Rank-Based Score Normalization for Multi-Biometric Score Fusion

Panagiotis Moutafis and Ioannis A. Kakadiaris
Computational Biomedicine Lab, Department of Computer Science
University of Houston, 4800 Calhoun Rd., Houston, TX 77004
{pmoutafis,ioannisk}@uh.edu

***Abstract*—The matching score distributions produced by different biometric modalities are heterogeneous. The same is true for the matching score distributions obtained for different probes. Both of these problems can be addressed by score normalization methods that standardize the corresponding distributions. In our previous work we demonstrated that, in the case of multi-sample galleries, the matching score distributions are also heterogeneous between different subsets of matching scores obtained for the same probe. In this paper, we use this result to propose a rank-based score normalization framework for multi-biometric score fusion. Specifically, in addition to normalizing the matching scores produced for each biometric modality independently, we propose to further join them to form a single set. This set is then partitioned to subsets using a rank-based scheme. The theory of stochastic dominance demonstrates that the rank-based scheme imposes the distributions of the subsets to be ordered. Hence, by normalizing the matching scores of each subset independently, better normalized scores are produced. The normalized scores can be fused using any fusion rule. Experimental results using face and iris data from the CASIA-Iris-Distance database demonstrate the improvements obtained.**

***Index Terms*—Score Normalization, Score Fusion, Multi-Biometric Systems**

I. Introduction

Biometric systems use measurable biological and behavioral characteristics to perform automated recognition. Some of the most popular biometric traits are the face, iris, voice, and fingerprints. Systems that rely on a single trait are usually called *unimodal biometric systems*. The recognition performance of such systems is usually degraded for several reasons, such as lack of uniqueness and noisy data. Multi-biometric systems, also known as multimodal biometric systems, address these problems by utilizing multiple modalities. That is, data from two or more biometric traits are used for each subject (e.g., face and iris). Fusing the information obtained from the different modalities is not a trivial problem. On the contrary, it has a significant impact on the overall recognition performance. The most common way to address this problem is to perform feature level or matching score level fusion. Feature level fusion methods utilize more information compared to the matching score level fusion approaches. However, they are computationally expensive and require large training datasets. On the other hand, score level fusion methods rely on matching scores (i.e., one number for each pairwise comparison) and thus they are very efficient. In addition, the matching score distributions usually provide sufficient information to perform effective fusion. In most cases though, the distributions produced by the different modalities are heterogeneous, which complicates the fusion process. There are two ways to address this problem. First, some methods utilize the training data to learn optimal weights for the matching scores produced by each modality. Second, other approaches normalize the matching score distributions of each modality independently. As a result, the corresponding distributions become homogeneous and the fusion process is simplified. Score normalization techniques, though, go beyond that. Each biometric sample is subject to distortions during the data acquisition. Consequently the matching score distributions obtained for different probes are heterogeneous. This phenomenon degrades the performance of both unimodal and multi-biometric systems. By normalizing the matching scores obtained for each probe this problem is alleviated and better performance is obtained. In our previous work [1], [2], we focused on unimodal biometric systems with multi-sample galleries. We demonstrated that the matching scores obtained for a single probe can be partitioned into subsets in such a way that the corresponding distributions are heterogeneous. The theory of stochastic dominance guarantees this result. Hence, by normalizing the matching scores of each subset individually, better normalized scores are obtained on a per probe basis. We named that method RBSN (i.e., Rank-Based Score Normalization) and some of its advantages include: (i) it can improve the performance of any score normalization method by utilizing the existing information more effectively; (ii) it improves the recognition performance on a per probe basis (i.e., the transformation of the scores is non-linear); and (iii) it increases the discriminability of the matching scores across probes. An overview of RBSN is provided in Fig. 1.

