ARTICLE IN PRESS
JID: PATREC [m5G;April 25, 2018;7:27]
Pattern Recognition Letters 000 (2018) 1–10
Contents lists available at ScienceDirect
Pattern Recognition Letters
journal homepage: www.elsevier.com/locate/patrec

Score-level fusion for cancelable multi-biometric verification

Rudresh Dwivedi*, Somnath Dey

Discipline of Computer Science and Engineering, Indian Institute of Technology Indore, Indore 453552, India

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Biometric
Multibiometric system
Fusion
Score-level fusion

A B S T R A C T

Integration of scores from multiple biometric modalities has become promising to alleviate the limitations of unibiometric systems such as sensitivity to outliers, erroneous authentication caused by inter-class and intra-class variability and low verification performance due to poor quality. In this work, we propose a two-level score level fusion approach for integrating the scores obtained from cancelable templates of different biometric modalities. As a result, we achieve a significant improvement in overall recognition performance providing secure authentication for the different application. At the first level, scores from multiple matchers are combined using a novel mean-closure weighting (MCW) technique to achieve the desired score for a particular biometric modality. The proposed solution is based on the region of uncertainty between the genuine and imposter distribution. Further, the derived scores from different modalities are integrated using a novel rectangular area weighting (RAW) technique at the second level to obtain the overall fused score. Overall, the proposed two-level cancelable score fusion method improves the performance over unimodal cancelable systems and are more robust to the variability of scores and outliers. The evaluation has been performed on two virtual databases and is compared with the existing weighting, density and classification based score fusion techniques. Experimental results show that the proposed two-level cancelable score fusion improves the overall performance over unibiometric system satisfying the requirement of secure authentication.

1. Introduction

1.1. Background

Biometric-based authentication systems based on a single source of information (unimodal biometric systems) suffer from limitations such as the lack of uniqueness and non-universality of the chosen biometric trait, noisy data, and spoof attacks [31]. The idea of information fusion, which combines the information or evidence presented by multiple sources, has been effectively used to overcome the limitations of unimodal biometric systems and to improve the recognition performance of unimodal biometric systems [31,35]. The use of multiple pieces of evidence to verify a user's identity is often referred to as multibiometrics. To achieve the goal of claimed performance enhancement, fusion rule must be selected based on the type of application, biometric traits, and level of fusion [19]. Fusion can be performed at five different levels of information namely: sensor, feature, decision, rank and match-score levels [31]. Among all of the five fusion levels [31], fusion at the sensor, feature, and decision levels have been extensively studied in the literature [37]. Biometric systems that perform fusion at an early stage of processing are believed to be more effective than those systems which perform fusion at a later stage [23]. Sensor-level fusion combines the data acquired from multiple sensors and derives a new data from which features can be extracted. Sensor-level fusion addresses the problem of noisy data, but all other limitations associated with unimodal biometric systems remain intact. Since the features contain significant information about the input biometric data than the match score or the decision of a matcher, fusion at the feature level provide better recognition performance than other levels of fusion. Feature level fusion computes the feature vector from the same or different input image. Next, feature vectors are combined (concatenated) to form a new feature set. However, fusion at the feature level is difficult to achieve in practice as the relationship between the feature spaces of different biometric systems are unknown, the concatenation of feature vectors may lead to a very large dimensionality, and the inaccessibility of the feature vectors of most commercial biometric systems. Rank-level fusion assigns multiple ranks to an identity and determines a new rank to make a final decision. Fusion at rank-level has not drawn much attention since it is only applicable for user's identification [30]. In decision level fusion, resulting feature vectors are individually classified into two classes: accept or reject. Next, majority voting or other fusion rules are applied to draw the

* Corresponding author.
E-mail address: phd1301201006@iiti.ac.in (R. Dwivedi).

https://doi.org/10.1016/j.patrec.2018.04.022
0167-8655/

Please cite this article as: R. Dwivedi, S. Dey, Score-level fusion for cancelable multi-biometric verification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.04.022

final decision. Fusion at the decision level is too strict since only a limited amount of information or decisions are available. In score level fusion, multiple classifiers output a set of match scores which are fused to derive a single scalar score. Next, this score is utilized to verify the user. Among all levels, fusion at the score level is the most promising Since integrating information calculated from individual trait's score seems both practical and feasible as it allows more freedom to apply the best-suited feature extraction and matching algorithms for different biometric traits. Further, it offers the best trade-off between information and ease of fusion [30].

1.2. Existing approaches

There are different schemes for performing score level fusion based on different techniques. These include transformation based, density based and classifier based fusion schemes [31].

1.2.1. Transformation based approaches

In these approaches, first match scores are normalized to a common domain (e.g. in the interval [0,1]). Next, the fused score is derived by weighting different match scores and this combined scores is utilized for authentication. The approaches proposed in [12,14,16,23,25,27,28,33,35], have utilized the transformation based approaches for match-score level fusion. Toh et al. [35] utilized weighted-sum rule-based fusion onto hand geometry, fingerprint and voice biometric for verification. The four learning and decision scenarios have been investigated. The experiments carried out shows of about 50% improvement over equal error rate (EER). Snelick et al. [33] proposed an adaptive normalization procedure which is utilized to increase the separation between genuine and imposter distributions in the range [0,1]. This procedure is carried out using three functions: Two-Quadrics (QQ), Logistic, and Quadric-Line-Quadric. Next, different fusion techniques such as simple-sum, min-score, max-score, matcher weighting, and user weighting are applied for fusion. Jain et al. [14] analyzed the performance of existing normalization techniques and fusion methods for face, fingerprint and hand geometry-based multimodal biometric system. Further, they reported that tanh normalization followed by the simple sum of score fusion method gives better performance in comparison to other normalization techniques [15]. Lobrano et al. [16] introduced a new index named Score Decidability Index which computes the coefficients of the linear combination for each classifier. Next, mean-rule fusion method is utilized for fusion. Poh et al. [27] proposed a novel normalization procedure which first, categorizes the user scores based on joint density computation. Next, F-norm is applied to estimate group-specific mean by maximum-a-posteriori (MAP) computation. Hanmandlu et al. [12] first, normalized the scores in the interval of [0,1] using fuzzy triangular membership function. Further, product rule is applied for score-level fusion. Poh et al. [28] applied a Gaussian Mixture Model-based fusion classifier. Next, B-ratio for each possible combination is evaluated and OR-switcher is utilized for fusion. Peng et al. [25] proposed constructed a virtual database of multiple finger biometric sources named: finger vein, fingerprint, finger-shape, and finger-knuckle. Before fusion, matching scores of the four biometric features are normalized into the range [0,1]. Then, t-norm based score-level fusion is applied for authentication. Nanni et al. [23] proposed a novel normalization method which computes the ith matcher's performance p_i ($p_i = 1 - EER_i$) and means of the matchers performance ($P_\mu = \sum_i \frac{(1-EER_i)}{N}$). Next, the weight of each matcher is quantized by its recognition performance. Further, different fusion algorithms are applied and compared with the existing state-of-the-art approaches. Wang et al. [38] applied Aczél-Alsina (AA) t-norm [38] to fuse score information for dual iris and face biometric modalities expanding the interval between genuine and imposter distribution. Transformation-based methods do not involve any training except the appropriate normalization and computation of combination weights. Further, they require exhaustive empirical evaluation.

