Why Rank-Level Fusion? And What is the Impact of Image Quality?

Emanuela Marasco

Lane Department of Computer Science and Electrical Engineering,
West Virginia University,
PO Box 6109 Morgantown, WV, USA
E-mail: emanuela.marasco@mail.wvu.edu

Ayman Abaza

West Virginia High Technology Consortium Foundation,
1000 Technology Drive, Suite 1000,
Fairmont, WV USA
E-mail: aabaza@wvhtf.org
Biomedical Engineering Department,
Cairo University
Giza, Egypt.

Bojan Cukic

Department of Computer Science,
University of North Carolina at Charlotte,
9201 University City Blvd Charlotte,
NC 28223-0001 USA
E-mail: bcukic@uncc.edu

Abstract: The goal of a biometric identification system is to determine the identity of the input probe. In order to accomplish this, a classical biometric system uses a matcher to compare the input probe data against each labeled biometric data present in the gallery database. The output is a set of similarity scores that are sorted in decreasing order and ranked. The identity of the gallery entry corresponding to the highest similarity score (or lowest rank) is associated with that of the probe. In multibiometric systems, the outputs of multiple biometric classifiers are consolidated. Such a fusion can be accomplished at the score-level or rank-level (apart from other levels of fusion). Recent research has established benefits of rank-level fusion in identification systems; however, these studies have not compared the advantages, if any, of rank-level fusion schemes over classical score-level fusion schemes. In the presence of low quality biometric data, the genuine match score is claimed to be low and expected to be an unreliable individual output. Conversely, the rank assigned to that genuine identity is believed to remain stable even when using low quality biometric data. However, to the best of our knowledge, there is not a deepen investigation on the stability of ranks.

In this paper, we analyze changes of the rank assigned to the genuine identity in multimodal scenarios in presence of low quality data. The contribution of this paper is two-fold: i) investigating the rank stability in both unimodal and multimodal biometric systems; and ii) comparing the identification performance of rank-level and score-level fusion in the presence of low quality data. The performance is evaluated using two datasets: (i) The first dataset is a subset of the database Face and Ocular Challenge Series (FOCS) collection (the Good, Bad and Ugly database), composed of three frontal faces per subject for 407 subjects. (ii) The second dataset was collected at West Virginia University, composed of rolled fingerprints for 494 subjects (70 of these 494 are low quality). Results show that a variant of the highest rank fusion scheme, which is robust to ties, performs better than the other non-learning based rank-level fusion methods explored in this work. However, experiments demonstrate that score-level fusion yields better identification accuracy than existing rank-level fusion schemes.

Keywords: Multi-biometrics; Rank-Level Fusion; Biometric Image Quality.

1 Introduction

Information fusion in biometrics entails the combination of different sources of evidence and, it has been extensively shown that it is able to enhance the recognition accuracy compared to a biometric system which exploits only a single modality [1] [2] [3]. This is required when dealing with low quality biometric images which remain a challenge [4] [5]. Designing such a system requires the implementation of an effective fusion scheme to integrate the evidence [6] [7]. Evidence can be integrated *before matching*, at sensor or feature level; or, *after matching* at decision, rank or score level. While the amount of information to integrate progressively decreases from the sensor-level to the decision-level, the degree of noise also decreases [8] [9]. This work focuses on fusion at the score-level and rank-level, which are described below:

Score-level fusion. Match scores output by different biometric matchers are fused. This approach has been widely used since match scores are easy to access and combine. However, match scores output by different biometric matchers may not be homogeneous: each matcher can conform to different scales and they may not have the same interpretation across different matchers (they can be distances, confidences, etc.) [10]. Thus, before integration, each matcher may have to be transformed into a common domain via an effective normalization scheme [11]. A fusion rule which is commonly used in the literature and that is employed in this work is the *simple mean* formulated in Eqn. (1), where s_k is the match score output by the k^{th} matcher.

$$S_{mean} = (\sum_{k=1}^{K} s_k)/K \tag{1}$$

Rank-level fusion. Ranks describe the relative order of the gallery identities, and they carry less information than the true values of match scores, as nothing is retained of the notion of distance (or similarity) between the probe and each gallery. Ranks can be referred to as *ordinal* variables since they only carry information about the relative ordering of the different identities. There are cases where the information about how the different identities are ranked can be useful. First, match scores may be not available for those systems that output only a list of candidate identities [12]. Second, when conducting statistical parametric tests, distributions of match scores are assumed to be normal [13]. These tests may be heavily sensitive to the normality assumption and fail when the considered distributions are not normal. Further, in cases where monotonous transformation are applied to match scores, the corresponding ranks are kept unchanged. Ranks do not change when the scale on which the corresponding numerical measurements changes [14]. Finally, as stated in the introduction, when combining multiple modalities, the fusion of ranks does not require a normalization phase as typically needed with heterogeneous match scores. Each matcher ranks the identities in the gallery based on the match scores between the input probe and these gallery identities [15] [16]. Let $\mathbf{R}$=$[r_{ij}]$ be the rank matrix in a multi-biometric system where r_{ij} is the rank assigned to the identity I_i by the j^{th} matcher, i=1…N and j=1…K. A reordered statistic r_i is computed for each user I_i such that the highest consensus rank is assigned to the user with the lowest value of r.

The rank assigned to the genuine identity is expected to remain stable even in the presence of low quality biometric data [17]. However, this statement has been argued but not experimentally demonstrated. The main contribution of this paper, is to analyze the robustness of rank level and score level fusion schemes in presence of low quality data.

This paper is organized as follows: Section 2 discusses benefits and drawbacks of ranks. Section 3 presents the approaches for fusion at rank level used to conduct this study. Section 4 describes the technique adopted to synthetically degrade the quality of the fingerprint images and the actual low quality face samples used in our experiments. Section 5 reports results and Section 6 summarizes the conclusions of this work.

2 Rank Information Related Works

Several works have focused on the problem of enhancing the performance of rank level fusion schemes in adverse operational environments (i.e., noise input data, etc.).

Monwar and Gavrilova presented a Markov chain approach for combining rank information in multimodal biometric systems comprising face, ear and iris [18]. Their experiments showed the superiority in accuracy and reliability over other biometric rank aggregation methods. They reported a rank-1 multimodal identification accuracy of 98.5% compared to the unimodal accuracies of 87%, 92% and 94% for ear, face and iris respectively. However, this improvement may be due to the presence of the iris modality. Later, the same authors combined face, ear and signature modalities using Borda Count and Logistic Regression [19]. First, they reported rank-1 identification accuracy is 87.03% when applying the Borda Count and 90.24% when applying the Logistic Regression fusion rule [20]. Second, they integrated results of iris, ear and face unimodal biometric matchers. The best performance was achieved by the Logistic Regression with an accuracy of 98.8%.