In this paper, we extend this approach to the case of multi-biometric score fusion. Existing approaches normalize the matching scores of each modality independently and then employ a fusion rule. We argue that, as in the case of multi-sample galleries, we can utilize the multiple matching scores produced for each subject by the different modalities more effectively. In particular, the first step normalizes the matching score of each modality independently. This step alleviates the differences of the matching score distributions between the modalities. Then, we propose to join the normalized score sets to form a single set and employ RBSN to produce "twice" normalized scores. This step utilizes the information from the multiple matching scores obtained for each subject

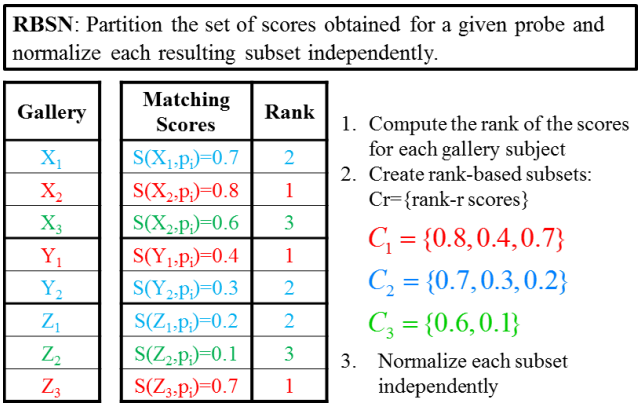

Fig. 1: Overview of the Rank-Based Score Normalization framework. The notation $S(X_1, p_i)$ is used to denote the score obtained by comparing a probe p_i to the biometric sample 1 of a gallery subject labeled X. [1], [2]

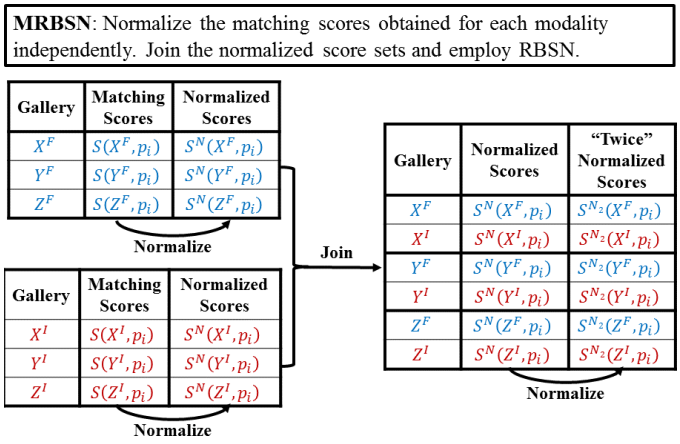

Fig. 2: Overview of the Multi-Rank-Based Score Normalization framework. The capital letters X, Y, and Z denote the labels of three different subjects, while the superscripts *F* and *I* denote face and iris biometric traits, respectively. The notation $S(X^F, p_i)$ is used to denote the matching score obtained by comparing a probe p_i to the facial biometric sample of X. The notations $S^N(X^F, p_i)$ and $S^N(X^F, p_i)$ denote the normalized and "twice" normalized scores, respectively.

more effectively. Finally, the resulting scores can be fused using any fusion rule. We call the proposed framework Multi-Rank-Based Score Normalization (MRBSN). An overview of MRBSN is provided in Fig. 2. Experimental results using face and iris data from the CASIA-Iris-Distance database [11] demonstrate the benefits of multi-biometric fusion.

The rest of this paper is organized as follows: Sec. II reviews the fusion rules and score normalization techniques used in our experimental evaluation; Sec. III offers an overview of the theory of stochastic dominance and the rank-based score normalization framework and describes the proposed multi-biometric rank-based score normalization; Sec. IV presents the experimental results; and Sec. V concludes the paper.

II. Related Work

In this section, we review the methods used in our experimental evaluation. For a comprehensive overview of score normalization methods and score fusion rules for multi-biometric systems we refer the readers to Jain *et al.* [3].

A. Score Normalization Techniques

Score normalization techniques make the matching score distributions homogeneous between: (i) different biometric samples (i.e., probes), (ii) different modalities, and (iii) different subsets of matching scores obtained for the same biometric sample.

