

# QPLC: A Novel Multimodal Biometric Score Fusion Method

**Jayanta Basak**

*NetApp Advanced Technology Group  
Bangalore, India*

*basakjayanta@yahoo.com*

**Kiran Kate**

*IBM Research, Singapore*

*kiran.kate@gmail.com*

**Vivek Tyagi**

*IBM Research, New Delhi, India*

*vivetyag@in.ibm.com*

**Nalini Ratha**

*IBM T J Watson Research Center  
Hawthorne, USA*

*ratha@us.ibm.com*

---

## Abstract

In biometrics authentication systems, it has been shown that fusion of more than one modality (e.g., face and finger) and fusion of more than one classifier (two different algorithms) can improve the system performance. Often a score level fusion is adopted as this approach doesn't require the vendors to reveal much about their algorithms and features. Many score level transformations have been proposed in the literature to normalize the scores which enable fusion of more than one classifier. In this paper, we propose a novel score level transformation technique that helps in fusion of multiple classifiers. The method is based on two components: quantile transform of the genuine and impostor score distributions and a power transform which further changes the score distribution to help linear classification. After the scores are normalized using the novel quantile power transform, several linear classifiers are proposed to fuse the scores of multiple classifiers. Using the NIST BSSR-1 dataset, we have shown that the results obtained by the proposed method far exceed the results published so far in the literature.

---

## 1. INTRODUCTION

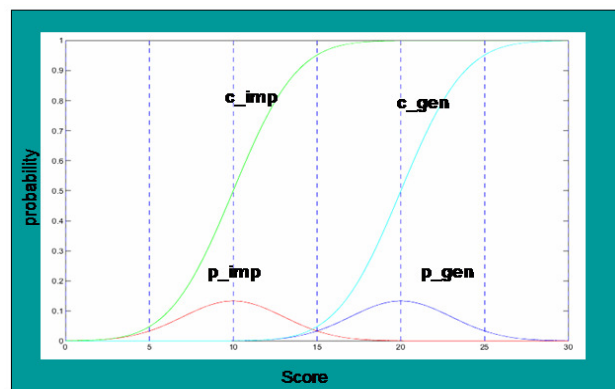
Biometrics-based authentication systems have been shown to be extremely useful in many security applications because of the non-repudiation functionality. However, these systems suffer from many shortcomings: the errors associated with the biometrics such as the false accept rate and false reject rate can impact the performance of the system; the failure to acquire and failure to enroll error rates can also impact the coverage of the population; fake biometrics e.g., latex fingers, face masks etc. can be used to fool biometrics systems. In order to overcome these problems, multi-biometrics systems have been proposed which is also known as biometric fusion. The fusion can be at various levels: signal (data), features, and classifiers. Several examples of biometric fusion methods have been reported in the literature. Fusion could involve more than one biometrics modality such as finger and face; involve more than one classifier e.g., face with two different matchers; involve more than one sample of a biometrics e.g., two samples of the same finger; involve more than one sensing modality in a particular mode e.g., face acquisition using infra red imaging and regular color cameras. Each method of fusion described above would have some advantage over a unimodal system.

The biometrics fusion problem is very interesting problem from a research and practical use perspective. The general area of fusion in the computer vision community has been studied extensively while its application to biometrics has been a relatively recent phenomenon. Early

research in this area dealt with decision level fusion using majority, and, or rules. While only few papers have appeared in the area of feature level fusion, the score-level fusion has received considerable attention in the literature. In order for feature-level fusion to work, the description of the features used in the underlying unimodal biometrics system needs to be reported. Many commercial vendor-based systems aren't comfortable with this. While the impact of a pure decision is limited, the feature level fusion looks hard because of non-standard features used in commercial systems. The score level fusion has been proposed as the optimal level as most vendors produce a score from a biometrics template pair matching. The score is available for making a final decision. The only challenge in a score level fusion has been score normalization. Even within the same mode (e.g., face), every matcher provides a score within its own range and interpretation. Many score normalization methods have been proposed before the standard sum rule or other simple fusion rules can be applied [11, 15].

In this paper, we propose a novel method of score transformation before the classifiers can fuse them. Many score normalization methods depend on the range of the scores produced by the classifiers. Even small change in the scores, can cause the normalization methods to vary significantly. Often the quantile transform has been used in many statistical data analysis to suppress the impact of outliers. In a biometrics system, there are two score distributions: genuine and impostor as shown in Fig. 1.