Abaza and Ross proposed a quality-based Borda Count scheme that is able to increase the robustness of the traditional Borda Count in the presence of low quality images without requiring a training phase [21]. Marasco *et al.* proposed a predictor-based approach to perform a reliable fusion at rank level. In such a scheme, a predictor (classifier) was trained using both rank and match score information for each modality and designed to operate before fusion [22]. In order to evaluate the robustness of ranks and scores in the presence of low quality data, Marasco *et al.* [23] [24] introduced a concept of rank stability. We will apply this method for experiments in this paper.

3 Usage of Ranks to Combine Multiple Evidence

The existing approaches for fusion at rank level can be categorized as *learning-based* schemes or *non learning-based* schemes, see Table 1. Methods which require a learning phase may be biased to a specific training set; while results on other sets may change dramatically. Thus, the evaluation is conducted by considering approaches that combine ranks without learning. However, in order to observe the impact of the choice of the training set on the performance on the test set, we implement one *learning-based* fusion rule.

In multi-biometric systems, rank-level fusion combines different biometric systems. Scores output by each matcher are arranged in a descending order to form the ranking list of matching identities. Let K be the number of modality matchers to be combined and N be the number of enrolled users. Let r_{ij} be the rank assigned to the j^{th} identity in the gallery by the i^{th} matcher, $i = 1 \dots K$, and $j = 1 \dots N$.

Traditional Highest Rank. This method obtains consensus ranking by sorting the identities according to their highest rank as follows:

$$R_i = \min_{k=1}^{K} r_{ik}, \quad i = 1, 2, ...N \tag{2}$$

The advantage of this method lies in utilizing the best of the combined matchers. However, this method may lead to one or multiple ties. This negatively impacts the reliability of the final decision [33]. This drawback can be effectively addressed by applying a variant of the traditional highest rank fusion rule, referred to as *Modified Highest Rank*, and formulated below [21]:

$$R_i = \min_{k=1}^{K} r_{ik} + \epsilon_i, \quad i = 1, 2, ...N \tag{3}$$

where

$$\epsilon_i = \sum_{k=1}^{K} r_{ik} \tag{4}$$

In case of fusing two classifiers, an example of how ties are resolved is as follows: the ranks for the true match user j = 1 $r_{11} = 1$ and $r_{21} = 2$, while for another user j = 2 $r_{12} = 3$ $r_{22} = 1$. According to equation 1, $R_1 = 1$ and $R_2 = 1$ result in a tie. The result will be the same even if r_{12} changes from 3 to 30 or 300. Returning to the above mentioned example, and assuming K = 100, $R_1 = 1 + 3/100$ and $R_2 = 1 + 4/100$. In generating the fused rank, R_1 (the true match) will have a lower rank than R_2. The criteria of selecting the lowest rank is maintained and the *epsilon* factor is just used to break the ties.

Traditional Borda Count. In this method the consensus ranking is obtained by summing the ranks assigned by the individual matchers, see Eqn. (5). This approach is highly vulnerable to the effect of weak matchers since it assumes that all the matchers present the same classification capability. This method assumes that the ranks assigned to the identities in the gallery are statistically independent and that the performance of all the matchers are uniform [34] [35].

$$R_i = \sum_{k=1}^{K} r_{ik}, \quad i = 1, 2, ...N \tag{5}$$

Quality-based Borda Count. This method is a redefinition of the traditional Borda Count where the input data quality is incorporated in the fusion scheme as follows:

$$R_i = \sum_{k=1}^{K} Q_{ik} r_{ik}, \quad i = 1, 2, ...N \tag{6}$$

Q_{ik} is defined as $Q_{ik} = min(Q_i, Q_k)$, where Q_i and Q_k are quality of the probe and the gallery data, respectively [21]. Adding weights, to the matcher outputs, leads to a reduction of the effect of poor quality biometric samples.

Logistic Regression. In this method, the fused rank is calculated as a weighted sum of the individual ranks [6], and is defined in the following equation:

$$R_i = \sum_{k=1}^{K} w_{ik} r_{ik}, \quad i = 1, 2, ...N \tag{7}$$

In order to combine modality matchers with non-uniform performances, the ranks produced by each modality matcher should be appropriately weighted. The weights reflect the relative significance of the corresponding unimodal output [30]. The weight w assigned to the different matchers is determined in the training phase of Logistic Regression. This fusion rule is robust in the presence of matchers with significant differences in performance; however, its main drawback lies in requiring a learning phase to determine the weights.

4 Input Image Quality

The biometric acquisition process can be affected by several factors such as noise of the biometric device, the interaction between the user and the device, external factors due to the environment and the intrinsic variation of biometric. These factors generally lead to degradation in image quality that significantly impacts the matching accuracy [36]. Several studies have shown that degradation in image quality leads to a reduction of the accuracy of a biometric system [37].

A recent research trend focuses on reducing the adverse effect of the low quality data on the performance. In the context of a biometric system, the term quality refers to the utility of the acquired image to an automated system and an image is considered to have a

Table 1 Approaches for rank level fusion

Learning-based Methods	Non Learning-based Methods
Markov Chain [25], [18]	Borda Count [26], [27]
Logistic Regression [6], [28]	Quality-based Borda Count [21]
Bayesian Approach [29]	Traditional Highest Rank [30], [15]
Weighted Borda Count [31], [32]	Modified Highest Rank [21]
Non-linear Weighted Ranks [31], [32]	
Bucklin Majority Voting [31]	

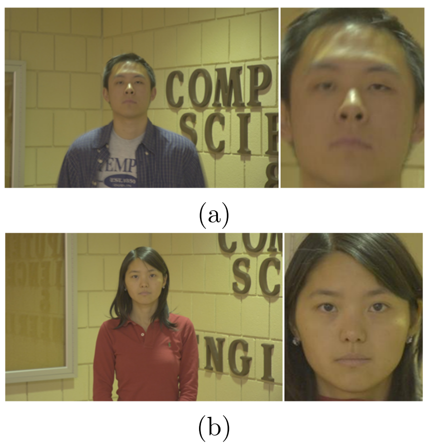


Figure 1: Examples of face images taken from the Face and Ocular Challenge Series (FOCS) collection (the Good, Bad and Ugly database): a) sample image from the Ugly partition; b) sample image from the Good partition.

good quality if is suitable for automated matching [5]. In this work, we consider sources of noisy input data that may arise during the image capture where the image quality can be impacted for example by an incorrect presentation of the biometric sample to the system [38]. For this purpose, we use real data for face modality and synthesized data for fingerprint.