Z-score: This method is very easy to implement and usually yields significant improvements in terms of recognition performance. Therefore, it is widely used and its properties are well examined. Specifically, Z-score relies on second order statistics. That is, it relies on the assumption that the location and scale parameters of the matching score distribution can be approximated by the mean and standard deviation in a satisfactory manner. When the underlying distribution is Gaussian, the Z-score transformation can retain the shape of the distribution. However, the normalized scores are not bounded. Moreover, Z-score is sensitive to outliers as the mean and standard deviation estimators are not robust to observations with extreme values. Finally, since this is a linear operation, it does not change the ordering of matching scores obtained for a single probe. Hence, it does not affect the rank-k recognition performance.

W-score: This score normalization method was proposed by Scheirer *et al.* [4] and normalizes the matching scores by modeling the tail of the non-match scores distribution. The normalized scores take values in the interval $[0, 1]$. This approach relies on the Extreme Value Theory (EVT). The necessary conditions to invoke EVT are detailed in Section 3 of Scheirer *et al.* [4]. The strongest points of this approach are that it does not make any assumptions about the distribution of the matching scores and it yields good performance. However, it is not clear how many non-match scores should be selected to model the tail of the non-match scores distribution. In the literature, it is reported that selecting as few as five scores should be enough for that task. In our experience, though, using a small number of scores results in discretized normalized scores. The implication of this fact is that the user cannot assess the recognition performance at low false acceptance rates. On the other hand, if the user selects too many scores then the necessary conditions to invoke the Extreme Value Theorem are violated. In addition to the requirement of selecting the number of non-match scores used to model the tail of the non-match scores distributions, W-score has another limitation. It assumes that each matching score corresponds to a single gallery subject. As a result, it cannot be directly applied to multi-sample galleries. One possible solution would be to fuse the scores for each subject before applying W-score. However, we demonstrate that the rank-based score normalization framework proposed in our previous work can naturally extend the use of W-score to multi-sample galleries and improve its performance.

B. Fusion Rules

The general problem of combining various classifiers, or equivalently fusing evidence from multiple measurements, is an area that has been well studied. The work by Kittler *et al.* [5] focuses on the statistical background of such fusion rules. Even though the corresponding results refer to likelihood values, these rules are usually applied to matching scores as well. Specifically, as demonstrated in the literature, these rules work well whether the user combines multiple matching scores per subject from a single modality or one matching score per subject from multiple modalities [6], [1], [2]. In our experiments, we use the *sum* rule. Under the assumption of equal priors the *sum* rule is implemented by employing the *addition* operator. When this assumption is not true the *mean* operator is employed instead. Even though this rule makes restrictive assumptions, it appears to yield good performance as demonstrated in the relevant literature [5], [3].

III. Rank-Based Score Normalization for Multi-Biometric Score Fusion

In this section, we first review the theory of stochastic dominance and the RBSN framework. These two are inseparable components of the proposed approach. Then, we describe the MRBSN framework.

A. Stochastic Dominance Theory

The theory of stochastic dominance is a branch of decision theory. It is most often applied in portfolio analysis for financial applications. We focus only on the results that help us support the proposed framework.

Definition: The notation $X \gtrsim_{FSD} Y$ denotes that X first order stochastically dominates Y, that is

$$Pr\{X > z\} \geq Pr\{Y > z\}, \quad \forall z. \tag{1}$$

This definition implies that the corresponding distributions will be ordered. The following lemma makes this observation more clear.

Lemma: Let X and Y be any two random variables, then

$$X \gtrsim_{FSD} Y \Rightarrow E[X] \geq E[Y]. \tag{2}$$

The proof of this lemma can be found in [7]. Figure 1 of Wolfstetter *et al.* [7] depicts an illustrative example of first order stochastic dominance, where $\bar{F}(z)$ and $\bar{G}(z)$ are two functions such that $\bar{F}(z) \gtrsim_{FSD} \bar{G}(z)$. It is relatively easy to show that a first order stochastic dominance relationship implies all higher orders as well [8]. Moreover, as it has been implicitly illustrated by Birnbaum *et al.* [9], this relation is known to be transitive. Finally, the first order stochastic dominance is often referred to as stochastic ordering of random variables.

