu)/\sigma > 0$ and $1 + kx > 0$. μ, σ and k are the mean, scale and shape parameters of the distribution. Given a GEV distribution, normalised score is estimated as a probability measure, considering that a given score is an outlier of the distribution. This is computed from the value of CDF of the distribution, as follows:

$$S_i^N = G(S_i, \mu, \sigma, k) \tag{8}$$

where S_i and S_i^N are the matching scores obtained from the ith classifier and its corresponding normalised score.

3.2 Score-level fusion technique

The DSmT can formally combine any independent sources of information represented in the form of belief functions. DSmT has the capacity to solve complex, static or dynamic fusion problems as opposed to the DST framework, especially when a conflict between the sources is high or refinement of the frame (Θ) becomes inaccessible because of vague, relative and imprecise nature of elements in it.

In a biometric system, the frame of discernment used is $\Theta = \{\theta_{\text{gen}}, \theta_{\text{imp}}\}$. DSmT framework operates on the hyper-power set of the frame of discernment which is defined as $D^\Theta = \{\phi, \theta_{\text{gen}}, \theta_{\text{imp}}, \theta_{\text{gen}} \cup \theta_{\text{imp}}, \theta_{\text{gen}} \cap \theta_{\text{imp}}\}$. The proposed method of fusion based on DSmT involves the following three steps: (i) estimation of gbba, (ii) combination of gbba through a DSmT-based combination rule and (iii) decision for accepting or rejecting a person. The following three sections describe these steps.

3.2.1 Estimation of gbba: Here, gbba is a mapping function $m(.)$ defined as: $D^\Theta \to [0, 1]$, such that

$$\sum_{\theta_j \in D^\Theta} m(\theta_j) = 1 \tag{9}$$

For transforming the normalised matching scores into gbba, we have used the method described in [30]. According to the method, if a matching score is greater than the maximum of impostor scores (Max_{imp}), it belongs to the genuine class (if matchers scores are similarity scores), as shown in Fig. 1. In addition, if the matching score is less than the minimum of genuine scores (Min_{gen}), it belongs to the impostor class. There is an uncertainty of the matching score lying in the region of conflict $[\text{Min}_{\text{gen}}, \text{Max}_{\text{imp}}]$. Using these observations, matching scores are mapped to gbba corresponding to all the elements of D^Θ, as follows:

$$\begin{array}{ll}
\text{if}(S_i^N > \text{Max}_{\text{imp}}) & : m_i(\theta_{\text{gen}}) = \Psi(S_i^N) \\
 & \quad m_i(\theta_{\text{gen}} \cup \theta_{\text{imp}}) = 1 - \Psi(S_i^N) \\
\text{if}(S_i^N < \text{Min}_{\text{gen}}) & : m_i(\theta_{\text{imp}}) = 1 - \Psi(S_i^N) \\
 & \quad m_i(\theta_{\text{gen}} \cup \theta_{\text{imp}}) = \Psi(S_i^N) \\
\text{if}(S_i^N \in [\text{Min}_{\text{gen}}, \text{Max}_{\text{imp}}]) & \\
\quad \text{if}(S_i^N > T_i) & : m_i(\theta_{\text{gen}} \cup \theta_{\text{imp}}) = \Psi(S_i^N) \\
 & \quad m_i(\theta_{\text{gen}}) = 1 - \Psi(S_i^N) \\
\quad \text{if}(S_i^N < T_i) & : m_i(\theta_{\text{gen}} \cup \theta_{\text{imp}}) = 1 - \Psi(S_i^N) \\
 & \quad m_i(\theta_{\text{imp}}) = \Psi(S_i^N) \\
m_i(\phi) = 0 & \\
m_i(\theta_{\text{gen}} \cap \theta_{\text{imp}}) = 0 &
\end{array}$$