4.1 Low Quality Face Images

Images from the Face and Ocular Challenge Series (FOCS) database are of frontal faces taken under uncontrolled illumination, both indoors and outdoors. The partitions of interest are referred to as *Good* and *Ugly*, that have an average identification accuracy of 0.98 and 0.15 respectively [`http://www.nist.gov/itl/iad/ig/focs.cfm`]. Figure 1 shows examples of real low quality face images.

4.2 Synthetic Fingerprint Image Quality Degradation

The study conducted in this paper involves the analysis of how poor quality fingerprint images impact the accuracy of both rank and score level fusion. This performance can be estimated on low quality images which do not have sufficient ridge details to execute a reliable matching process. The required degradation effect can be simulated by adopting a gray-scale saturation technique which converts fingerprint pixels corresponding to the ridges into background pixels [21]. The gray-scale saturation level (SL) indicates the gray level value above which pixels are saturated to white (255). Figure 2 shows examples of low quality fingerprint images obtained by decreasing the saturation level.

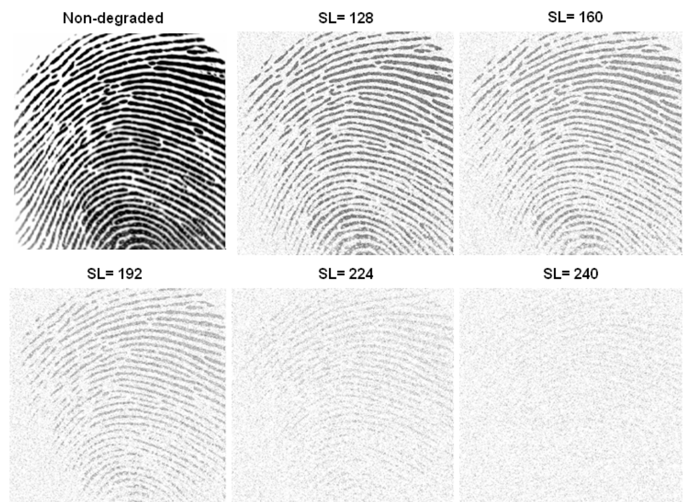


Figure 2: Examples of low quality fingerprint images artificially degraded by using five different noise saturation levels ST = [128, 160, 192, 224, 240].

5 Experimental Results

This section starts by discussing the database, demonstrates various experiments to compare the performance of score and rank level fusion schemes using low quality input data, and discusses the results of these experiments.

5.1 Datasets

Evaluation was carried out on three different data sets. The first data set is a subset of a database collected at West Virginia University (WVU). In particular, images used in our experiments pertain to fingerprints of left thumb [FL1], right thumb [FR1], left index [FL2], right index [FR2]). Fingerprints were collected using an optical scanner, without explicitly controlling the quality [39]. The entire dataset was divided into five sets:

the first sample of each identity was used to compose the *gallery* and the remaining four samples of each identity were used as *probes* (P_1, P_2, P_3, P_4). VeriFinger [`http://www.neurotechnology.com/verifinger.html`] software was used for generating the fingerprint scores. Match scores were generated by comparing high quality gallery versus probe image degraded at different levels. Figure 3 illustrates the decrease in matching performance of the unimodal system when the quality of the probe image at different quality levels. Quality measures were extracted using the IQF software, developed by MITRE [`http://www.mitre.org/tech/mtf/`]. This quality factor (Q) ranges from 0 to 100, with 0 being the lowest and 100 being the highest quality.

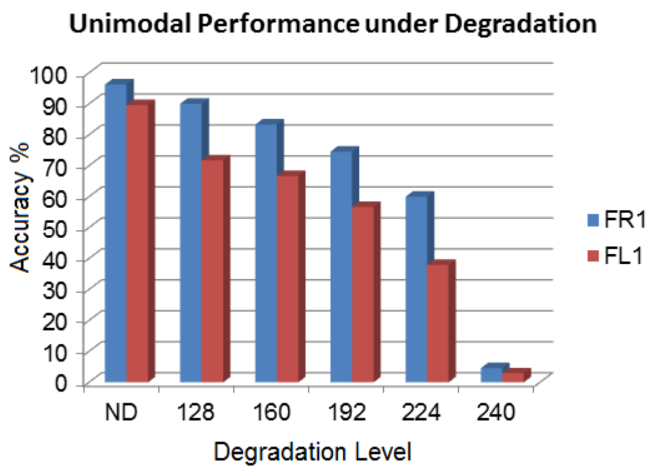


Figure 3: The unimodal performance decreases when degrading the quality of the probe image at different levels. ND indicates the case with no degradation.

The second data set is a subset of the Face and Ocular Challenge Series (FOCS) collection (the Good, Bad and Ugly database). Frontal face images taken under uncontrolled illumination, both indoors and outdoors are considered. The partitions of interest are referred to as *Good* and *Ugly*, that have an average identification accuracy of 0.98 and 0.15, respectively [`http://www.nist.gov/itl/iad/ig/focs.cfm`]. The used dataset composed of 407 subjects, three frontal instances of faces:

- two high quality images (from the Good dataset)
- one actual low quality image (from the Ugly dataset)

PittPatt [`http://www.pittpatt.com/`] software was used for generating the face match scores. These results are generated from two different matching scenarios: both the gallery and probe are not degraded, referred to as *Good-Good* and the gallery is not degraded, but the probe is degraded, referred to as *Good-Ugly*.

The third dataset, collected at West Virginia University, consists of fingerprint images from 494 users. They were acquired using a live-scan optical device. Users provided two sets of fingerprints, in sequence. Fingerprints were collected without controlling the quality in acquisition. Match scores were extracted employing Identix BioEngine SDK. Fingerprint image quality is estimated by the IQF software by MITRE [40]. Figure 4 shows examples of high and low quality images described above.

5.2 *Experiments*

We conducted several experiments to evaluate the stability of ranks in challenging unimodal scenarios where the quality of the probe image decreases. We experimentally analyzed the effectiveness of a fusion scheme at rank level when the quality of the probe image of a subset of the combined modalities is low.

Figures 5 a) and b) illustrate the histograms of rank values assigned to the genuine identity, for various probes. The original image, with no quality degradation, corresponds to a scenario with high quality input images; while the degradation in probe image quality, using different saturation levels, corresponds to a scenario with low quality probe images (a higher SL value causes a lower quality of the image). In case of no degradation, 224 out of 250 are the identities appropriately identified as being at rank-1. The number of identification errors increases when increasing the degradation level of the probe image. The rank-1 identification rate decreases. For SL= 128 the number of rank-1 identifications is 173 out of 250.