1.2.2. Classification based approaches

In these approaches, scores obtained from different matchers for the same subject are concatenated to form a feature vector. Next, a classifier is utilized to differentiate between genuine and imposter users. The approaches proposed in [17,32,36], have utilized the Classification based approaches for match-score level fusion. Ma et al. [17] proposed an improved classifier using a tree model in a random forest. Each tree in a forest represents a classification axiom. The tree grows until the terminal node with the decision is reached based on the splitting rule. Dempster-Shafer (DS) theory is another classical classification technique widely studied and utilized in the literature [10,24]. DS theory combines evidence by computing the basic belief assignment (BBA) of each individual's score. Singh et al. [32] applied DS theory onto four different classifier outputs for fingerprint: Minutiae-based verification, ridge-based verification, fingercode based verification and pore-based verification. For each fingerprint image, each of the four classifier assigns a label true i.e. 1 to proposition i if $i \in T$ and remaining classes are assigned a label false i.e. 0. Next, the respective BBA are computed and combined using DS rule for verification. Tronci et al. [36] proposed a method which dynamically selects the matching score to provide a better separation between genuine and imposter. The selectors provide a better separation in terms of area under ROC curve. In classifier-based fusion, an incorrect decision may occur in scenarios where sufficient training samples are absent as the genuine match scores supplied to the classifier for training is $O(N)$ but impostor scores are of $O(N^2)$, where N is the number of subjects. Further, the cost of false acceptance may differ from the cost of false rejection and the selection of an optimal classifier for a given data set is a challenging task.

1.2.3. Density based approaches

In these approaches, scores from multiple matchers are concatenated to form a feature vector. Next, match-score densities are evaluated to form a set of training score. Most of the existing approaches [9,22] utilize Kernel density estimator (KDE) or a mixture of Gaussians (MOG). The approaches proposed in [5,21,22], have employed the Classification based approaches for match-score level fusion. Nandakumar et al. [21] proposed a Gaussian mixture model for genuine and imposter density estimation. Next, likelihood ratio fusion rule is applied to match scores and estimated densities. Further, Tao et al. [34] introduced a method to compute likelihood ratio directly by ROC curves of individual classifiers involved in the fusion. Nanni et al. [22] utilized mixture of Gaussian [9] estimation using expectation-maximization (EM) algorithm. Next, support vector machine (SVM) and AdaBoost of neural network classifier are applied on the different biometric training sets and sum rule is used for fusion. Dass et al. [5] proposed a modified kernel density estimator in which the marginal density is computed as a mixture of continuous and discrete components. Next, the joint density is estimated using copula functions. Density estimation is categorized into parametric or non-parametric methods. In parametric methods, the density function is known and the parameters of the function are estimated from the training data. Non-parametric methods represent score distribution and estimate probability density function (PDF) by kernel density estimation (KDE). In parametric density fusion methods, the assumption of incorrect models for both genuine and impostor scores may lead to deficient fusion rules. In case of nonparametric methods, an insufficient number of training samples (i.e. genuine scores) affects the design of an effective fusion rule.

Please cite this article as: R. Dwivedi, S. Dey, Score-level fusion for cancelable multi-biometric verification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.04.022

1.3. Motivations and contributions

As described above, an insufficient number of training samples and complex density estimation are the two critical issues in classification and density based approaches, respectively. Among the three categories described above, we adopt the transformation based fusion to mitigate the limitations of the previous score level fusion approaches in order to improve the performance in addition to overall accuracy. Traditional score fusion methodologies have not considered the scores from cancelable biometric template yet. Here, we aim to create a multimodal biometric system when the scores from cancelable biometric templates are fused to verify the user's identity. Since cancelable biometric based authentication systems lead to performance degradation in comparison to original biometric authentication systems, the prime motivation becomes to perform score level fusion for cancelable biometric modalities thereby reducing performance degradation. Also, the literature suffices one of the limitations of unimodal biometric systems are the intraclass variations. To alleviate this, the multi-modal biometric systems integrating high-performance biometric modalities as iris and fingerprint can be utilized to obtain a zero EER [1]. Earlier, we have proposed the techniques to derive cancelable iris and cancelable fingerprint template. Therefore, we integrate the scores obtained from multiple matchers applied onto cancelable iris and cancelable fingerprint template in this work. However, the proposed fusion framework could be extended to other biometric traits also.

To the best of our knowledge, our method is the first to incorporate the two-level fusion utilizing novel mean closure weighting (MCW) and rectangular area weighting (RAW) method to estimate weights for protected modalities. In the proposed two-level fusion scheme, first we apply mean-closure weighting (MCW) method considering complete set of scores (including genuine and imposter scores) for each matcher. However, The existing weighting methods (such as equal error rate weighting (EERW) [33], *d*-prime weighting (DPW) [33], mean-to-extrema weighting (MEW) [4] and Fisher's discriminant ratio weighting (FDRW) [26]) for score level fusion only involves the scores outside the uncertainty region to estimate the weight which results the performance sensitive to outliers. Furthermore, EER cannot be considered as weight estimation factor since a matcher with a lower EER may have higher FMR than the other matchers involved in fusion. MCW decides whether the user's score is closer to genuine or imposter distribution for any matcher. This user-specific score weighting scheme performs better for the scores lying in the region of uncertainty by assigning more weight to the scores lying into the confidence region and vice-versa. Next, we apply a novel rectangular area weighting (RAW) method to evaluate the fused scores, where the weights are estimated based on the rectangular area containing the region of uncertainty for the individual modalities. The modality containing less rectangular area is assigned more weight since it has less uncertainty region and vice-versa. Further, the computation of rectangular area requires less computation as compared to area under ROC curve (AUC) or area approximation methods. Finally, the performance of the proposed method is better than that of the existing weighting, density and classification based methods.

In a nutshell, we highlight the contribution remarks of our work as follows:

1. We perform two-level fusion onto cancelable (protected) scores utilizing mean-closure and rectangular area based weighting method to reduce the performance degradation due to cancelable transformation and to ensure diversity and revocability for the multibiometric template.
2. In this work, normalization is not required as the scores evaluated from the different matchers are in the interval of [0,1]. Score level fusion is performed directly onto raw protected scores which contain richer information than the normalized score.
3. The weighting techniques (i.e. MCW and RAW) incur the minimal computational complexity without any learning and fulfill both requirements of performance improvement and privacy preservation.
4. Experimental evaluation onto two different virtual datasets is carried out to explore the potential robustness of the proposed method for multimodal biometrics. Also, we have compared our method with state-of-the-art transformation based, density based and classification based approaches. The experimental results conclude that our method outperforms over the existing density based, classification based and existing weighting approaches.
5. We have computed the performance in two scenarios; (i) performance with fused scores obtained from cancelable templates (ii) performance with fused scores obtained from unprotected (original) templates. The experimental evaluation confirms a minor performance degradation with respect to unprotected multibiometric system. Also, we achieve a significant relative performance improvement over the unimodal cancelable biometric systems.