Key Remarks: If a variable stochastically dominates another then we can conclude that the corresponding distributions are going to be ordered (i.e, heterogeneous). In our previous work, we used this result to illustrate that a rank-based partitioning of the matching scores for a single probe will result in subsets of scores with ordered distributions. Hence, by normalizing the matching scores of each subset independently the corresponding distributions will become homogeneous and better normalized scores will be obtained. In this work, the same idea is adopted for a set of scores obtained for a single probe but from multiple modalities.

B. Rank-Based Score Normalization

The main idea of RBSN [1], [2] is to partition the set of matching scores obtained for a single probe into subsets and then normalize the matching scores of each subset independently. This is a framework that can be used in conjunction with any score normalization method. It relies on the assumption that multiple samples per subject are available. We offer an overview of the steps performed in *Algorithm 1*. The notation to be used throughout this paper is as follows:

- S: the set of matching scores for a given probe when compared against a given gallery
- S_i: the set of matching scores that correspond to the gallery subject with *identity=i*, $S_i \subseteq S$
- $S_{i,r}$: the ranked-r score of S_i
- S^N: the set of normalized scores for a given probe
- C_r: the rank-r subset, $\bigcup_r C_r = S$
- $|d|$: the cardinality of a set d
- U: the set of unique gallery identities
- Z: a given score normalization technique

Algorithm 1 Rank-Based Score Normalization

procedure *RBSN*($S = \bigcup_i \{S_i\}$, Z)
  Step 1: Partition S into subsets
    $C_r = \{\emptyset\}, \forall r$
    for $r = 1 : max_i\{|S_i|\}$ **do**
      for all $i \epsilon U$ **do**
        $C_r = C_r \bigcup S_{i,r}$
      end for
    end for ▷ (i.e., $C_r = \bigcup_i S_{i,r}$)
  Step 2: Normalize each subset C_r
    $S^N = \{\emptyset\}$
    for $r = 1 : max_i\{|S_i|\}$ **do**
      $S^N = S^N \bigcup Z(C_r)$
    end for
    return S^N
end procedure

For a detailed description of each step of the algorithm we refer the readers to our previous work [1], [2]. Here, we focus on some key remarks.

Key Remarks: The subsets C_r obtained from Step 1 include at most one matching score for each gallery subject. Also, the theory of stochastic dominance ensures that the corresponding distributions are ordered. By construction we have that

$$S_{x,i} \geq S_{x,j}, \forall i \leq j \quad \text{and} \quad \forall x. \tag{3}$$

Let X_i and X_j be the variables that correspond to $S_{x,i}$ and $S_{x,j}$ (i.e., C_i and C_j), respectively. Hadar and Russell [10] have demonstrated that this condition is sufficient to conclude

that $X_i \gtrsim_{FSD} X_j$. By invoking the corresponding results from Sec. III-A, it is clear that, if $i \neq j$, the density distributions P_{X_i} and P_{X_j} will be ordered. Hence, by normalizing the subsets C_r independently better normalized scores can be obtained. Even though there are other ways to define the subsets C_r (e.g., ranking by illumination or pose), only the rank-based scheme can guarantee that the corresponding density distributions will be ordered. Moreover, conventional score normalization methods such as Z-score do not change the ordering of the scores and therefore do not affect the rank-1 identification performance. The rank-based framework addresses this problem as the order of the normalized scores S^N is different from the input matching scores S. Finally, since the subsets C_r include at most one score per subject, W-score can be employed without having to first fuse the matching scores. However, since $|C_r| < |S|$, less information is available for the parameter estimation required by some score normalization techniques (e.g., estimates of the mean and standard deviation values for Z-score).