Figures 5 c) and d) represent the histograms of the match score value of the genuine identity using high and low quality probe images. Figures 6 a) and b) represent the histograms of rank values assigned to the genuine identity for the face modality using probe of high quality and low quality. Figures 6 c) and d) represent the histograms of the match score value of the genuine identity for high and low quality of the face probe image. The histograms of the difference between the rank (and the score) assigned to the genuine identity in the presence of a high and low quality probe image can be visualized in Figures 7 and 8.

5.2.1 *Comparison of Score and Rank level Fusion*

In this section, we report results obtained when integrating ranks in multimodal biometric systems, and compare them to the performance achieved using scores. Figure 10 shows the accuracy achieved by rank and score level fusion schemes involving four fingerprint modalities under quality degradation of one fingerprint image. The modified highest rank exhibits the best robustness in the presence of one degraded image. It achieves a rank-1 identification rate of 97.08% when the noise saturation level applied to one fingerprint image in every pair is 128 and 85.00% when increasing the noise saturation level to 240. However, the performance of the score sum exceeds that obtained by rank level fusion by achieving a rank identification rate of 99.17% in both non-degraded and degraded conditions. Its accuracy decreases to 97.50% only when the noise saturation level is 240.

Figure 9 shows the accuracy achieved by rank and score level fusion schemes when combining four fingerprint modalities using low quality probes of two fingerprints. Also in this scenario, in which the number of the degraded images has been doubled, the modified highest rank exhibits the best performance among the considered rank level fusion schemes; the achieved rank identification rate decreases from 92.08% to 57.5% when increasing the degradation factor. Regarding the rank identification rate obtained by the score sum, it decreases from 99.17% to 86.67% when increasing the degradation factor.

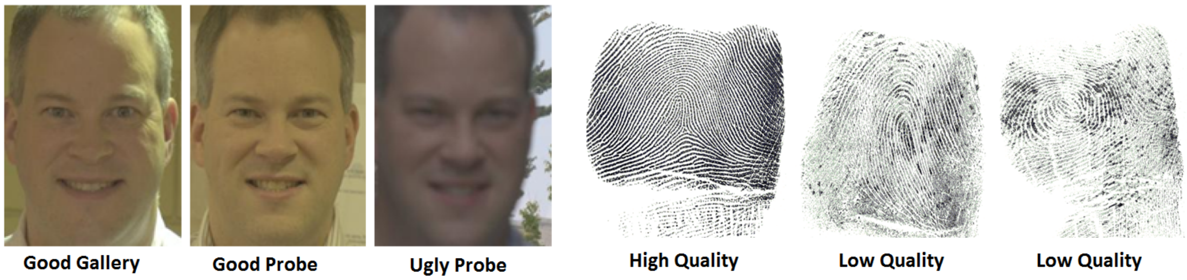
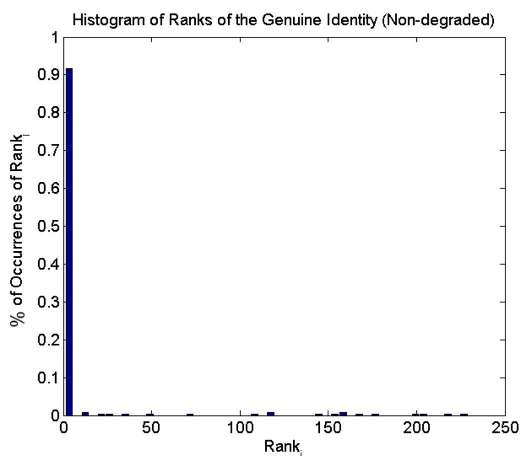
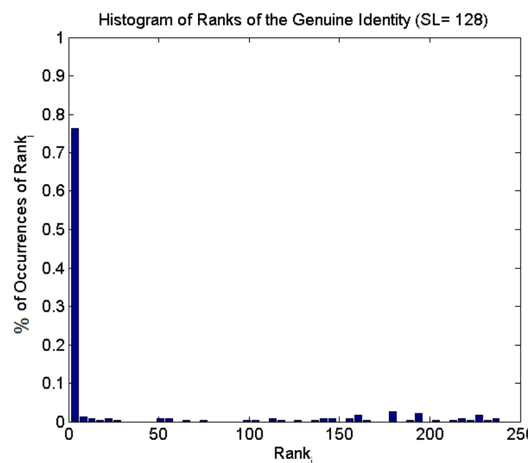


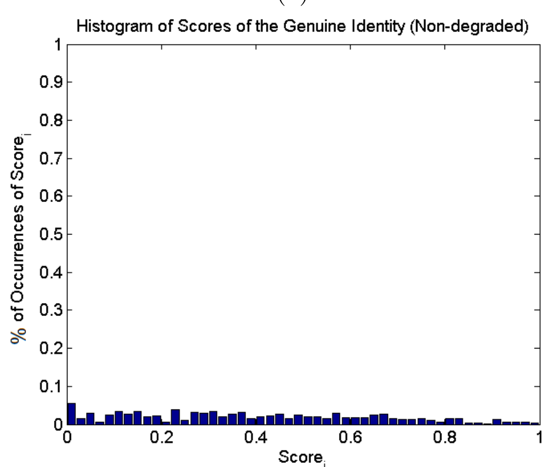
Figure 4: Examples of a high quality (Good) and low quality (Ugly) face images taken from the FOCS collection; and, high and low quality fingerprint taken from the WVU database.



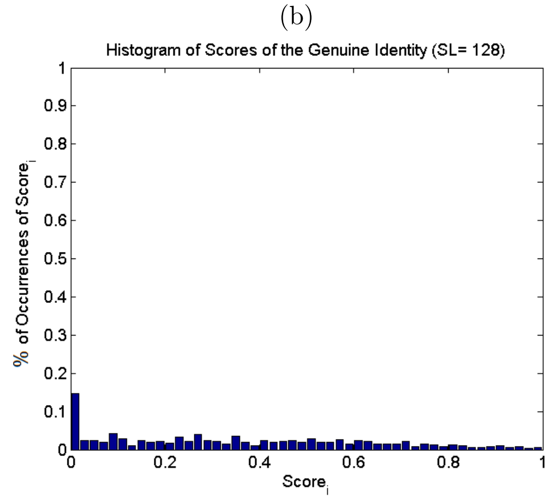
(a)



(b)