The block diagram of the proposed fusion framework is illustrated in Fig. 1. The remainder of this paper is organized as follows. Section 2 briefly describes the existing methods utilized to compute match scores. The proposed methodology for score level fusion is described in Section 3. Section 4 demonstrates and analyzes experimental results as well as compares the proposed method with some existing score-level fusion approaches. Finally, Section 5 summarizes our findings and concludes the paper.

2. Match score computation

In this section, we present the distance/similarity metrics utilized to compute match scores for cancelable iris and fingerprint biometric. The computation of match scores involves the comparison between cancelable enrolled and cancelable query template. The scores computed from individual biometric systems represent either dissimilarity (distance) or similarity measure. Therefore, it is needed to convert all the scores alike in nature. In this work, we transform all the scores into similarity measure following the common practice.

2.1. Cancelable iris match scores computation

The comparison between cancelable enrolled and cancelable query template is performed to evaluate match scores for iris biometric. To derive cancelable iris template, we apply the similar methodology proposed in our earlier work [8]. For convenience, we briefly describe this method. First, iris images are pre-processed using Masek's [18] and Daugman's [6] techniques. Then, IrisCode features are extracted in the form of a 0–1 matrix using 1-D Log-Gabor filter [18] with phase quantization from the pre-processed iris images. Thereafter, rotation-invariant IrisCode is generated from the original IrisCodes and the rotation invariant IrisCode is transformed into a row vector. Next, the decimal vector is derived by partitioned the row vector into fixed size blocks. Finally, a Look-up table is created to map the decimal-encoded vector and to generate the cancelable template. Finally, matching between protected enrolled and protected query iris template is performed in the transformed domain to measure the match score. First, we compute the similarity in Hamming domain for its simplicity in the evaluation. Hamming distance (*HD*) is the sum of non-equivalent bits (exclusive-OR) between stored and query

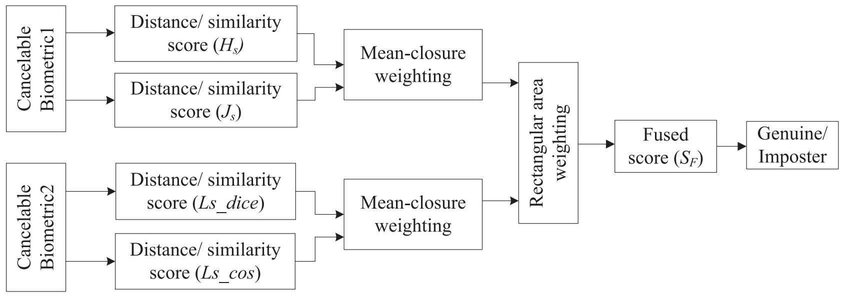

Fig. 1. Block diagram of the proposed fusion framework.

templates. The Hamming similarity is computed by subtracting normalized Hamming distance by one, as defined in Eq. (1):

$$\text{Hamming similarity}(H_s) = 1 - \frac{1}{N}\sum_{i}^{N} E_i \oplus Q_i \tag{1}$$

where, Q_i and E_i are the ith bits of the query and enrolled templates, respectively. N is the total number of bits in the template.

Next, Jaccard similarity is evaluated between protected stored and protected query iris template. Jaccard similarity is the overlap of bits in E and Q except the ill condition i.e.0-0 overlap as defined in Eq. (2). Jaccard similarity is computed to elude the ill match condition (0-0 match) between the protected query and protected enrolled templates.

$$\text{Jaccard similarity } (J_s) = \frac{N_{11}}{N_{01} + N_{10} + N_{11}} \tag{2}$$

where,

N_{11}: Number of positions where E, Q both have a value of 1,

N_{01}: Number of positions where value in E is 0 and Value in Q is 1,

N_{10}: Number of positions where value in E is 1 and Value in Q is 0.

2.2. Cancelable fingerprint match scores computation

Cancelable enrolled and cancelable query fingerprint templates are compared to calculate match scores for fingerprint biometric. To derive cancelable fingerprint template, we apply the method as proposed in our earlier work [7]. For reader's clarity, we describe the method briefly. First, the input fingerprint image is preprocessed to obtain the thinned image and to extract the minutiae information. Next, we form a nearest-neighbor structure around each minutiae point using the ridge-based co-ordinate system and compute the ridge features from the thinned image and minutiae information. Thereafter, we apply cantor pairing function to encode the ridge features uniquely. Finally, the random projection is applied to the paired output to derive the protected template. In the verification stage, the same procedure is followed to generate the protected template from the query fingerprint.

The matching between enrolled and query templates is performed in the transformed domain to maintain security. We adopt the inner product based similarity measures since similarity computation requires measuring the likelihood between the rows in the protected enrolled template (E) to the rows in protected query templates (Q). First, we utilize Dice coefficient to measure the local similarity (Ls_dice) between each row of enrolled and that of query templates as utilized in [39] as defined in Eq. (2).

$$Ls\_dice\,(i, j) = \frac{2E(i,:) \cdot Q(j,:)}{||E(i,:)||^2 + ||Q(j,:)||^2} \tag{3}$$

Further, we apply cosine similarity ($Ls\_\cos$) between each row of the enrolled and that of query templates to compute normalized dot product as defied here:

$$Ls\_cos\,(i, j) = \frac{E(i,:) \cdot Q(j,:)}{\sqrt{||E(i,:)||^2}\sqrt{||Q(j,:)||^2}} \tag{4}$$

Next, we re-evaluate each element in local similarity matrix to avoid double matching. For this purpose, we acquire those positions where the maximum scores in $E(i, :)$, and $Q(j, :)$ coincides to obtain filtered similarity matrix. Next, the global similarity score is obtained by summing up the entries in filtered matrix and dividing by a minimum of the minutiae points in E and Q. Finally, the likelihood of the enrolled and query template being the two fingerprint of the same subject is measured to compute global similarity scores.

3. Proposed score level fusion

After, evaluation of match scores from cancelable modalities, there is a need for score normalization such that match scores are transformed into a common interval (e.g. in the interval of [0,1]). In this work, normalization is not needed since the methods utilized in score computation generate the scores in the interval of [0,1]. However, the proposed work can be extended to the situations where the scores from different biometric modalities follow different distribution range or scores derived through different matchers may have a different range instead of [0,1]. In these situations, we utilize the RHE normalization [13] to guarantee a meaningful score integration since it is found to be sensitive to outliers. RHE minimizes the score-sets to be normalized because of the fact that raw scores have richer information content than the normalized score. In the following, we have proposed a novel mean-closure (MC) weighting mechanism followed by rectangular area (RA) weighting method for optimal weight estimation. The proposed model achieves the optimal weights for different matchers corresponding to each of the regions present in the FMR/FNMR curve including the uncertainty region. Further, the fused score is utilized for multimodal verification. In this work, we evaluate the scores from protected iris and protected fingerprint biometric to explore the potential significance of cancelable multimodal biometrics with respect to security, privacy and performance improvement over the unimodal biometric system.