where S_i^N and T_i are the normalised score and optimal threshold of the ith classifier, ϕ is the empty set, $m_i(\theta_{\text{gen}})$ is a gbba corresponding to the θ_{gen} element and $\Psi(S_i^N)$ is a monotonically non-decreasing function which maps a matching score in the range of [0,1]. We have used the sigmoid function in our experiments. $m_i(\theta_{\text{gen}} \cup \theta_{\text{imp}})$ represents belief function for a region of conflict and $m_i(\theta_{\text{gen}} \cap \theta_{\text{imp}}) = 0$ is a constraint applied to biometric systems as a score cannot be genuine and impostor both at the same time. The works in [36, 37] define gbba using probability density function of a Gaussian distribution, whereas Vatsa *et al.* [38] estimated gbba using a multivariate Gaussian distribution over genuine, impostor and region of conflict scores separately. Abbas *et al.* [39] utilised an extended version of Appriou's model for belief assignment. Earlier score fusion methods based on DSmT focus on the region of conflict as a single entity and model it using one distribution, whereas in our proposed method we have focused on two parts (FAR and FRR regions as shown in Fig. 1) of the region of conflict separately and define gbba accordingly. This helps in the efficient modelling of the region of conflict which is a main source of error in biometric systems. We have applied GEV-based normalisation only on genuine scores as opposed to the earlier normalisation techniques based on EV distribution, as it is generally the genuine scores which forms extreme/rare values with respect to the whole distribution.

3.2.2 Combination of gbbas: Several rules of combination [35] are defined in the DSmT framework: classical DSm rule, hybrid DSm rule of combination, Smets' rule, Dubois and Prade's rule, Yager's rule and PCR rule (PCR 1–6). We have used PCR5 for combining belief assignments of different uni-biometric systems. The PCR5 formula for two uni-biometric systems is given as (see (10)) where $m_i(\theta_j)$ is the belief assignment (gbba) of the ith classifier (uni-biometric system) and the jth element of D^Θ, while $m_\cap(\theta_j)$ corresponds to the conjunctive consensus of θ_j between the two classifiers. A PCR6 rule is a generalised form of PCR5 rule (only applicable for two element frame of discernment) for the frame of discernment having more than two elements. PCR combination rules transfer (total or partial) conflicting masses to non-empty sets involved in the conflicts, proportionally with respect to the masses assigned to them by sources. Hybrid DSm rule of combination is used in [36, 37], whereas PCR5 rule is used in [38, 39] for the combination.

$$\begin{aligned} m_{\text{fus}}(\phi) &= 0 \\ m_{\text{fus}}(\theta_j) &= m_\cap(\theta_j) + \sum_{\theta_j \in D^\theta \setminus \{\theta_j\} \theta_i \cap \theta_j = \phi} \left[\frac{m_1(\theta_i)^2 m_2(\theta_j)}{m_1(\theta_i) + m_2(\theta_j)} + \frac{m_2(\theta_i)^2 m_1(\theta_j)}{m_2(\theta_i) + m_1(\theta_j)} \right] \end{aligned} \tag{10}$$

Table 2 VRs with various score-level fusion techniques on the multi-algorithm system, using NIST BSSR1 dataset

Techniques	VR	VR with GEV	Techniques	VR	VR with GEV
face G	93.7226	94.3380	one-class SVM	88.6667	96.7368
face C	94.7071	95.7947	RBF SVM	97.0429	99.0563
sum rule	93.8149	96.8910	Dempster–Shafer	71.6236	98.8873
likelihood ratio	95.8039	97.0030	Vatsa *et al.* [38]	96.8944	98.5723
HMM	99.9675	96.5950	our method	99.9390	99.9710

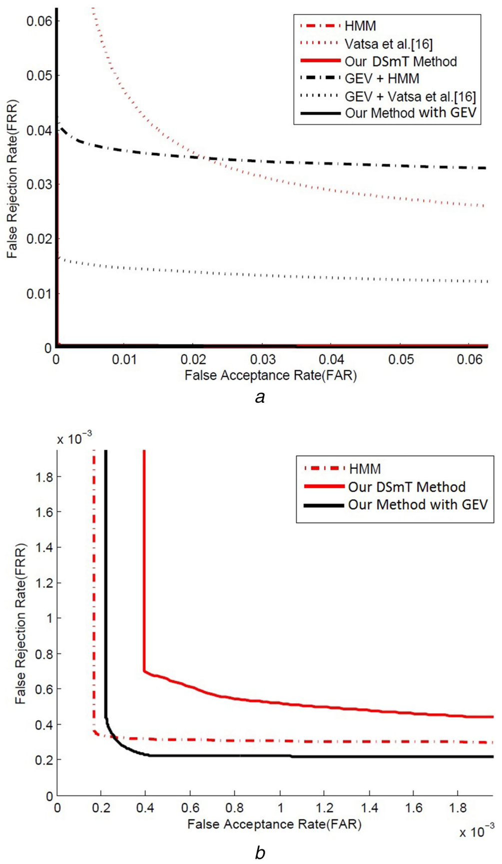


Fig. 3 *ROC curves for multi-algorithm biometric system*

(a) ROC curves of various methods on the multi-algorithm biometric system (NIST BSSR1 dataset). The best six performing ones are only shown along with our proposed approach, ***(b)*** Zoomed in version of Fig. 3*a*, for very low values (up to 10^{-4}) of FRR and FAR