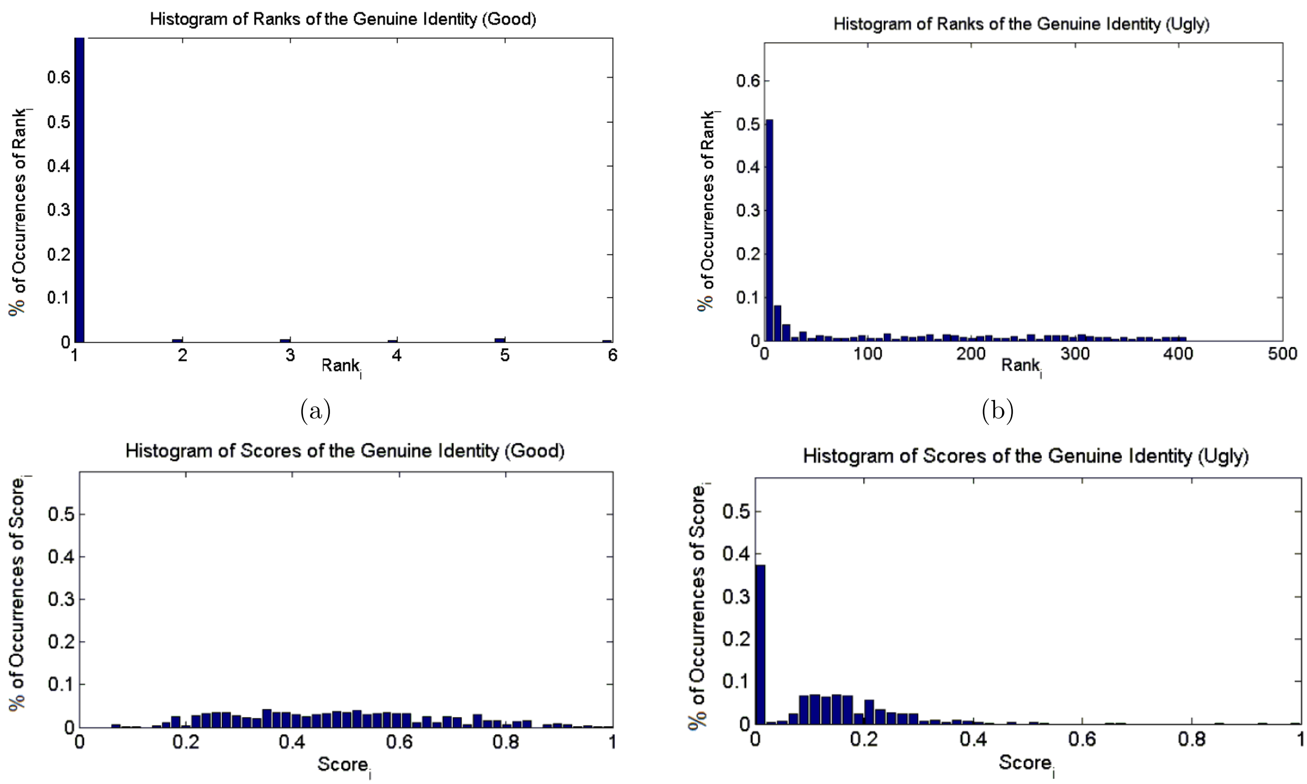


(c)



(d)

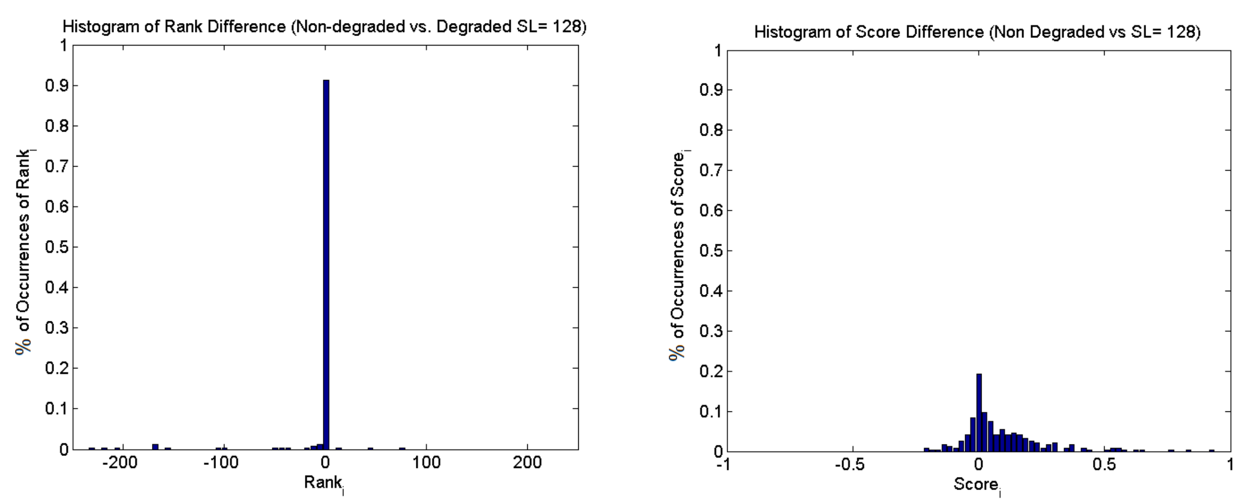
Figure 5: Row 1. Histograms of ranks assigned to the genuine identity in the gallery for the fingerprint modality FL1 taken from WVU database analyzed under different levels of degradation; (a) non degradation; (b) degradation with SL= 128. Row 2. Histograms of match scores of the genuine identity for the fingerprint modality FL1 taken from WVU database analyzed under different levels of degradation: (c) non degradation; (d) degradation with SL= 128.



(a) (b)

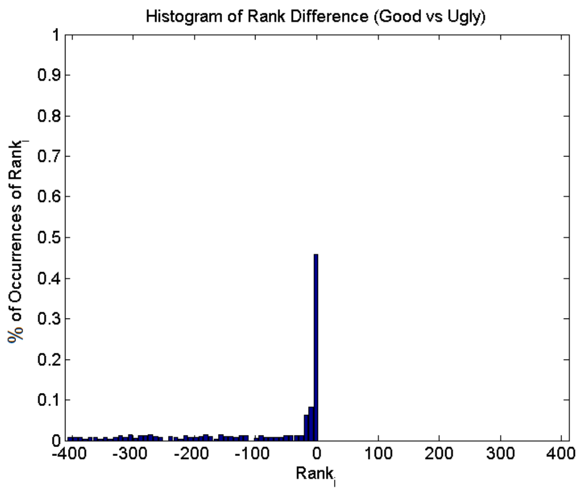
(c) (d)

Figure 6: Row 1. Histograms of ranks assigned to the genuine identity for the face modality taken from Good Ugly real database; a) non degradation (Good); b) degradation (Ugly). Row 2. Histograms of match scores of the genuine identity for the face modality taken from Good Ugly real database; c) non degradation (Good); d) degradation (Ugly).