3.1. Mean-closure (MC) based weighting

Let us consider, Hamming similarity and Jaccard similarity measure to be matcher 1 and matcher 2, respectively where the scores from two matchers (s_h, s_j) are to be integrated. On the basis of Fig. 2, we indicate five possible regions for the different scores of any matcher. The gray color regions (i.e. R_1 and R_2) represents the region of confidence where both the regions are able to classify the scores accurately. Region 3 (R_3), region 4 (R_4) and region 5 (R_5) falls into the uncertainty where it is very difficult to classify the match score. Therefore, it is necessary

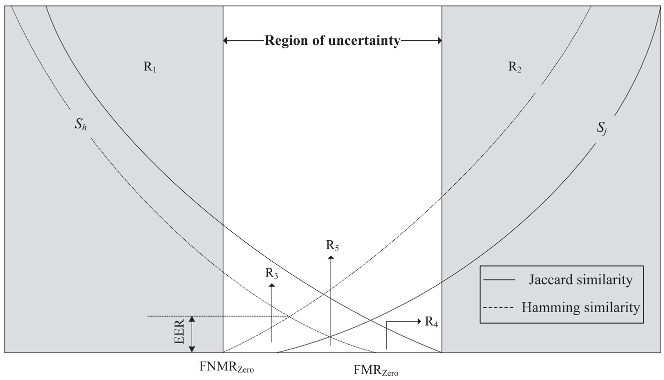

Fig. 2. Explanatory diagram Region of uncertainty present in FMR/FNMR curve; FMR_{Zero}, $FNMR_{Zero}$ and EER correspond to matcher 1 i.e. Hamming similarity.

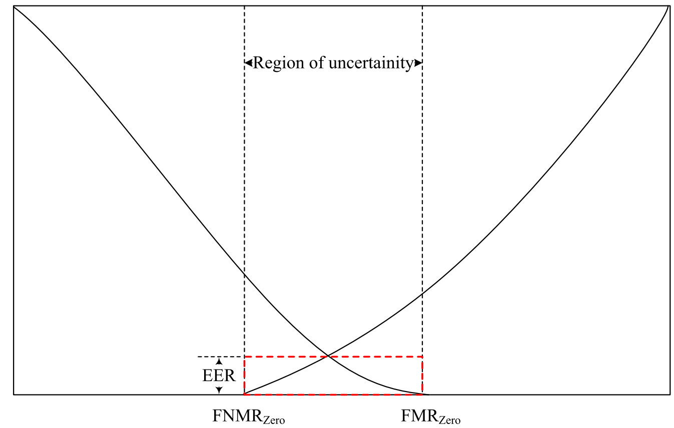

Fig. 3. Explanatory diagram for rectangular area containing region of uncertainty for a particular modality.

to assign more weight to the scores lying into the confidence region (i.e. R_1 and R_2) and relatively less weight to the scores in the region of uncertainty while evaluating the fused score. In this work, we estimate the weights on the basis of mean-closure metric which measures the separation of scores from the mean of the matcher's genuine and imposter distribution for different users. The ratio of these two decides whether the user's score is close to genuine or imposter distribution of matcher 1 or matcher 2. We represent these notations as (*i, m*) for every pair of user and matcher. The mean-closure (Mc_i^m) for a of user-matcher pair (*i, m*) in a multibiometric system is defined as:

$$Mc_i^m = \left(\frac{\mu_i^m(gen) - s_m}{\mu_i^m(imp) - s_m} \right)^2 \tag{5}$$

where, $\mu_i^m(gen)$ and $\mu_i^m(imp)$ represents the mean of genuine distribution and mean of imposter distribution, respectively. Further, the estimated weight for each matcher using MC weighting is computed as follows:

$$w_i^m = \frac{mc_i^m}{\sum_{i=1}^{M} mc_i^m} \tag{6}$$

where, w_i^m is the weight for matcher m and M is the number of matchers for a particular modality. $0 \le w_i^m \le 1$, $\forall i$, $\forall m$, and $\sum_{m=1}^{M} w_i^m = 1, \forall i$.

Here, the weights are proportional to the corresponding mean-closure i.e. the more accurate matcher attains higher weights than those of less accurate matcher for user i. This user-specific score weighting scheme deals optimal with the scores lying in the region of uncertainty. Applying the weights, we achieve fused scores from different matchers of a modality.

3.2. Rectangular area based weighting

In this method, the weights are estimated based on the rectangular area containing a region of uncertainty for the individual modalities in a multibiometric system. The rectangular area (RA) is evaluated as:

$$RA = EER \times (FMR_{zero} - FNMR_{zero}) \tag{7}$$

where FMR_{zero} and $FNMR_{zero}$ are the points where FMR and FNMR become zero, respectively as shown in Fig. 3.

Assuming that the estimated weight for modality k is represented as w_k, the estimated weight for modality N in a multibiometric system using RA weighting technique is computed as follows:

$$w_k = \frac{\frac{1}{RA_k}}{\sum_{k=1}^{N} \frac{1}{RA_k}} \tag{8}$$

where, $0 \le w_k \le 1$, $\forall k$, $\forall N$. The estimated weights are applied in an inversely proportional manner for the available scores of the modalities i.e. lower weight for the modality that provides a larger rectangular area and vice versa.

4. Experimental results and analysis

To perform successful multimodal verification, we present a number of experiments to demonstrate the performance of our proposed score level fusion method. Section 4.1 describes the databases utilized in our work to integrate scores from different modalities. Section 4.2 narrates the experimental settings and performance metrics to quantify the results for each database. Furthermore, the performance of the proposed method is presented in Section 4.3. Also, we compare the proposed methodology with the other approaches in order to specifically measure the effectiveness of the proposed approach in Section 4.4 including statistical evaluation in Section 4.5.

4.1. Database

We evaluate the performance of our method onto two virtual databases involving iris and fingerprint modalities. The virtual databases are created due to the underlying cost and efforts related to multimodal database creation. Most of the multibiometric system proposed in literature utilize a virtual database constructed by pairing a user from one modality with a user from another modality. This pairing assumes that biometric traits of a user are independent. For iris, we use the CASIA V-3-Interval [3] database maintained by the Chinese Academy of Science and Multimedia university database (MMU1) [20]. The CASIA V-3-Interval database contains 2639 high-quality iris images from 249 users collected in two different sessions while MMU1 comprises of left and right iris images for 46 users. Considering left iris and right iris as different subjects, we find that there are 117 left and 121 right iris subjects from 348 total subjects from 249 users of CASIA V-3-Interval which contain at least 7 samples per subject. In MMU1 database, we consider a dataset of 92 users with 5 iris samples assuming left and right iris as a different subject. For fingerprint, we use datasets DB1, DB2 of FVC2002 [11] database containing a total of 800 images of 100 subjects with eight samples each.

The first virtual database (i.e. Virtual_A) comprises of 100 subjects where iris images are randomly selected from 121 right iris subjects of CASIA V-3 Interval and fingerprint from FVC2002DB1 with 7 samples per subject. The second virtual database (i.e. Virtual_B) comprises of 92 subjects where iris images are selected

Please cite this article as: R. Dwivedi, S. Dey, Score-level fusion for cancelable multi-biometric verification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.04.022

from MultiMedia university version-1 (MMU1) database [20] and fingerprint from FVC2002DB2 [11] with 5 samples per subject.