3.2.3 Decision for verification: For decision making, the $m_{\text{fus}}(\theta_{\text{gen}} \cup \theta_{\text{imp}})$ score is distributed between $m_{\text{fus}}(\theta_{\text{gen}})$ and $m_{\text{fus}}(\theta_{\text{imp}})$, proportionally based on their weights. Weights are assigned according to their distance from the optimal threshold. Confidence scores for genuine and impostor are estimated using fused gbba as follows:

$$
\begin{aligned}
d(\theta_{\text{gen}}) &= m_{\text{fus}}(\theta_{\text{gen}}) + \frac{\text{Max}_{\text{imp}} - T}{\text{Max}_{\text{imp}} - \text{Min}_{\text{gen}}} m_{\text{fus}}(\theta_{\text{gen}} \cup \theta_{\text{imp}}) \\
d(\theta_{\text{imp}}) &= m_{\text{fus}}(\theta_{\text{imp}}) + \frac{T - \text{Min}_{\text{gen}}}{\text{Max}_{\text{imp}} - \text{Min}_{\text{gen}}} m_{\text{fus}}(\theta_{\text{gen}} \cup \theta_{\text{imp}})
\end{aligned} \tag{11}
$$

where T is the optimal threshold and $d(.)$ is the confidence score that the query belongs to either genuine and impostor. Min_{gen}, Max_{imp} and T are calculated from the training data. The query sample belongs to the one having maximum confidence score. Singh *et al.* [36] directly apply estimated threshold for decision making on gbba. Vatsa *et al.* in [37] applied pignistic probability transform and likelihood ratio on gbba in [38] for final decisions. Abbas *et al.* [39] utilised DSm probabilistic (DSmP) transform and likelihood ratio for making a decision. In all previous works based on DSmT, probability/gbba measures of genuine and impostor are utilised and region of conflict was ignored. We have utilised gbba over a region of conflict ($m_{\text{fus}}(\theta_{\text{gen}} \cup \theta_{\text{imp}})$) by distributing it proportionally to the gbba of the impostor and genuine, as in (11).

The overall novelty of the proposed approach lies in combining GEV-based normalisation and DSmT-based fusion technique as a unified scheme for score fusion and a unique blend of three stages of the DSmT framework. Most earlier published work had either worked on GEV distribution-based normalisation or DSmT-based fusion, providing the limited power of score-level fusion in the multi-biometric system. There is a limitation of the DSmT framework which also applies to our proposed method. The framework works efficiently only in verification mode, whereas it is intractable in identification mode as the cardinality of hyper-power set (see Section 3.2) increases rapidly. There is also no consensus in decision making in a DSmT framework which keeps this area open for research.

4 Experiments and analysis

To analyse the performance of GEV-based score normalisation technique and its combination with DSmT-based score fusion technique, we have performed experiments on multi-algorithm, multi-unit and multi-modal biometric systems. Three publicly available databases were used for experiments: (i) NIST BSSR1 [5] multi-modal biometric score database, (ii) FRGC V2.0 [6] face database and (iii) LG4000 [7] iris database. Experiments were repeated ten times (ten folds) and an average of the performance values are shown as results. All the experiments were performed on a core-*i*5, 2.3 GHz, 8 GB random access memory machine. For comparative study of performance, following score-level fusion techniques are used: sum rule [44], likelihood ratio-based method [19], one-class SVM [52], Radial basis function (RBF) kernel SVM [53], HMM [26], Dempster–Shafer [30] and the method proposed in Vatsa *et al.* [38].