The quantile transform is applied to both the distributions. In order to improve the separability between the two distributions, we apply a non-linear transform. After the scores are normalized, we apply many special linear classifiers e.g., model-based, SVM etc. We learn the needed parameters from a training set and use the models on test data. The proposed method has been tested using a publicly available multi-modal score set from NIST. Our results outperform the published results in the literature.



**FIGURE 1:** Genuine and impostor score distribution and their cumulative distributions.

The rest of the paper is organized as follows. Section 2 discusses recent work in the area of biometric fusion. Section 3 describes the basic QPLC transform technique. Results of the proposed method are described in Section 4. Finally in section 5, we analyze the performance of the system and provide conclusions.

## 2. RELATED WORK

There have been several interesting tutorial like articles in the broad area of biometrics fusion [8]. Several decision level fusion methods have been described in [10]. Kittler et al. [9] wrote one of the most influential papers involving general classifier fusion techniques. The methods described in this classic paper can be applied to biometrics classifiers. However, before the various rules can be applied for fusion of biometrics engines, one has to go through a set of score normalization methods. Several score normalization techniques such as min-max, Z-

normalization, Median, Median Absolute Deviation, double sigmoid, tanh have been described in [11, 15]. It is quite well known that min-max, Z-normalization and similar score transformation methods are sensitive to outliers while tanh and sigmoid based transforms are robust to outliers. Ulery et al. [12] have studied several score level fusion methods for a large public score set and concluded that product of log likelihood ratios and logistic regression outperformed other techniques. Rank level fusion techniques like Borda count [13] have been applied to the biometric fusion problem in the recent past [14]. Poh, Kittler, and Bourlai [15] have proposed a quality based score normalization and subsequently applied it to multimodal fusion. In this quality-based score normalization, Poh et al. incorporated the qualitative device information. In case the device information is not available, the technique can still be used, but with the qualitative device information, the technique outperforms the other competitive methods. Vatsa et al. [16] also separately computed quality scores from fingerprint images and augmented these scores with the classifier scores and finally fused them using DS<sub>m</sub> theory to improve the performance of the resultant verification engine. Vatsa et al. [17] incorporated the likelihood-ratio test statistic in an SVM framework to fuse various face classifiers towards improved verification scores. Singh et al. [18] also used SVM for multimodal biometric information fusion. Vatsa et al. fused textural level matching scores and topological level matching scores to produce an improved iris recognition system in [19].

### 3. METHOD

In this section, we describe the data transformation and the modeling that we used for the multimodal biometric authentication.

#### 3.1 Data Transformations

We transform the data such that the outliers do not affect the distribution. In the literature [5, 11, 15], three different kinds of data transformation have been used. These are min-max transformation, Bayesian approach, and non-linear transformation using sigmoid (tanh(.)) function. We perform non-linear transformation of the data using quantile transformation. For each modality, we compute  $q$  quantiles (where  $q$  is an input variable) and then represent these  $q$  quantiles as  $q+1$  bins. For example,  $i$ -th bin is the range of values between quantile  $i-1$  and  $i$ . In this process, if there is an outlier far from the distribution then also it is mapped to either 1<sup>st</sup> or the last bin.

In our multimodal biometric dataset, the samples are highly imbalanced. For example, if there are  $M$  individuals then we have only  $M$  genuine scores and  $M(M-1)$  imposter scores. Therefore, we have only  $100/M$  % genuine scores and the rest are the imposter scores. For a large value of  $M$ , most of the distribution appears from the imposter data. Therefore if we compute the quantiles over the entire dataset including genuine and imposter then almost all the bins will be occupied by the imposter samples, and only one bin or only part of one bin will be occupied by the genuine samples which results in poor classification.

In order to suitably transform both genuine and imposter samples, we compute the quantiles of the imposter distribution and the genuine distribution separately. We use equal number of quantiles for both the imposter and the genuine distribution. Note that, it may not be necessary to have equal number of quantiles for both imposter and genuine distributions, however, we use the same number of quantiles in our data transformation.

Let  $x$  be any score for a modality  $i$ . Let the quantile values computed from the genuine scores be  $[y_1, y_2, \dots, y_q]$  where  $q$  is the number of quantiles. Similarly let the quantile values computed from the imposter distribution be  $[z_1, z_2, \dots, z_q]$ . We transform  $x$  using the quantile values of the

genuine distribution to  $k_{gen}(x)$  where  $y_{k_{gen}(x)} \leq x < y_{k_{gen}(x)+1}$ . If  $x \geq y_q$  then  $k_{gen}(x) = q + 1$ .