4.2. Experimental settings

Cancelable template for iris and fingerprint are generated and are compared to derive intra-class (i.e. genuine) scores and inter-class (i.e. imposter) scores. We adopt the following protocol to obtain the match scores:

To obtain genuine scores, the first sample of a subject is compared with all remaining samples of the same subject whereas the first sample of a subject is compared with the first sample of remaining subjects to measure imposter scores. For Virtual_A database, the experiment is performed on 100 subjects resulting into 2100 intra-class comparison and 4950 inter-class comparisons. For Virtual_B database, 920 intra-class comparisons and 4186 inter-class comparisons are measured to evaluate the performance. First, the match scores from iris and fingerprint from different matchers are integrated by applying the proposed MCW method. As a result, the fused iris and fused fingerprint scores are obtained. Next, we apply RAW method to combine these scores. Further, the performance of our method is evaluated using the following metrics:

FMR: The probability of getting a positive comparison decision for an imposter;

FNMR: The probability of getting a negative comparison decision for a genuine user;

GMR: Can be measured as 1-FNMR;

EER: The error rate where FMR and FNMR hold equality.

Note that the focus of this work is the score fusion for cancelable multimodal biometric, details of cancelable template generation is not reviewed here.

4.3. Performance evaluation

To estimate the performance of our method, we evaluate EER values and Receiver Operating Characteristic (ROC) curves for unimodal and multimodal databases. Further, we evaluate the performance in term of GMR @ 0.01% FMR since a biometric system deployed in a security application is considered to be efficient if it has low EER and high GMR at low FMR [29].

4.3.1. Virtual_A

The multimodal biometric performance of Virtual_A is evaluated utilizing the scores obtained from CASIA V-3 Interval and FVC2002DB1. First, the performance for individual modalities (i.e. iris and fingerprint) taking part in fusion is evaluated. Next, we evaluate the performance for the multimodal biometric system. The ROC curve for Virtual_A multimodal database is shown in Fig. 4 which demonstrate the performance for the scores obtained in the unprotected and protected domain. From Fig. 4 it has been observed that the performance of the multibiometric system is better than that of a unimodal biometric system utilizing the proposed approach for both domains.

4.3.2. Virtual_B

In a similar manner, the Virtual_B database comprising MMU1 iris and FVC2002DB1 is tested against our method. Figure 5 illustrates the ROC curve for the Virtual_B database. We also demonstrate the ROC curves for individual modalities comprising Virtual_B. It can be noticed from Fig. 5 that the proposed multibiometric system outperforms over the unimodal system for scores obtained through original and cancelable biometric systems.

We have also evaluated the performance with respect to the scores obtained using original biometric templates along with the protected template. From Figs. 4 and 5, it is evident that the performance is degraded by 0.29% and 0.47% for Virtual_A and Virtual_B datasets, respectively. Therefore, we conclude that performance degradation produced by the cancelable transformation is very low. Further, we also evaluate the performance of our method in terms of GMR @ 0.01% FMR and results are reported in Table 1 for Virtual_A and Virtual_B databases, respectively. From Table 1, it is evident that the performance of the multibiometric system using the proposed method is better that that of unimodal systems.

The performance for the Virtual_B database is higher than that of Virtual_A since there is a relative minimal overlap between the genuine and imposter score distributions. The extent of overlap is evaluated by decidability index d', which is defined as:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{2}}} \tag{9}$$

where, μ_1 and μ_2 represent the genuine mean and imposter mean distributions, respectively; and the variances of the genuine and imposter score distributions are represented by σ_1 and σ_2 respectively. The value of d' should be higher if the genuine and imposter distributions are more separable. We achieve the d' of 2.74 and 3.01 for Virtual_A and Virtual_B databases, respectively. The score distributions for Virtual_A and Virtual_B databases are shown in Fig. 6. From Fig. 6, it is evident that the proposed fusion scheme achieves the optimal separation between genuine and imposter distribution for both the virtual databases.

4.4. Comparison with other state-of-the-art methods

To evaluate the robustness of the proposed fusion approach, we have implemented four other well-established weighted fusion methods for comparison in addition to the relevant density and classification based fusion techniques:

4.4.1. EER weighted (EERW)

In this method [33], the weights are assigned based on the EER of the individual matchers. EER is the value at which the FMR and FNMR hold equality. The weight for modality k using EERW is evaluated as:

$$w_k = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{N} \frac{1}{EER_k}}$$

where, EER_k is the EER for matcher k.

4.4.2. D-prime weighted

The d-prime based weighting technique [33] measures the separation between the genuine and impostor scores. Larger separation between the genuine and the impostor scores corresponds to the better performance of a biometric system. The d-prime metric for modality k, d'_k is defined as:

$$d'_k = \frac{\mu_k^G - \mu_k^I}{\sqrt{\sigma_k^{G^2} + \sigma_k^{I^2}}}$$

where, μ_k^G and μ_k^I are the mean for genuine and imposter distributions, respectively whereas σ_k^G and σ_k^I are the standard deviations of the genuine and impostor score distributions, respectively. In the d-prime weighted (DPW) technique, the estimated weight of modality k is defined as follows:

$$w_k = \frac{d'_k}{\sum_{k=1}^{N} d'_k}$$

4.4.3. Fisher's discriminant ratio weighted (FDRW) technique

In this technique [26], the weights of the matchers are estimated based on the separability of the genuine and impostor