4.1 Multi-algorithm biometric system

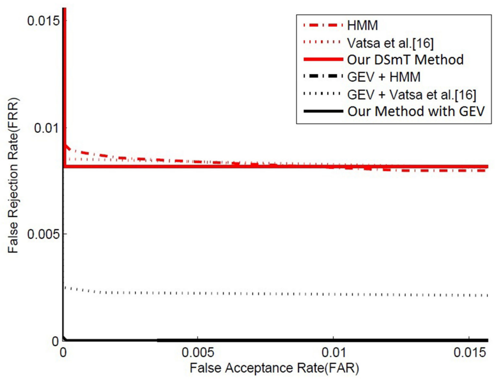
In NIST BSSR1 [5] dataset, for two face algorithms (labelled as C and G), similarity score vectors of 6000 samples for 3000 subjects were given. We have used half of the scores for the estimation of training parameters and rest for the testing. Verification rate (VR) of the various methods is given in Table 2 and the ROC curves of only six high-performance techniques (HMM [26], Vatsa *et al.* [38] and the proposed method) with and without GEV normalisation are shown in Fig. 3*a*. Zoomed in version of Fig. 3*a* for extremely low values of FAR and FRR, is shown in Fig. 3*b*, to exhibit the marginally better performance of the proposed method over that of HMM [26].

4.2 Multi-unit biometric system

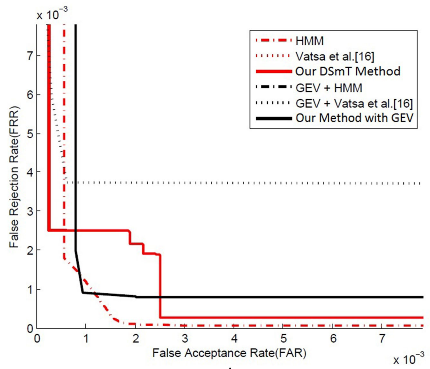
In NIST BSSR1 [5] dataset, for multiple units of the fingerprint (labelled as Ri and Li), similarity score vectors of 6000 samples for 6000 subjects were given. We have used half of the scores for the estimation of training parameters and rest for the testing. VR of the various methods are given in Table 3 and the ROC curves of only six high-performance techniques are shown in Fig. 4*a*. From Table 3 and Fig. 4*a*, it can be concluded that our proposed method outperforms other methods in the multi-unit biometric system, and the second-best performing method is the work published in Vatsa *et al.* [38].

Table 3 VRs with various score-level fusion techniques on the multi-unit biometric system, using NIST BSSR1 dataset

Techniques	VR	VR with GEV	Techniques	VR	VR with GEV
fingerprint Li	92.4096	93.7071	one-class SVM	96.6419	98.3000
fingerprint Ri	94.8335	95.9711	RBF SVM	97.8167	99.5190
sum rule	97.1876	98.2553	Dempster–Shafer	97.9944	99.3978
likelihood ratio	97.1901	98.2500	Vatsa *et al.* [38]	99.1691	99.7764
HMM	99.1797	98.3089	our method	99.1863	99.9971



a



b

Fig. 4 *ROC curves for multi-unit biometric system*
(a) ROC curves of various methods on the multi-unit biometric system (NIST BSSR1 dataset). The best six performing ones are only shown along with our proposed approach, ***(b)*** ROC curves for best six performing multi-modal biometric system with face (FRGC v2.0) and iris (LG 4000) datasets

Table 4 VRs with various score-level fusion techniques, for face and fingerprint multi-modal biometric system, using NIST BSSR1 dataset

Techniques	VR	VR with GEV	Techniques	VR	VR with GEV
face G	93.9117	94.5235	one-class SVM	97.6798	99.0667
fingerprint Li	92.2216	93.5976	RBF SVM	97.8608	99.8378
sum rule	97.3844	98.9875	Dempster–Shafer	88.1064	98.9798
likelihood ratio	98.4200	99.0000	Vatsa *et al.* [38]	96.7712	99.7279
HMM	99.9846	93.6057	our method	99.5007	99.9865

4.3 Multi-modal biometric system

Here, for evaluation in a multi-modal biometric scenario, three modalities were used in two combinations: (i) combination of face and fingerprint biometric traits from NIST BSSR1 [5] and (ii) face and iris modalities from FRGC v2.0 [6] and LG 4000 [7] datasets, respectively. From NIST BSSR1 [5] dataset, similarity scores of 3000 samples (face algorithm G and left index fingerprint Li) for 3000 subjects are taken we have used half of the scores for the estimation of training parameters and rest for the testing. VR of various methods are shown in Table 4 and the ROC curves are shown in Fig. 5*a*. Our method performs the best as can be concluded from Table 4. Zoomed in version of Fig. 5*a* for extremely low values of FAR and FRR is shown in Fig. 5*b* as the pair of plots run very close to the coordinate axis (as near ideal performance). The second-best performing method is the RBF kernel SVM [53]. ROC curves (Fig. 5*b*) show similar performances of our method and that too using HMM [26].