Similarly we obtain the transformation of  $x$  to  $k_{imp}(x)$  using the imposter quantile values ( $[z]$ ). We then obtain the resultant transformed score as

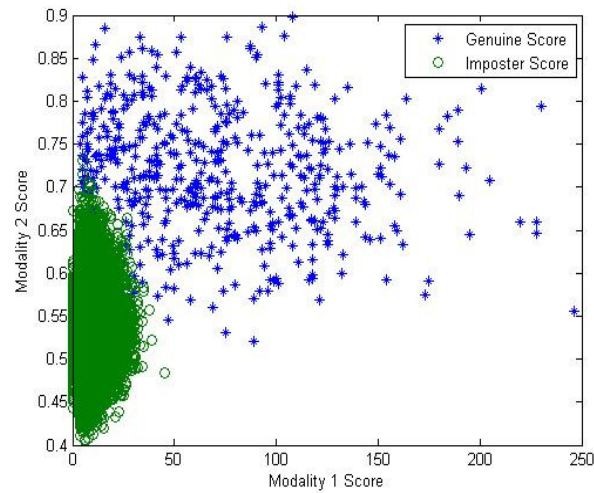
$$k(x) = k_{gen}(x) + k_{imp}(x) \quad (1)$$

Once we obtain the transformed values, we normalize  $k$  by  $2q+2$ , i.e.,  $k(x) = k(x)/(2q+2)$ , since  $k$  can attain a maximum value of  $2q+2$ . We first compute the transformed scores for the training data. We preserve the quantile information for all modalities derived from the training data. We then perform the model fitting on the transformed training data. For a test sample, we use the quantile information as derived from the training data and transform the test sample in the same way as in Equation (1) using the quantile information from the training data.

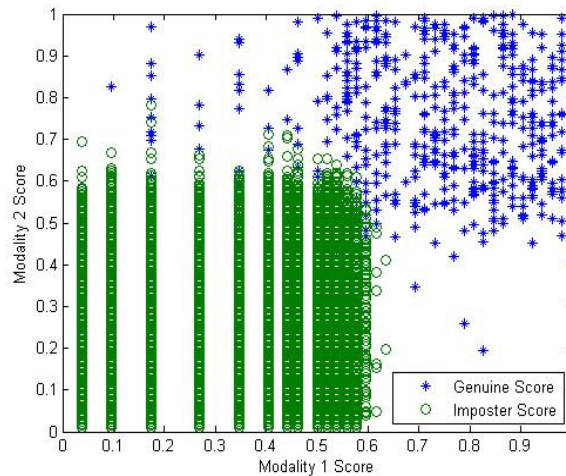
Ideally, if the genuine samples are separated from the imposter samples for a specific modality then after transformation, the transformed imposters will take values in the range  $[0,0.5]$  and the genuine samples will take values in the range  $[0.5,1]$ . This is illustrated in Fig. 3. Fig. 2 is the original score distribution of the two of the modalities of the NIST-BSSR1 dataset and Fig. 3 shows the effect of quantile transform on these scores. In the multi-modal score distribution, we can view the transformed scores to be bounded in a four-dimensional hypercube. The imposter samples will be roughly confined in the box defined by  $[(0, 0, 0, 0), (0.5, 0.5, 0.5, 0.5)]$  and the genuine samples will occupy rest of the volume. Once we compute the normalized transformed scores, we raise the scores to a certain positive power  $p$  i.e.

$$K(x) = k^p(x) \quad \text{where } p > 1 \quad (2)$$

With the increase in  $p$ , the volume occupied by the imposter samples in the hypercube decreases and the volume occupied by the genuine samples increases. In other words, the imposter sample distribution gets squeezed and the genuine sample distribution expands. This is evident in Fig. 4 which shows the score distribution of the transformed NIST-BSSR1 scores for two of the modalities (the original score distribution is as shown in Fig. 2). We perform the quantile power transformation (QPT) as in Equation (2) and subsequently use linear classifier to classify the multi-modal scores. We denote QPT along with the linear classifier explained in section 3.2 as QPLC.