Implementation Details: Any ties on the matching scores can be broken arbitrarily without affecting the final outcome. Moreover, Alg. 1 can be implemented using parallel programming. For example, the ranking of the matching scores for each subject can be performed in parallel. In addition, the matching score contained in each subset can be simultaneously normalized as they are independent operations. In real life applications, the galleries used might include a different number of samples per subject. Hence, some of the subsets defined might contain only a few matching scores. In these cases, it is better not to perform any normalization. We have found that replacing these scores with *Not a Number (NaN)* and ignoring them at a decision level does not affect the performance. The reason is that this phenomenon is likely to happen for subsets C_r of a low rank. Hence, the information omitted is not useful. Finally, there are many factors that affect the final performance, such as the quality of the input matching scores, the score normalization technique and fusion rule used, or the order in which they are applied. In this paper, we always normalize the matching scores before we fuse them.

C. Multi-Rank-Based Score Normalization

In this section, we first describe the conventional way of utilizing score normalization methods for intuitive score fusion, and then we present the proposed MRBSN for multi-biometric systems.

Without loss of generality, we assume that the system at hand relies on face and iris biometric traits. We further assume that only one biometric sample per gallery subject is available for each modality. The notation S^F is used to denote the set of matching scores obtained by comparing a given probe with the gallery that comprises facial biometric samples, and S^I is used to denote the set of matching scores obtained by comparing the probe with the gallery that comprises iris biometric samples. According to the conventional approach, each set of matching scores is first normalized with a given score normalization technique. The corresponding sets with the normalized scores are denoted $S^{F,N}$ and $S^{I,N}$, respectively. Finally, the normalized scores can be fused using any given fusion operator, such as mean. We argue that by joining the normalized score sets obtained from different modalities we define an "artificial" unimodal biometric system with a multi-sample gallery. As a result, the RBSN algorithm can be utilized to normalize the scores more effectively. In our example, the two sets of normalized scores $S^{F,N}$ and $S^{I,N}$ can be joined to obtain $S = S^{F,N} \cup S^{I,N}$. The set S fulfills the required conditions to employ RBSN as two scores correspond to each gallery subject. The reason why we use $S = S^{F,N} \cup S^{I,N}$ instead of $S = S^F \cup S^I$ is that the matching score distributions of S^F and S^I are already ordered. Hence, the rank-based scheme would simply partition S to two subsets identical with S^F and S^I. The MRBSN implementation is described by Alg. 2. The notation to be used is as follows:

S^J: the set of matching scores obtained for a given probe using the modality denoted by J
$S^{J,N}$: the set of normalized scores for a given probe
S: the set of joined normalized score sets, $S = \bigcup_J S^{J,N}$
S^{N_2}: the set of "twice" normalized scores
R: a given fusion rule

Algorithm 2 Multi-Rank-Based Score Normalization

procedure *MRBSN*(S^J, Z, R)
Step 1: Normalize each S^J independently
  for all J **do**
    $S^{J,N} = Z(S^J)$
  end for
Step 2: Join $S^{J,N}$
  $S = \bigcup_J S^{J,N}$
Step 3: Employ RBSN
  $S^{N_2} = RBSN(S, Z)$
Step 4: Fuse the "twice" normalized scores
  $S^{N_2} = R(S^{N_2})$
  return S^{N_2}
end procedure

Step 1: The sets of matching scores obtained for each modality are normalized independently. Thus, the corresponding distributions become homogeneous.

Step 2: The sets of normalized scores obtained in Step 1 are joined to form a single set. As a result, the required conditions for RBSN are satisfied.

Step 3: The set of joined score sets obtained in Step 2 is normalized by employing RBSN to generate "twice" normalized scores.

Step 4: The "twice" normalized scores obtained in Step 3 are fused so that one score corresponds to each gallery subject.