In another experiment, 19,079 images of the face and 19,079 images of the left iris for 461 subjects were used from FRGC v2.0 face and LG 4000 iris datasets, respectively. The entire dataset was divided into three sets: training set, evaluation (validation) set and test set. Five images of the face (and left iris) per subject were selected for the training set; five images of the face (and left iris) per subject constitute the evaluation set, and the remaining $(28,938 = 19,079 \times 2 - 461 \times 10 \times 2)$ images of face and iris were included in the test set. Images were selected at random and do not overlap across sets. All iris images were resized to 640×480 and face images were cropped to 131×111 px^2. Gabor filters (eight orientations and five scales) were used for feature extraction from face images and a one-dimensional log-Gabor feature extraction technique was used for the iris images. SVM (RBF kernel) [54] and probabilistic neural network [54] classifiers were used for the generation of scores in the face and iris biometric systems, respectively. VR of various methods are given in Table 5 and the ROC curves are shown in Fig. 4*b*. Again, it can be deduced that our method works efficiently for a multi-modal system as shown in Table 5. RBF kernel-based SVM [53] method performs second best. Our method and HMM [26] shown similar performance in ROC curves (Fig. 4*b*).

Computational time required for the fusion of the matching scores obtained from two experts is shown in Fig. 6. From all four cases of experiments (Tables 2–5), one can conclude that GEV-based score normalisation technique boosts the performance of all categories of score-level fusion: transformation-based, likelihood ratio-based, classifier-based and belief function-based fusion techniques. Hence, our proposed approach needs negligibly more computational time (hence not heavy in cost) than other methods. There is an exception in the case of HMM-based [26] fusion technique as shown in Figs. 3*b* and 5*b*. However, the performance of our proposed method which is a combination of GEV-based normalisation and DSmT-based fusion outperforms all other fusion methods.

5 Conclusion

We have proposed a unified scheme for score-level fusion which is a combination of GEV distribution-based normalisation and DSmT-based fusion technique. We have also analysed the performance of transformation-based, classifier-based, likelihood ratio-based and belief function-based score fusion techniques after applying the GEV-based score normalisation, and one can conclude that GEV-based score normalisation technique not only performs well as compared with other normalisation techniques in given sum and SVM rule-based methods as shown in [43], but also enhances the performance of other score-level fusion techniques. A unique blend of belief assignment and decision-making methods in DSmT framework results in an efficient DSmT-based score fusion technique. This outperforms GEV as (i) GEV-based normalisation has now been applied only to genuine scores; (ii) region of conflict is modelled as two separate entities – using FAR and FRR; and (iii) distribution of the relevance of each such part proportionally to the belief assignment of impostor and genuine scores. Extensive experiments on multi-algorithm, multi-unit and multi-modal biometric systems show that combination of GEV-based score normalisation and DSmT-based score-level fusion outperforms many other recent state-of-the-art fusion techniques.