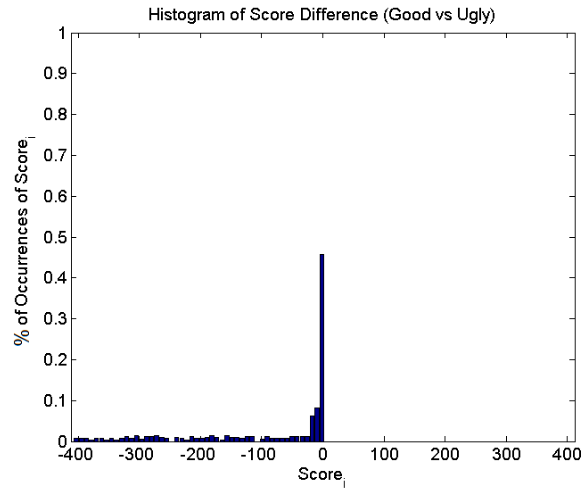


(a) (b)

Figure 7: Histograms of the difference between the rank (and the score) assigned to the genuine identity in the gallery before and after degradation of the probe image: a) Rank difference: Non Degraded vs degradation with SL= 128; b) Score difference: Non Degraded vs degradation with SL= 128.



(a)



(b)

Figure 8: Histograms of the difference between the rank (and the score) assigned to the genuine identity before and after degradation of the probe image: a) Rank difference: Good vs Ugly; b) Score difference: Good vs Ugly.

Further, a worst case scenario where all the fused modalities are of low quality, is analyzed. For this experiment, we used two low quality fingerprints (synthesized data). Figure 11 shows the change in performance of the fusion schemes in such a scenario. The traditional highest rank exhibits the best accuracy as rank level fusion approach by achieving 80.83% when the noise saturation level of both fingerprint images is 128 and 2.50% when increasing it to 240. The performance of the score sum varies from 93.75% when the noise saturation level of both fingerprint images is 128 and 6.67% when increasing it to 240.

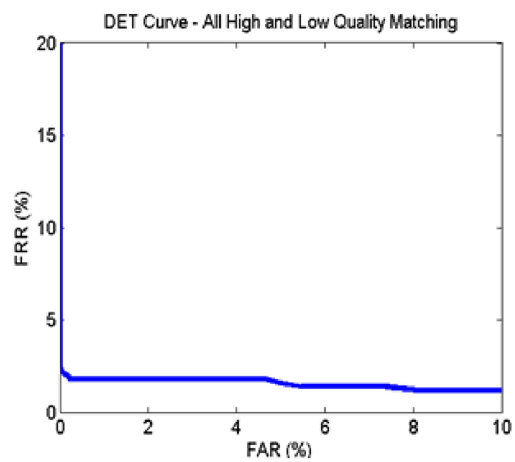
5.3 Discussion

The Detection Error Tradeoff (DET) curves pertaining to the WVU dataset are reported in Figure 12. Figure 6 shows histograms of the match score value of the genuine identity for high and low quality face probes, compared to the rank value.

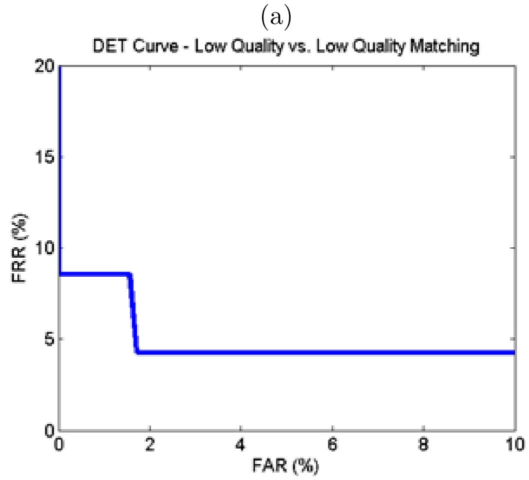
Ranks (scores) are stable if the rank (score) assigned to the genuine identity when using high quality probes does not change when using low quality probes. A difference in ranks (scores) between high and low quality equal to zero indicates stability. Visually, the distributions of the differences of ranks (and match scores) suggest that ranks are more stable than match scores.

We report results obtained when integrating ranks in multimodal biometric systems, and compare them to the performance achieved using scores.

For rank level fusion, the best rank-1 identification accuracy is achieved by the Modified Highest Rank method. For the traditional Borda Count fusion scheme, it is sufficient to combine only one individual output with high rank assigned to the genuine identity to have an incorrect identification. Even assigning low weights to the matchers with low quality input images, the multimodal identification is incorrect. When applying the Highest Rank rule a more robust accuracy is obtained since such a scheme requires that only one of the combined matchers assigns rank-1 to the genuine identity. Errors due to the ties are solved with its modified version. In a scenario where all the combined



(a)



(b)

Figure 12: Performance of the fingerprint system a) with both high and low quality images; b) with low quality images only.

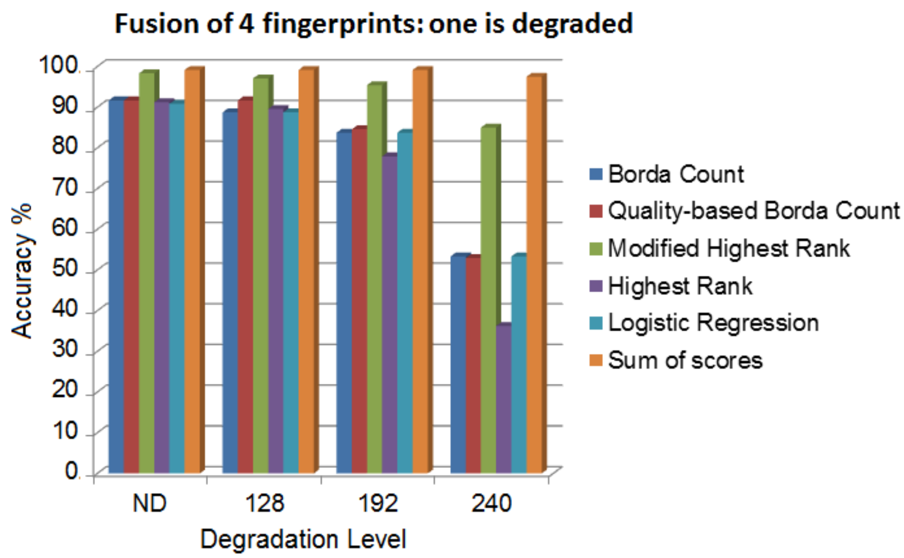


Figure 9: Fusion of four fingerprints when one of them is degraded: change in performance of different schemes at rank and score under different degradation levels.