Key remarks: The proposed approach implicitly requires that the rank-based scheme can partition the set of scores S in a meaningful way. That is, the distributions of the subsets C_r should be stochastically ordered. For the case of unimodal biometric systems with multi-sample galleries this implicit requirement is usually satisfied by the diversity of the gallery samples. Hence, the implicit requirement of MRBSN is that the variation due to the multiple samples per gallery subject is greater than the variation attributed to the heterogeneous behavior of the matching scores produced by

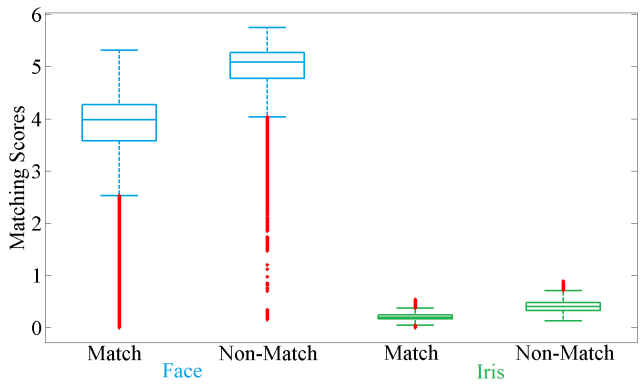

Fig. 3: Boxplots of the match and non-match scores used for each modality. The two boxplots on the left correspond to face scores, while the two boxplots on the right correspond to iris scores.

the different modalities. Step 1 minimizes the latter by making the corresponding score distributions homogeneous.

IV. Experimental Results

In this section, we describe the database used and provide information about implementation details. Finally, we present the corresponding experimental results.

CASIA-Iris-Distance Database: The CASIA-Iris-Distance [11] images were acquired using a long-range multi-modal biometric image acquisition and recognition system (LMBS) developed by the CASIA group. The high resolution camera used enables use of the captured images for both face and iris recognition. The images were captured in an indoor environment in a single session. Most of the 142 subjects are graduate students of the CASIA group. The number of samples per subject ranges from 10 to 23, resulting in 2,567 images.

Implementation details: The CASIA group provided us with pairwise distances for the face and iris traits (i.e., $2 \times 3,296,028$ distances). To transform the distances into matching scores the following formula was used $score = max(distance) - distance$. This way, the scaling of the score distributions was not altered. The obtained matching scores lie in the interval $[0, max(distance)]$. This formula was applied for the distances of each modality independently. The corresponding boxplots for the match and non-match scores are depicted in Fig. 3. As demonstrated, the corresponding distributions are heterogeneous. The face matching scores have a higher mean value than the iris matching scores. To assess the discriminative properties of the two modalities we computed the corresponding Receiver Operating Characteristic (ROC) curves. The Area Under the Curve (AUC) obtained for the face matching scores is 93.48%, while the AUC obtained for the iris matching scores is 94.17%. Since W-score is not directly applicable to multi-sample galleries, when needed, the corresponding matching scores were first fused using the mean operator. In all cases, 35 matching scores were used to fit the Weibull distribution. Finally, the matching scores of subsets C_r with a cardinality less than 10 or with a standard deviation less than 10^{-3} were replaced by NaN.

TABLE I: Summary of results for Experiment 1. The values are reported in the format: mean (standard deviation).

Modality	Method	Rank-1 (%)	AUC (%)
Face	Matching Scores	89.70 (2.31)	96.17 (3.47)
	Z-score	89.70 (2.31)	97.76 (0.82)
	RBSN:Z-score	90.73 (1.91)	98.29 (0.63)
	W-score	89.70 (2.31)	97.73 (0.87)
	RBSN:W-score	77.43 (3.66)	98.76 (0.30)
Iris	Matching Scores	92.49 (1.42)	96.97 (0.47)
	Z-score	92.49 (1.42)	99.20 (0.27)
	RBSN:Z-score	92.13 (1.45)	99.14 (0.27)
	W-score	92.49 (1.42)	98.51 (0.40)
	RBSN:W-score	84.63 (2.87)	98.87 (0.28)
Fused	Matching Scores	93.34 (1.85)	97.32 (0.71)
	Z-score	97.14 (1.06)	99.74 (0.19)
	RBSN:Z-score	97.33 (0.90)	99.78 (0.15)
	W-score	94.55 (1.72)	99.78 (0.11)
	RBSN:W-score	95.93 (1.23)	99.85 (0.07)