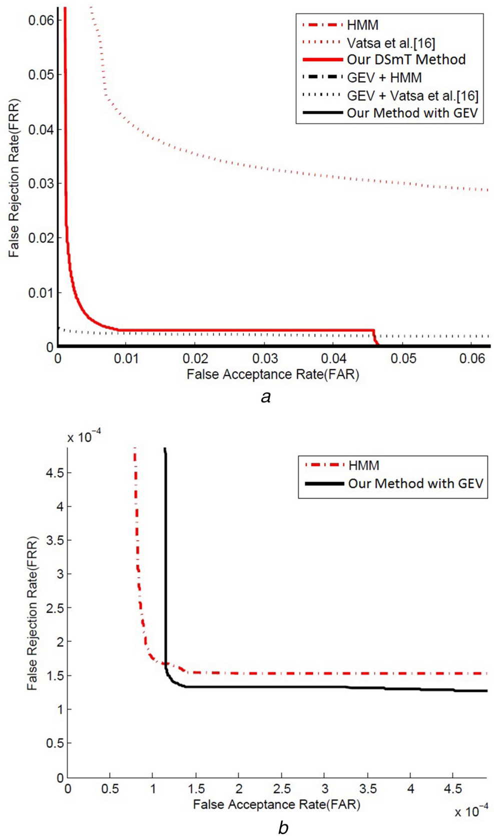


Fig. 5 *ROC curves for multi-modal biometric system*
(a) ROC curves of various methods on the multi-modal biometric system (NIST BSSR1 dataset) consisting of the face and fingerprint modalities. The best six performing ones are shown along with our proposed approach, ***(b)*** Zoomed in version of Fig. 5*a*, for very low values of FAR

Table 5 VRs with various score-level fusion techniques on the multi-modal biometric system, using face (FRGC v2.0) and iris (LG 4000) datasets

Techniques	VR	VR with GEV	Techniques	VR	VR with GEV
face	93.5337	95.9110	one-class SVM	98.5713	99.5363
iris	97.3742	98.0540	RBF SVM	99.7016	99.8173
sum rule	98.8016	99.5991	Dempster–Shafer	99.7831	99.7837
likelihood ratio	98.6012	99.5628	Vatsa *et al.* [38]	99.1706	99.6271
HMM	99.8927	95.8439	our method	99.7831	99.9062

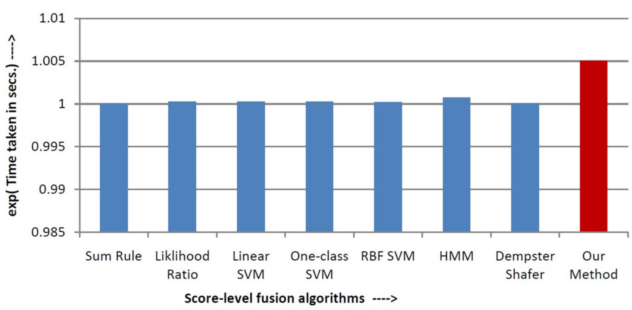


Fig. 6 *Average time taken (in seconds) by various score-level fusion techniques, for multi-algorithm biometric system (NIST BSSR1)*

6 References

[1] Scheirer, W., Rocha, A., Micheals, R., *et al.*: 'Robust fusion: extreme value theory for recognition score normalization'. European Conf. Computer Vision, 2010, pp. 481–495

[2] Li, Z., Vinyals, O., Baker, H., *et al.*: 'Feature learning using generalized extreme value distribution based k-means clustering'. Int. Conf. Pattern Recognition, 2012, pp. 1538–1541

[3] Fragoso, V., Sen, P., Rodriguez, S., *et al.*: 'EVSAC: accelerating hypotheses generation by modeling matching scores with extreme value theory'. IEEE Int. Conf. Computer Vision, 2013, pp. 2472–2479

[4] Gibert, X., Patel, V.M., Chellappa, R.: 'Sequential score adaptation with extreme value theory for robust railway track inspection'. IEEE Int. Conf. Computer Vision Workshops, 2015, pp. 131–138

[5] NIST: 'Biometric scores set (BSSR1)'. Available at https://www.nist.gov/itl/iad/image-group/nist-biometric-scores-set-bssr1, 2004

[6] Phillips, P.J., Flynn, P.J., Scruggs, T., *et al.*: 'Overview of the face recognition grand challenge'. Computer Vision and Pattern Recognition, 2005, vol. **1**, pp. 947–954

[7] LG 4000 iris image database. Available at http://biometrics.idealtest.org/, 2002

[8] Ross, A., Jain, A., Nandakumar, K.: 'Score normalization in multimodal biometric systems', *Pattern Recognit.*, 2005, **38**, (12), pp. 2270–2285

[9] Snelick, R., Uludag, U., Mink, A., *et al.*: 'Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (3), pp. 450–455

[10] Ross, A., Jain, A.: 'Information fusion in biometrics', *Pattern Recognit. Lett.*, 2003, **24**, (13), pp. 2115–2125

[11] Connaughton, R., Bowyer, K.W., Flynn, P.J.: '*Fusion of face and iris biometrics*' (Handbook of Iris Recognition, Springer London, 2016), pp. 397–415

[12] Chaa, M., Boukezzoula, N.-E., Attia, A.: 'Score-level fusion of two-dimensional and three-dimensional palmprint for personal recognition systems', *J. Electron. Imaging*, 2017, **26**, p. 13018