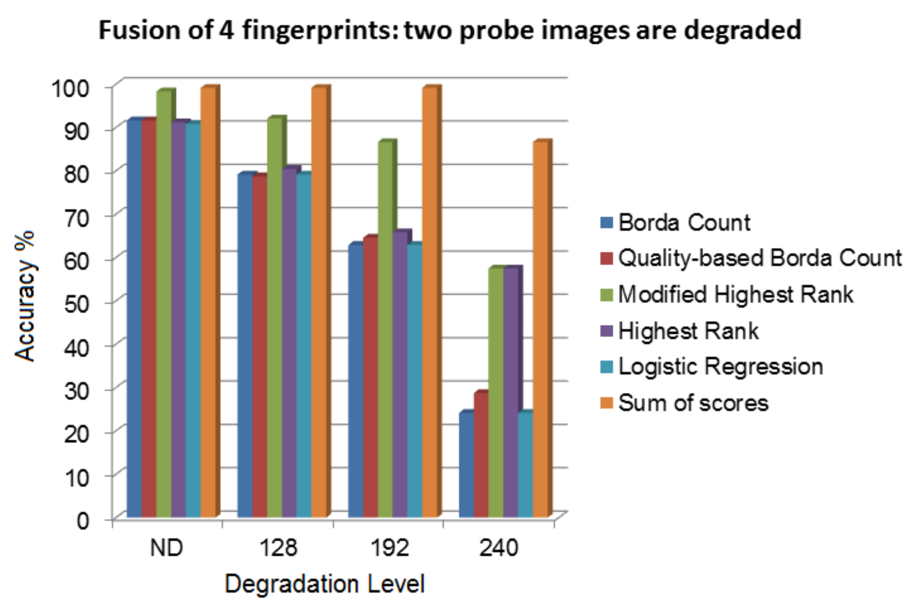


Figure 10: Fusion of four fingerprints when two of them are degraded: change in performance of different schemes at rank and score under different levels of degradation of two fingerprint probe images.

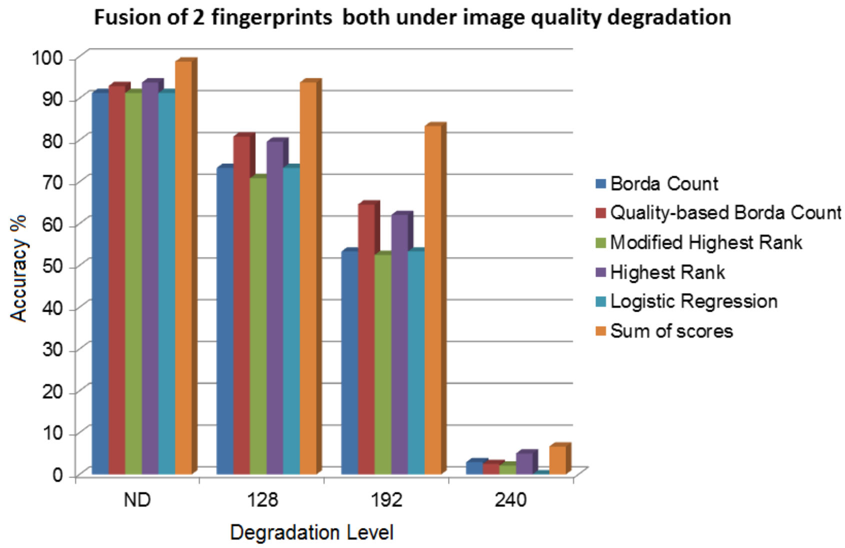


Figure 11: Fusion of two fingerprints both degraded: change in performance of different schemes at rank and score when combining two fingerprints both under different degradation levels.

modalities are using low quality probes, the Quality-based Borda Count is the most effective fusion scheme. However, fusion at score level outperforms fusion at rank level in all the considered scenarios, i.e., when only a subset or all the combined modality are low quality.

Regarding the Logistic Regression scheme, the weights have been determined using only high quality probes. When testing in a scenario where one probe out of four is of low quality, the accuracy decreases from 90.83% to 88.75% with SL= 128, and to 53.33% with SL= 240. When testing in a scenario where two probes out of four is of low quality, the accuracy decreases from 90.83% to 79.17% with SL= 128, and to 24.17% with SL= 240. Finally, when fusing two modalities and both probes are highly degraded, the accuracy is 0.03% which corresponds to the worst result obtained among the fusion schemes considered in this work.

Figure 13 illustrates face identification performance in the presence of high quality and compares it when the probe is low quality. In unimodal scenario the accuracy significantly decreases, while when fusing low quality with high quality probe sum of scores is able to achieve 100% of accuracy. The best fusion scheme at rank-level is Modified Highest Rank. Figure 14 illustrates the performance when combining low quality face with low quality fingerprint. Sum of scores and Modified Highest Rank are again the most accurate schemes.

6 Conclusions

This study carried out an investigation regarding the stability of the rank in the context of biometrics. Further, we analyzed different non learning-based rank level fusion schemes in the presence of both synthetically degraded fingerprint images and actual low quality face images. The experiments showed that rank is stable when the degradation level of the low quality image is not very significant. When the level of degradation is significant, both ranks and scores are not stable. Further, ranks are more stable than scores since they present a higher rank correlation coefficient value. (However, the performed study may be dependent upon the matcher used).

Conditions under which it is reasonable to use ranks can be expressed as follows:

- When match scores are not available, fusing ranks by applying the modified highest rank scheme leads to the best identification accuracy.
- When match scores are available, a better identification accuracy can be obtained by employing score level fusion.

7 Acknowledgments

The authors are grateful to West Virginia University and the Center for Identification Technology Research (CITeR). This material is supported by ONR under Contract No. N00014-09-C-0388 awarded to West Virginia High Technology Consortium Foundation.

References

[1] J. Kittler, Y. P. Li, J. Matas, and M. U. R. Sanchez. Combining evidence in multimodal personal identity recognition systems. *International Conference on Audio- and Video-based Biometric Person Authentication*, 1997.

[2] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain. Incorporating image quality in multi-algorithm fingerprint verification. *International Conference on Biometrics (ICB)*, pages 213–220, January 2006.

[3] J. Kittler and N. Poh. Multibiometrics for identity authentication: Issues, benefits and challenges. *IEEE Transaction On Pattern Analysis and Machine Intelligence*, 2009.

[4] M. Monwar and M. Gavrilova. Multimodal biometric system using rank-level fusion approach. *IEEE*

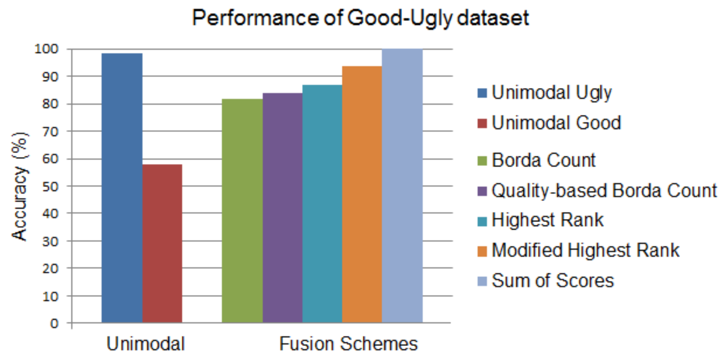


Figure 13: Fusion of one face of high quality and one face of low quality. The face modality is taken from the GBU data set of the FOCS database.