Experiment 1: The objectives of this experiment are to assess the impact of score normalization methods to the recognition performance of: (i) unimodal systems with multi-sample galleries, and (ii) score fusion for multi-biometric systems with multi-sample galleries. To this end, 71 subjects were used to define a gallery and for each subject 5 samples were selected for each modality. The rest of the biometric samples were used as probes, which resulted to an open-set problem. The matching scores were normalized using Z-score, W-score, RBSN:Z-score, and RBSN:W-score. The score normalization steps were followed by score fusion in all cases. For example, the Z-score normalized scores were fused for each modality independently. To obtain the fused results across modalities a subsequent fusion step was performed by using the fused normalized scores. In all cases the mean operator was employed. This process was repeated 50 times. The performance was computed in terms of: (i) Rank-1 Identification performance for probes that are part of the gallery; and (ii) AUC for the corresponding ROC curves. An overview of the obtained results is reported in Table 1. First, we focus on the unimodal settings. As demonstrated, the Rank-1 performance for Z-score and W-score is the same as that of the unprocessed matching scores. For Z-score, RBSN improves the Rank-1 performance for face but not for iris. In terms of AUC, it appears that normalizing the scores always improves the performance. For Z-score, RBSN improves the AUC for the face but not for the iris. The comparisons of W-score with RBSN:W-score are not considered because W-score cannot be applied to multi-sample galleries and the matching scores had to be fused before they are normalized. Fusing the matching scores appears to improve the performance for both Rank-1 and AUC, as higher values compared to both the face and iris settings are obtained. Furthermore, normalizing the scores with either Z-score or W-score results in improved performance compared to the unprocessed matching scores. Finally, RBSN:Z-score yields further improvements. In other words, the evidence indicates

TABLE II: Summary of results for Experiment 2. The values are reported in the format: mean (standard deviation).

Modality	Method	Rank-1 (%)	AUC (%)
Face	Matching Scores	83.01 (2.38)	93.06 (1.09)
	Z-score	83.01 (2.38)	95.27 (1.28)
	W-score	83.01 (2.38)	95.27 (1.25)
Iris	Matching Scores	81.17 (2.32)	93.88 (0.79)
	Z-score	81.17 (2.32)	96.75 (0.71)
	W-score	81.17 (2.32)	95.75 (0.83)
Fused	Matching Scores	86.93 (1.96)	95.06 (1.04)
	Z-score	90.04 (1.64)	98.36 (0.62)
	MRBSN:Z-score	90.01 (1.72)	98.58 (0.51)
	W-score	85.58 (1.99)	98.84 (0.38)
	RBSN:W-score	85.53 (1.96)	98.53 (0.41)

that the normalized scores produced by RBSN are of better quality for the task of multi-biometric fusion.