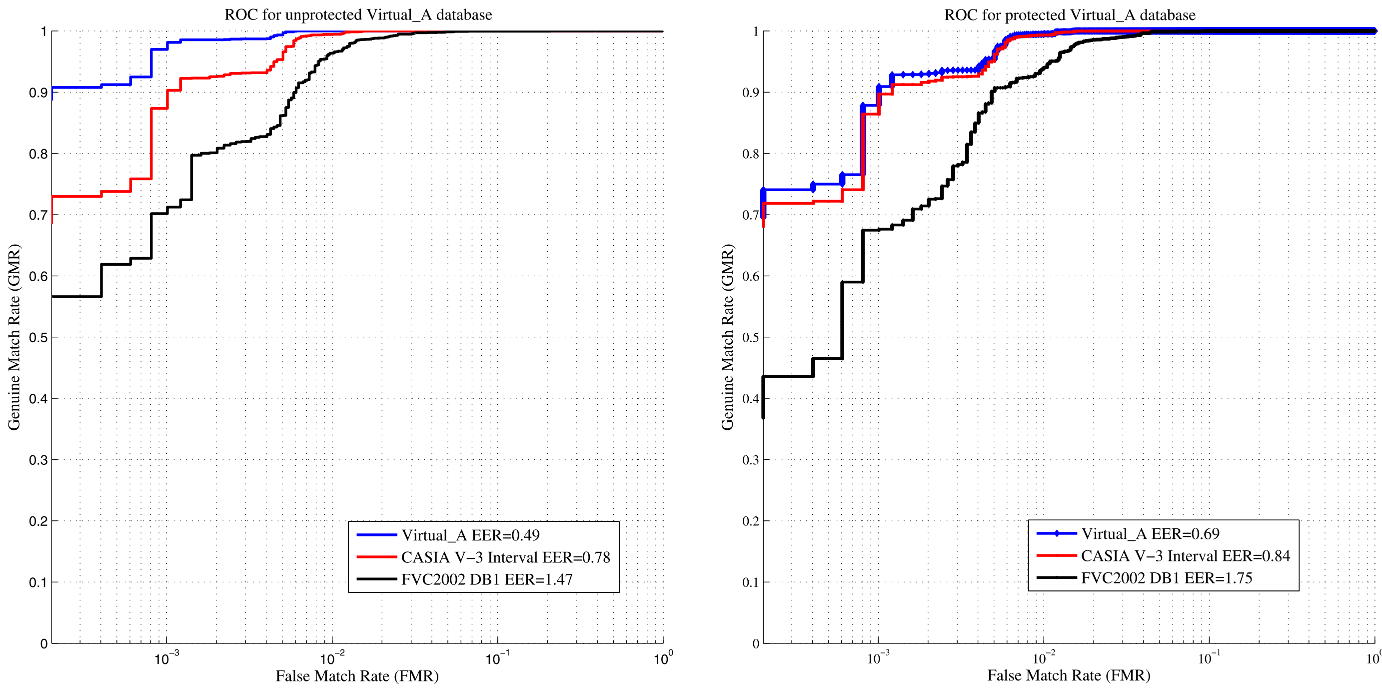

Fig. 4. ROC curves for Virtual_A database.

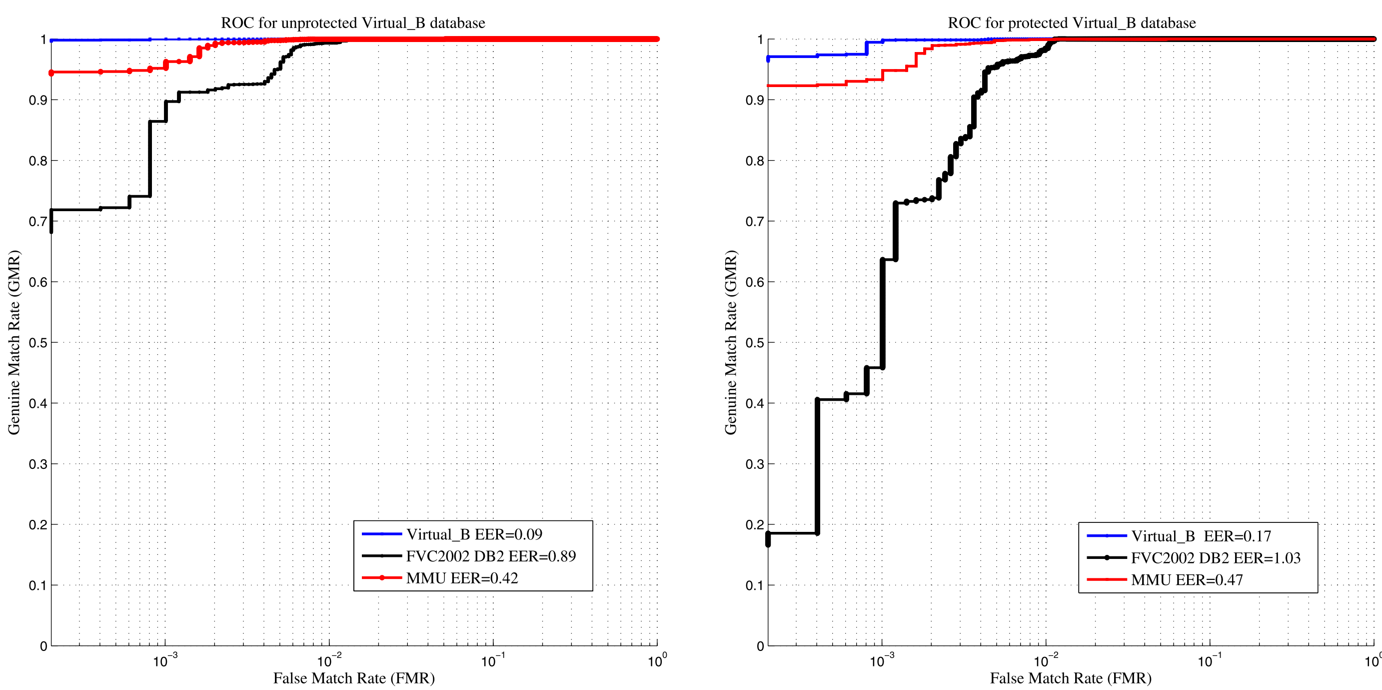

Fig. 5. ROC curves for Virtual_B database.

scores in a multi-biometric system. The FDR of a biometric system is defined as:

$$FDR_k = \frac{(\mu_k^G - \mu_k^I)^2}{\sigma_k^{G2} + \sigma_k^{I2}}$$

The estimated weight is proportional to the computed value of *FDR* i.e. a high-performance biometric authentication system has a high value of *FDR*. The weight based on *FDRW* technique for modality k is computed as:

$$w_k = \frac{FDR_k}{\sum_{k=1}^{N} FDR_k}$$

4.4.4. Mean-to-Extrema weighted (MEW) technique

In this technique [4], weights are estimated using the mean of the scores distribution and its maxima i.e. the two extremes of the overlap region. The mean-to-extrema (*ME*) for a matcher is computed as:

$$ME_k = \left(Max_k^I - \mu_k^I\right) + \left(\mu_k^G - Max_k^G\right)$$

For modality k, the weight using the MEW technique is computed using the equation:

$$w_k = \frac{MEW_k}{\sum_{k=1}^{N} MEW_k}$$

Table 1
Performance comparison of proposed method with existing weighting approaches (in %).

Methods	EER				GMR @0.01% FMR			
	Virtual_A		Virtual_B		Virtual_A		Virtual_B	
	unprotected	protected	unprotected	protected	unprotected	protected	unprotected	protected
Density-based methods								
Nandakumar et al. [21]	0.89	1.03	0.77	0.95	99.01	98.82	99.23	98.90
Nanni et al. [22]	1.25	1.48	1.03	1.20	98.65	98.45	98.91	98.75
Tao et al. [34]	0.98	1.19	1.08	1.31	98.90	98.70	98.83	98.69
Classification-based methods								
Tronci et al. [36]	1.35	1.62	0.55	0.68	98.53	98.30	99.43	99.25
Nguyen et al. [24]	1.52	1.72	0.97	1.28	98.39	98.13	98.93	98.70
Transformation-based (weighting) methods								
EERW [33]	0.59	0.77	0.21	0.32	99.11	98.53	99.63	99.06
DPW [33]	0.52	0.73	0.14	0.29	99.29	98.72	99.81	99.29
FDRW[26]	0.89	0.98	0.28	0.42	98.60	98.09	99.05	98.78
MEW [4]	0.91	1.09	0.34	0.49	98.39	97.90	98.81	98.71
Proposed score fusion	**0.49**	**0.69**	**0.09**	**0.17**	**99.59**	**98.89**	**99.97**	**99.64**