[13] Eisenbach, M., Kolarow, A., Vorndran, A.: 'Evaluation of multi feature fusion at score-level for appearance-based person re-identification'. Int. Joint Conf. Neural Networks, 2015, pp. 469–476

[14] Zheng, L., Wang, S., Tian, L., *et al.*: 'Query-adaptive late fusion for image search and person re-identification'. IEEE Conf. Computer Vision and Pattern Recognition, June 2015

[15] Kabir, W., Ahmad, M.O., Swamy, M.N.S.: 'Score reliability based weighting technique for score-level fusion in multi-biometric systems'. IEEE Winter Conf. Applications of Computer Vision, 2016, pp. 1–7

[16] Artabaz, S., Sliman, L., Dellys, H.N., *et al.*: 'Multibiometrics enhancement using quality measurement in score level fusion', in Janusz, K. (Ed.): '*Advances in intelligent systems and computing*', vol. **557** (Springer, 2017), pp. 1287–1297

[17] Dass, S.C., Nandakumar, K., Jain, A.K.: 'A principled approach to score level fusion in multimodal biometric systems'. Audio and Video-Based Biometric Person Authentication, 2005 (LNCS, **3546**), pp. 1049–1058

[18] Marval, P.L.R., Aoyama, S., Ito, K., *et al.*: 'Score level fusion of multibiometrics using local phase array'. Advances in Multimedia Information Processing, 2015, vol. **9315**

[19] Nandakumar, K., Chen, Y., Dass, S.C., *et al.*: 'Likelihood ratio-based biometric score fusion', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (2), pp. 342–347

[20] Mukherjee, S., Pal, K., Majumder, B.P., *et al.*: Differential evolution based score level fusion for multi-modal biometric systems, Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014, pp. 38–44

[21] Liang, Y., Ding, X., Liu, C., *et al.*: 'Combining multiple biometric traits with an order-preserving score fusion algorithm', *Neurocomputing*, 2016, **171**, pp. 252–261

[22] Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: 'Fusion of face and speech data for person identity verification', *IEEE Trans. Neural Netw.*, 1999, **10**, (5), pp. 1065–1074

[23] Ross, A.A., Nandakumar, K., Jain, A.K.: '*Handbook of multibiometrics (international series on biometrics*' (Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2006)

[24] Castrillón-Santana, M., Lorenzo-Navarro, J., Ramón-Balmaseda, E.: 'Multiscale score level fusion of local descriptors for gender classification in the wild', *Multimedia Tools Appl.*., 2017, **76**, (4), pp. 4695–4711

[25] Safavi, S., Mporas, I.: 'Improving performance of speaker identification systems using score level fusion of two modes of operation'. International Conference on Speech and Computer, 2017 (LNCS, **10458**), pp. 438–444

[26] Zheng, Y., Blasch, E.: 'An exploration of the impacts of three factors in multimodal biometric score fusion: score modality, recognition method, and fusion process', *J. Adv. Inf. Fusion*, 2014, **9**, (2), pp. 124–137

[27] Artabaz, S., Sliman, L., Benatchba, K., *et al.*: 'Score level fusion scheme in hybrid multibiometric system'. Advances in Visual Informatics: Int. Visual Informatics Conf., 2015, pp. 166–177

[28] Shafer, G.: '*A mathematical theory of evidence*' (Princeton University Press, Princeton, 1976)

[29] Al-Ani, A., Mohamed, D.: 'A new technique for combining multiple classifiers using the Dempster–Shafer theory of evidence', *J. Artif. Intell. Res.*, 2002, **17**, pp. 333–361

[30] Mezai, L., Hachouf, F., Bengherabi, M.: 'Score fusion of face and voice using Dempster–Shafer theory for person authentication'. Intelligent Systems Design and Applications (ISDA), 2011, pp. 894–899

[31] Singh, R., Vatsa, A., Noore, M., *et al.*: 'Dempster–Shafer theory based classifier fusion for improved fingerprint verification performance', *Comput. Vis., Graph. Image Process.*, 2006, **4338**, pp. 941–949

[32] Kisku, D.R., Sing, J.K., Gupta, P.: 'Multibiometrics belief fusion'. Int. Conf. Machine Vision, 2009, pp. 37–40