Experiment 2: The objective of this experiment is to assess whether the proposed MRBSN framework can produce better normalized scores for the task of fusion in multi-biometric systems. To this end, 71 subjects were used to define a gallery, and for each subject one sample was selected for each modality. The remaining biometric samples were used as probes, which resulted in an open-set problem. The matching scores were normalized using Z-score, W-score, MRBSN:Z-score, and MRBSN:W-score. In all cases the mean operator was employed. This process was repeated 50 times. As in experiment 1, the performance was computed in terms of Rank-1 Identification and AUC for the corresponding ROC curves. An overview of the obtained results is reported in Table 2. For the face and iris modalities, both score normalization methods appear to improve the baseline performance. For the multi-biometric scenario, Z-score appears to improve the performance of both AUC and Rank-1. W-score appears to improve the AUC performance as well, but this is not the case for Rank-1. According to our understanding, the normalized scores produced by Step 1 of Alg. 2 are homogeneous and thus the rank-based scheme does not work as well as it does for the case of multi-sample galleries for unimodal systems. Nevertheless, we computed the Verification Performance of Z-score, MRBSN:Z-score, W-Score, and MRBSN:W-score for a fixed False Acceptance Rate (FAR) value. Specifically, we set FAR equal to 10^{-2} because for lower values W-score becomes unstable (see Sec. II). The obtained mean and standard deviation values are 90.90% (1.81%), 91.08% (1.74%), 85.46% (2.12%), and 86.29% (1.94%), respectively. The verification rate (VR) values were used to perform one-sided, non-parametric Wilcoxon Signed-Rank tests. The null hypothesis was set to H_0: the MRBSN:Z(W)-score and Z(W)-score median VRs are equal, and the alternative to H_a: the MRBSN:Z(W)-score median VR is larger than the Z(W)-score median VR. The Bonferonni correction was used to ensure that the overall statistical significance level (i.e., $a = 5\%$) is not overestimated due to the multiple tests performed. That is, the statistical significance of each individual test was set to $\frac{a}{m}$, where m is the number of tests performed (i.e., $m = 2$). The corresponding p-values obtained are $4.7{*}10^{-3}$ for Z-score and $1.5{*}10^{-9}$ for W-score, respectively. Hence, the improvements obtained appear to be statistically significantly better.

V. Conclusion

In this paper, we proposed a rank-based score normalization framework for multi-biometric score fusion. Our approach: (i) normalizes the matching score from each modality independently; (ii) joins the normalized score set; (iii) defines subsets of scores using a rank-based scheme; and (iv) normalizes the matching scores of each subset independently. A statistically significantly better verification rate was obtained for both Z-score and W-score when the proposed framework was employed. However, the Rank-1 Identification Rate and the Area Under the Curve for the corresponding ROC curves appear to be comparable with the conventional approach.

VI. Acknowledgments

The authors would like to thank Prof. Z. Sun and his students for sharing their data. Portions of the research in this paper use the CASIA-IrisV4 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA). This research was funded in part by the US Army Research Lab (W911NF-13-1-0127) and the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

References

[1] P. Moutafis and I. Kakadiaris, "Can we do better in unimodal biometric systems? a novel rank-based score normalization framework for multi-sample galleries," in *Proc. 6th IARP International Conference on Biometrics*, Madrid, Spain, June 4-7 2013.

[2] ——, "Can we do better in unimodal biometric systems? A novel rank-based score normalization framework," *Trans. on Cybernetics*, vol. 1, no. 1, pp. 1–14, 2014 (In Press).

[3] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[4] W. Scheirer, A. Rocha, R. Micheals, and T. Boult, "Robust fusion: extreme value theory for recognition score normalization," in *Proc. European Conference on Computer Vision*, vol. 6313, Crete, Greece, September 5-11 2010, pp. 481–495.

[5] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[6] G. Shakhnarovich, J. Fisher, and T. Darrell, "Face recognition from long-term observations," in *Proc. European Conference on Computer Vision*, Copenhagen, Denmark, May 27 - June 2 2002, pp. 851–865.

[7] E. Wolfstetter, *Stochastic dominance: theory and applications.* Humboldt University of Berlin, School of Business and Economics, 1993.

[8] S. Durlauf, L. Blume *et al.*, *The new palgrave dictionary of economics*. Palgrave Macmillan, 2008.

[9] M. Birnbaum, J. Patton, and M. Lott, "Evidence against rank-dependent utility theories: tests of cumulative independence, interval independence, stochastic dominance, and transitivity," *Organizational Behavior and Human Decision Processes*, vol. 77, no. 1, pp. 44–83, 1999.

[10] J. Hadar and W. Russell, "Rules for ordering uncertain prospects," *The American Economic Review*, vol. 59, no. 1, pp. pp. 25–34, 1969.

[11] Z. Sun and T. Tan. (2012, August 14, 2014) CASIA iris image database. [Online]. Available: http://biometrics.idealtest.org/