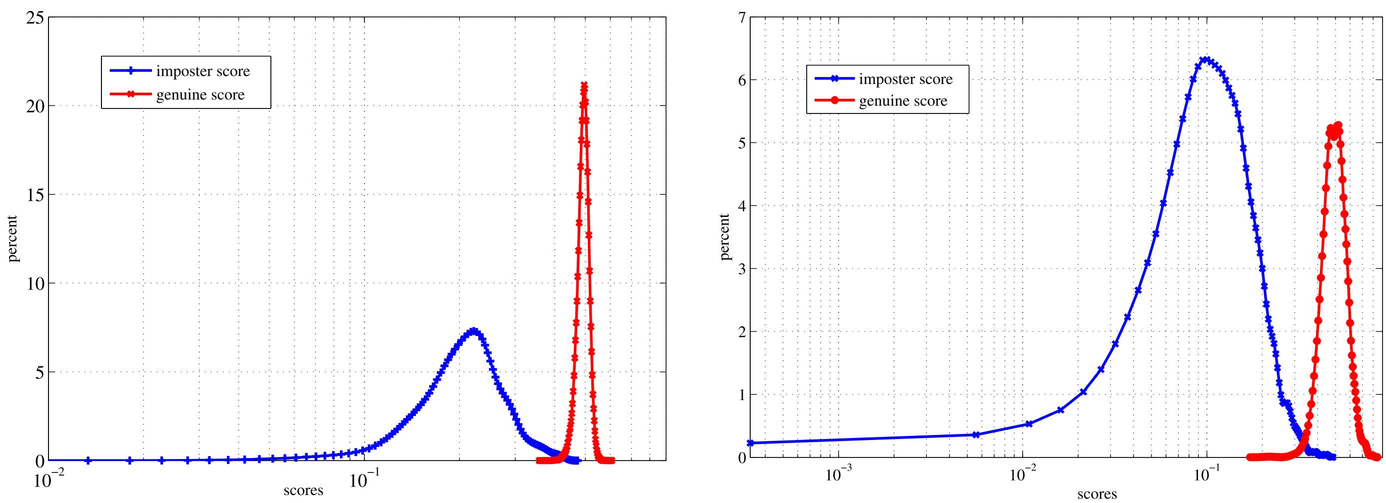

Fig. 6. Distribution curves of the fused matching scores.

It has been analyzed that DPW [33], MEW [4] and FDRW [26] techniques only involves the scores outside the uncertainty region to estimate the weight which results the performance sensitive to outliers. Furthermore, EER cannot be considered as weight estimation factor since a matcher with a lower EER may have higher FMR than the other one. In this work, the weights for individual matcher's are estimated based on the rectangular overlap area in order to assign the less weight to the weak matcher. Hence, the performance of the proposed method is better than that of the EERW [33], DPW [33], FDRW [26], MEW [4] methods.

We also implemented few relevant density-based[1] and classification-based[2] methods to perform a robust comparative analysis. Table 1 reports the EER(%) and GMR @ 0.01% FMR, obtained using the proposed and existing weighting (transformation) [4,26,33], density based [21,22,34] and classification based [24,36] approaches. From Table 1, it has been observed that the proposed multi-biometric system (i.e. cancelable iris - cancelable fingerprint system), provides lower EER and higher GMRs @ 0.01% FMR than that of the existing fusion techniques. Also, the best performance in terms of EER (i.e. 0.69% and 0.17%) and GMR @ 0.01% FMR (i.e. 98.89% and 99.64%) are achieved using the proposed weighting technique for the two virtual multimodal databases additionally providing secure authentication.

4.5. Statistical evaluation of proposed score fusion method

The performance of any biometric system is affected by the size of the database and image comprising the database. ROC curves and verification performance are not enough to validate the overall performance for the multibiometric system. In the literature, the statistical significance of the achieved performance is evaluated by a commonly used method proposed in [2] which utilizes the Half Total Error Rate (HTER) and Confidence Interval (CI). Hence, we test our method against these two parameters. HTER is computed as:

$$HTER = \frac{FMR + FNMR}{2}$$

In order to compute CI around HTER, we look for the bound $\sigma \times z_{\alpha/2}$. Here, σ and $z_{\alpha/2}$ are defined as [2]:

$$\sigma = \sqrt{\frac{FMR(1-FMR)}{4 \cdot NI} + \frac{FNMR(1-FNMR)}{4 \cdot NG}}$$

$$z_{\alpha/2} = \begin{cases} 1.645 & \text{for} \quad 90\%\ CI \\ 1.960 & \text{for} \quad 95\%\ CI \\ 2.576 & \text{for} \quad 99\%\ CI \end{cases}$$

where, NG and NI represents the total number of intra-class comparisons and the total number of inter-class comparisons,

[1] https://msu.edu/dingyaoh/WebpageofGUI/FusionTool.htm, http://www.lx.it.pt/mtf/mixturecode.zip.
[2] http://www.ti3.tu-harburg.de/rump/intlab/.

Please cite this article as: R. Dwivedi, S. Dey, Score-level fusion for cancelable multi-biometric verification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.04.022

Table 2
Confidence interval (CI) around HTER of the *d*-prime weighting (DPW) and proposed fusion methods.

Methods	HTER		Confidence interval (%) around HTER for					
			90%	95%	99%	90%	95%	99%
	Virtual_A	Virtual_B	Virtual_A			Virtual_B		
D-prime weighting (DPW) [33]	0.73	0.27	0.03	0.045	0.059	0.047	0.061	0.078
Proposed method	**0.67**	**0.13**	**0.02**	**0.038**	**0.047**	**0.04**	**0.049**	**0.052**

respectively. We evaluate HTER and CI for both of the virtual databases using the FMR and FNMR. The statistical evaluation is carried out at 0.01% FMR and results are reported in Table 2. From Table 2, it has been observed that HTER lies between 0.02 ± 0.05 with 95% confidence for both of the virtual databases. This validates the achieved performance in our method. In the runner-up DPW [33] method, the HTER for Virtual_A and Virtual_B database is 0.73 and 1.09, respectively. Also, the CI around HTER lies in between 0.03 ± 0.08 for both virtual databases which is inferior than the proposed method. This confirms the statistical soundness of the proposed fusion method over the state-of-the-art.

Further, the comparative analysis shows that the proposed fusion method outperforms over the existing weighting approaches. Also, we obtain a substantial improvement over recognition performance through the efficient fusion of match scores from cancelable biometric modalities providing secrecy over different applications. As described above, the substantial improvement over runner-ups i.e. existing fusion methods lies in achieving (i) minimal EER amongst all, (ii) Highest GMR@ 0.01%FMR (required for security applications), (iii) No requirement of learning, (iv) Deals optimally with the region of uncertainty and weight computation utilizing all set of scores.

5. Conclusion

In this paper, the score level fusion is performed onto the match scores obtained from cancelable biometric templates. The proposed two-level fusion method applies MC weighting and RA weighting at the first and second level, respectively. RA weighting utilizes the rectangular area containing region of uncertainty for each modality while MC weighting computes the optimal score for each matcher to be fused. The weighting techniques incur the minimal computational complexity without the need of any learning. Experimental evaluations vindicate that the proposed two-level cancelable multibiometric fusion method attains better performance compared to the cancelable unimodal biometric systems in terms of EER, d' and GAR for both the virtual databases. Further, the comparative analysis shows that the proposed fusion method outperforms over the existing weighting approaches. Also, we obtain a substantial improvement over recognition performance through the efficient fusion of match scores from cancelable biometric modalities providing secrecy over different applications. It is hoped that the proposed approach would be tested onto large databases containing 1000 subjects with more than two modalities. Additionally, we are also focusing on decision level fusion for cancelable multimodal biometric systems in future.