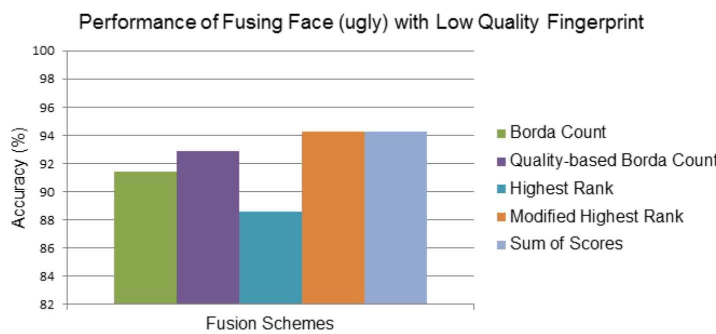


Figure 14: Fusion of one face of low quality with on fingerprint of low quality. The face modality is taken from the GBU data set of the FOCS database, fingerprints are taken from the WVU database.

Transactions on Systems, Man, and Cybernetics, 39:867–878, August 2009.

[5] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Transaction On Pattern Analysis and Machine Intelligence*, 29(4):531–543, 2007.

[6] A. Ross, K. Nandakumar, and A. Jain. *Handbook of MultiBiometrics*. Springer, 2006.

[7] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[8] A. Ross, A. Jain, and K. Nandakumar. *Introduction to Biometrics: A Textbook*. Springer, 2011.

[9] A. Ross and A. Jain. Multimodal biometrics: an overview. *12th European Signal Processing Conference (EUSIPCO)*, pages 1221–1224, 2004.

[10] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005.

[11] W. Scheirer, A. Rocha, R. Micheals, and T. Boult. Robust fusion: extreme value theory for recognition score normalization. *The 11th European Conf. Computer Vision*, pages 481–495, 2010.

[12] S. Labovitz. The assignment of numbers to rank order categories. *American Sociological Review*, 35(3):515–524, June 1970.

[13] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, December 1937.

[14] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, December 1945.

[15] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters 24*, pages 2115–2125, 2003.

[16] E. Marasco C. Sansone. An experimental comparison of different methods for combining biometric identification systems. *International Conference on Image Analysis and Processing (ICIAP)*, pages 255–264, 2011.

[17] S. Tulyakov and V. Govindaraju. Combining biometric scores in identification systems. pages 1–34, 2006.

[18] M. Monwar and M. Gavrilova. Markov chain model for multimodal biometric rank fusion. *Signal, Image and Video Processing*, pages 1863–1703, 2011.

[19] M. Monwar and M. Gavrilova. FES: A system for combining face, ear and signature biometrics using rank level fusion. *Fifth International Conference on Information Technology: New Generations, ITNG*, pages 922–927, 2008.

[20] M. Monwar and M. Gavrilova. Integrating monomodal biometric matchers through logistic regression rank aggregation approach. pages 1–7, 2008.

[21] A. Abaza and A. Ross. Quality-based rank level fusion in biometrics. *Third IEEE International Conference on Biometrics: Theory, Applications and Systems*, September 2009.

[22] E. Marasco, A. Ross, and C.Sansone. Predicting identification errors in a multibiometric system based on ranks and scores. *Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems*, September 2010.

[23] E. Marasco and A. Abaza. On the stability of ranks to low image quality in biometric identification systems. *Lecture Notes in Computer Science*, 8156:260–269, 2013.

[24] E. Marasco, A. Abaza, L. Lugini, and B. Cukic. Impact of biometric data quality on rank-level fusion schemes. *Algorithms and Architectures for Parallel Processing*, pages 209–216, 2013.

[25] M. Monwar and M. Gavrilova. Robust multimodal biometric system using markov chain based rank level fusion. pages 458–463, 2010.

[26] A. Jain, S. Prabhakar, and S. Chen. Combining multiple matchers for a high security fingerprint verification system. *Pattern Recognition Letters*, 20:1371–1379, 1999.

[27] J. Benediktsson and P. Swain. Consensus theoretic classification methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4):688–704, 1992.

[28] F. Roli and G. Fumera. Analysis of linear and order statistics combiners for fusion of imbalanced classifiers. pages 252–261, 2002.

[29] K. Nandakumar, A. Jain, and A. Ross. Fusion in multibiometric identification systems: What about the missing data? *3rd International Conference on Biometrics ICB*, June 2009.

[30] T. Ho, J. Hull, and S. Srihari. Combination of decisions by multiple classifiers. *Structured Document Image Analysis*, pages 188–202, 1992.

[31] A. Kumar and S. Shekhar. Personal identification using multibiometrics rank-level fusion. *IEEE Transactions on Systems, Man, and CyberneticsPart C: Applications and Reviews*, 41(5):743–752, September 2011.

[32] N. Mittal, M. Hanmandlu, J. Grover, and R. Vijay. Rank-level fusion of multispectral palmprints. *International Journal of Computer Applications*, 38(2):33–37, January 2012. Published by Foundation of Computer Science, New York, USA.

[33] F. Moscato, V. Vittorini, F. Amato, A. Mazzeo, and N. Mazzocca. Solution workflows for model-based analysis of complex systems. *IEEE Transactions on Automation Science and Engineering*, 9(1):83–95, 2012.

[34] E. Marasco. *Secure Multibiometric Systems*. PhD thesis, Università degli studi di Napoli Federico II, 2010.

[35] M. van Erp and L. Schomaker. Variants of the borda count method for combining ranked classifier hypotheses. *The 7th International Workshop on Frontiers in Handwriting Recognition*, pages 443–452, 2000.

[36] N. Poh and J. Kittler. A unified framework for biometric expert fusion incorporating quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):3–18, 2012.

[37] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Quality-based score level fusion in multibiometric systems. *International Conference on Pattern Recognition (ICPR)*, 4:473476, August 2006.

[38] N. Damera-Venkata, T. Kite, W. Geisler, and B. Evans. Image quality assessment based on a degradation model. *IEEE Transactions On Image Processing*, 9(4):636–650, April 2000.

[39] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak. A protocol for multibiometric data acquisition, storage and dissemination. *Technical Report, West Virginia University*, 2007.

[40] N. Nill. Mitre technical report IQF (Image Quality of Fingerprint) software application. 2007.