[33] Mezai, L., Hachouf, F.: 'Score-level fusion of face and voice using particle swarm optimization and belief functions', *IEEE Trans. Hum.-Mach. Syst.*, 2015, **45**, (6), pp. 761–772

[34] Nguyen, K., Denman, S., Sridharan, S., *et al.*: 'Score-level multibiometric fusion based on Dempster–Shafer theory incorporating uncertainty factors', *IEEE Trans. Hum.-Mach. Syst.*, 2015, **45**, (1), pp. 132–140
[35] Smarandache, F., Dezert, J.: '*Advances and applications of DSmT for information fusion*', vol. 3 (American Research Press, United States, 2009), pp. 760
[36] Singh, R., Vatsa, M., Noore, A.: 'Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition', *Pattern Recognit.*, 2008, **41**, (3), pp. 880–893
[37] Vatsa, M., Singh, R., Noore, A.: 'Unification of evidence-theoretic fusion algorithms: a case study in level-2 and level-3 fingerprint features', *IEEE Trans. Syst., Man Cybern. A, Syst. Hum.*, 2009, **39**, (1), pp. 47–56
[38] Vatsa, M., Singh, R., Noore, A., *et al.*: 'On the dynamic selection of biometric fusion algorithms', *IEEE Trans. Inf. Forensics Sec.*, 2010, **5**, (3), pp. 470–479
[39] Abbas, N., Chibani, Y., Hadjadji, B.: 'A DSmT based combination systems for handwritten signature verification', in Smarandache, F., Dezert, J. (Eds.): '*Advances and applications of DSmT for information fusion. Collected works*', (American Research Press, United States, 2015), vol. **4**, pp. 423–440
[40] Jasang, A., Munch-Maller, T.H.: 'Biometric data fusion based on subjective logic'. Int. Conf. Information Fusion, 2014, pp. 1–8
[41] Wasnik, P., Schafer, K., Ramachandra, R., *et al.*: 'Fusing biometric scores using subjective logic for gait recognition on smartphone'. Int. Conf. Biometrics Special Interest Group, 2017, pp. 1–5
[42] Fakhar, K., Aroussi, M.E., Saidi, M.N., *et al.*: 'Fuzzy pattern recognition-based approach to biometric score fusion problem', *Fuzzy Sets Syst.*, 2016, **305**, pp. 149–159
[43] Sharma, R., Das, S., Joshi, P.: 'Score normalization in multimodal systems using generalized extreme value distribution'. British Machine Vision Conf. (BMVC), 2015, vol. **1**, pp. 131.1–131.12
[44] He, M., Horng, S.-J., Fan, P., *et al.*: 'Performance evaluation of score level fusion in multimodal biometric systems', *Pattern Recognit.*, 2010, **4**, (5), pp. 1789–1800
[45] Dempster, A.P.: 'A generalization of Bayesian inference' Journal of the Royal Statistical Society, 1968, vol. **30**, pp. 73–104
[46] Zadeh, L.A.: 'A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination', *AI Mag.*, 1986, **7**, (2), pp. 85–90
[47] Dubois, D., Prade, H.: 'On the unicity of Dempster rule of combination', *Int. J. Intell. Syst.*, 1986, **1**, (2), pp. 133–142
[48] Voorbraak, F.: 'On the justification of Dempster's rule of combination', *Artif. Intell.*, 1991, **48**, (2), pp. 171–197
[49] Smets, P.: 'The combination of evidence in the transferable belief model', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990, **12**, (5), pp. 447–458
[50] Saaty, R.W.: 'The analytic hierarchy process – what it is and how it is used', *Math. Model.*, 1987, **9**, (3), pp. 161–176
[51] Brunelli, M.: 'Introduction to the analytic hierarchy process', in (Eds.): '*Springer briefs in operations research*' (Springer International Publishing, United States, 2015), p. 83
[52] Bergamini, C., Oliveira, L.S., Koerich, A.L., *et al.*: 'Combining different biometric traits with one-class classification', *Signal Process.*, 2009, **89**, (11), pp. 2117–2127
[53] Gutschoven, B., Verlinde, P.: 'Multi-modal identity verification using support vector machines (SVM)', Information Fusion, 2000, vol. **2**, pp. 3–8
[54] Duda, R.O., Hart, P.E., Stork, D.G.: '*Pattern classification*' (Wiley, United States, 2000, 2nd edn.)