Conflict of interest

This research is sponsored by SERB (ECR/2017/000027), Department of Science & Technology, Govt. of India. I have disclosed this information in acknowledgment. Rest, there is no conflict of interest.

Acknowledgments

The authors are thankful to SERB (ECR/2017/000027), Department of science & Technology, Govt. of India for providing financial support. Also, We would like to acknowledge Indian Institute of Technology Indore for providing the laboratory support and research facilities to carry out this research.

References

[1] H. Benaliouche, M. Touahria, Comparative study of multimodal biometric recognition by fusion of iris and fingerprint, Sci. World J. 2014 (2014).
[2] S. Bengio, J. Mariéthoz, A statistical significance test for person authentication, in: Proceedings of Odyssey 2004: The Speaker and Language Recognition Workshop, 2004, pp. 237–244.
[3] CASIA V-3-Interval database, Casia iris image database version 3.0, http://www.cbsr.ia.ac.cn/Databases.htm.
[4] N. Damer, A. Opel, A. Nouak, Biometric source weighting in multi-biometric fusion: towards a generalized and robust solution, in: 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 1382–1386.
[5] S.C. Dass, K. Nandakumar, A.K. Jain, A principled approach to score level fusion in multimodal biometric systems, in: International Conference on Audio- and Video-Based Biometric Person Authentication, Springer, 2005, pp. 1049–1058.
[6] J. Daugman, How iris recognition works, in: International Conference on Image Processing, 1, 2002, pp. 33–36.
[7] R. Dwivedi, S. Dey, A non-invertible cancelable fingerprint template generation based on ridge feature transformation, Unpublished manuscript (2017).
[8] R. Dwivedi, S. Dey, R. Singh, A. Prasad, A privacy-preserving cancelable iris template generation scheme using decimal encoding and look-up table mapping, Comput. Secur. 65 (2017) 373–386.
[9] M.A.T. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 381–396.
[10] M. Fontani, T. Bianchi, A.D. Rosa, A. Piva, M. Barni, A framework for decision fusion in image forensics based on Dempster Shafer theory of evidence, IEEE Trans. Inf. Foren. Secur. 8 (4) (2013) 593–607.
[11] FVC2002 fingerprint verification competition, https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/Home.aspxp.
[12] M. Hanmandlu, J. Grover, A. Gureja, H. Gupta, Score level fusion of multimodal biometrics using triangular norms, Pattern Recognit. Lett. 32 (14) (2011) 1843–1850.
[13] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M.K. Khan, K.O. Sentosa, Performance evaluation of score level fusion in multimodal biometric systems, Pattern. Recognit. 43 (5) (2010) 1789–1800.
[14] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern. Recognit. 38 (12) (2005) 2270–2285.
[15] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[16] C. Lobrano, R. Tronci, G. Giacinto, F. Roli, Dynamic linear combination of two–class classifiers, in: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), Springer, 2010, pp. 473–482.
[17] Y. Ma, B. Cukic, H. Singh, A classification approach to multi-biometric score fusion, in: International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 2005, pp. 484–493.
[18] L. Masek, Recognition of Human Iris Patterns for Biometric Identification, Technical Report, University of Western Australia, 2003.
[19] M.M. Monwar, M.L. Gavrilova, Multimodal biometric system using rank-level fusion approach, IEEE Trans. Syst., Man, and Cyber., Part B (Cyber.) 39 (4) (2009) 867–878.
[20] Multimedia University, MMMU1 iris image database, www.cs.princeton.edu/~andyz/downloads/MMUIrisDatabase.zip.
[21] K. Nandakumar, Y. Chen, S.C. Dass, A. Jain, Likelihood ratio-based biometric score fusion, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 342–347.
[22] L. Nanni, A. Lumini, S. Brahnam, Likelihood ratio based features for a trained biometric score fusion, Expert Syst. Appl. 38 (1) (2011) 58–63.
[23] L. Nanni, A. Lumini, M. Ferrara, R. Cappelli, Combining biometric matchers by means of machine learning and statistical approaches, Neurocomputing 149 (2015) 526–535.
[24] K. Nguyen, S. Denman, S. Sridharan, C. Fookes, Score-level multibiometric fusion based on Dempster Shafer theory incorporating uncertainty factors, IEEE Trans. Hum. Mach. Syst. 45 (1) (2015) 132–140.

[25] J. Peng, A.A.A. El-Latif, Q. Li, X. Niu, Multimodal biometric authentication based on score level fusion of finger biometrics, Optik-Int. J.Light Electron Opt. 125 (23) (2014) 6891–6897.
[26] N. Poh, S. Bengio, A study of the effects of score normalisation prior to fusion in biometric authentication tasks, Idiap-RR Idiap-RR-69-2004, IDIAP, 2004.
[27] N. Poh, J. Kittler, A. Rattani, M. Tistarelli, Group-specific score normalization for biometric systems, in: International Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 38–45.
[28] N. Poh, A. Ross, W. Lee, J. Kittler, A user-specific and selective multimodal biometric fusion strategy by ranking subjects, Pattern Recognit. 46 (12) (2013) 3341–3357.
[29] S. Prabhakar, A.K. Jain, Decision-level fusion in fingerprint verification, Pattern Recognit. 35 (4) (2002) 861–874.
[30] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognit. Lett. 24 (13) (2003) 2115–2125.
[31] A.A. Ross, K. Nandakumar, A.K. Jain, Handbook of Multibiometrics, Springer-Verlag New York, Inc., 2006.
[32] R. Singh, M. Vatsa, A. Noore, S.K. Singh, Dempster-Shafer theory based classifier fusion for improved fingerprint verification performance, in: 5th Indian Conference on Computer Vision, Graphics and Image Processing, Springer, 2006, pp. 941–949.
[33] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 450–455.
[34] Q. Tao, R. Veldhuis, Robust biometric score fusion by naive likelihood ratio via receiver operating characteristics, IEEE Trans. Inf. Foren. Secur. 8 (2) (2013) 305–313.
[35] K.-A. Toh, X. Jiang, W.-Y. Yau, Exploiting global and local decisions for multimodal biometrics verification, IEEE Trans. Signal Process. 52 (10) (2004) 3059–3072.
[36] R. Tronci, G. Giacinto, F. Roli, Dynamic score selection for fusion of multiple biometric matchers, in: 14th International Conference on Image Analysis and Processing, IEEE, 2007, pp. 15–22.
[37] M. Vatsa, R. Singh, A. Noore, A. Ross, On the dynamic selection of biometric fusion algorithms, IEEE Trans. Inf. Foren. Secur. 5 (3) (2010) 470–479.
[38] N. Wang, Q. Li, A.A.A. El-Latif, X. Yan, X. Niu, A novel hybrid multibiometrics based on the fusion of dual iris, visible and thermal face images, in: 2013 International Symposium on Biometrics and Security Technologies, 2013, pp. 217–223.
[39] W.J. Wong, A.B. Teoh, M.D. Wong, Y.H. Kho, Enhanced multi-line code for minutiae-based fingerprint template protection, Pattern Recognit. Lett. 34 (11) (2013) 1221–1229.