**FIGURE 2:** Score distribution of two of the modalities of the NIST-BSSR1 dataset.



**FIGURE 3:** Quantile transformed score distribution of two of the modalities of the NIST-BSSR1 dataset (before the power transform)

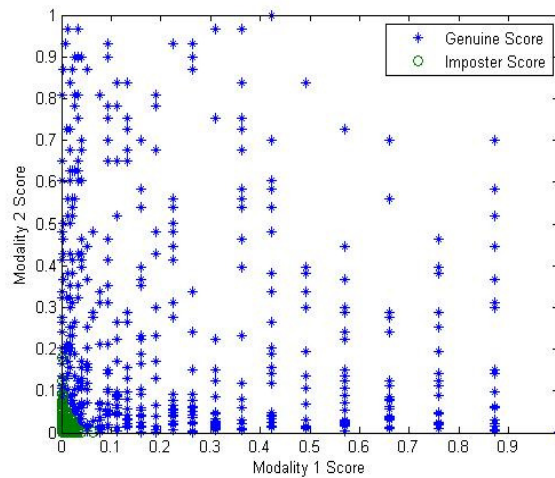
### 3.2 QPLC Model Fitting

We first transform the scores of the training data using the quantile mapping and then normalize the scores. We then raise the normalized transformed scores to certain positive power and then perform linear classification. In order to find out the linear classification boundary, it is possible to perform various techniques which include logistic regression and linear SVM. However, the cost of misclassification for the genuine samples and imposter samples are not the same in our classification task. The objective here is to attain the maximum possible TAR with minimum possible FAR. We restrict the FAR to certain low value and find the optimum classification boundary to increase TAR as much as possible.

As we mentioned before, we have four different modalities namely the left index, right index, and scores produced by two different matchers. Let us represent the separating hyperplane by  $[w_1, w_2, w_3, w_4, \theta]$  where first four parameters define the orientation of the hyperplane in the four-dimensional space and the last parameter defines the intercept. We constrain the orientation

parameters as  $\|w\|^2 = 1$  such that we have four free variables including the intercept. We then perform search over a four-dimensional hypersphere to obtain the orientation parameters. We search over the hypersphere in steps of certain  $\Delta w$ , and compute the ROC (FAR vs. TAR) for each such model.

We then obtain the set of models which produces the maximum TAR for a certain low FAR (FAR = 0.01%). Once we obtain the set of such models, we compute the AUC (area under the ROC curve) for each such model in the subset. We select one model from the subset which produces the maximum AUC. It is possible that more than one model in the subset produces the maximum AUC, and we randomly select one of such models. The overall approach is shown in Fig. 5. Once we obtain a model computed from the training data, we apply the same model on the test data. We vary the intercept to obtain the ROC on the test data.



**FIGURE 4:** QPT transformed score distribution of two of the modalities of the NIST-BSSR1 dataset ( $\rho = 7$ )

### 3.3 Quantile Transformation Applied To SVM

Vector Machine (SVM) classifier has been quite successfully applied to a diverse set of classification problems. To further validate the effectiveness of the proposed QP transformation, we have used the transformed dataset to train a linear kernel SVM [3]. Libsvm [2] library has been used to train the following two linear SVMs.

1. SVM trained on original dataset
2. SVM trained on QP transformed dataset with  $p = 7$

In our experiments we have found that QP transformed SVM performs better than the SVM trained on the original data. This may be attributed to the better suitability of QP transformed data for linear classification. The detailed results are presented in the following section.

```
=====
for  $w_1 = 0 : \Delta w : 1$ ,
     $R_1 = \sqrt{1 - w_1^2}$ ;
    for  $w_2 = 0 : \Delta w : R_1$ ,
```

```


$$R_2 = \sqrt{1 - w_1^2 - w_2^2},$$

for  $w_3 = 0 : \Delta w : R_2,$ 

$$w_4 = \sqrt{1 - w_1^2 - w_2^2 - w_3^2};$$

compute the  $ROC(w_1, w_2, w_3, w_4);$ 
end
end
end

Obtain the subset  $W$  of models from  $ROC$  which produces maximum TAR for FAR 0.01%;
for each  $(w_1, w_2, w_3, w_4) \in W,$ 
compute the  $AUC(w_1, w_2, w_3, w_4)$ 
Select a model vector  $w^* \in W$ 
where  $w^* = \arg \max_{w \in W} AUC(w);$ 
=====

```

**FIGURE 5:** Linear classifier of QPLC

## 4. RESULTS

The performance of the QPLC method was evaluated on a public-domain dataset NIST-BSSR1[1]. This dataset contains multimodal (two fingerprint and two face) scores for 517 users (NIST-517). It also contains two fingerprint matchers' scores for 6000 persons (NIST-Fingerprint) and two face matchers' scores for 3000 persons (NIST-Face). The first set of experiments was performed on the multimodal 517 users dataset using 20-fold cross validation. The results reported are the average values over these 20 folds.

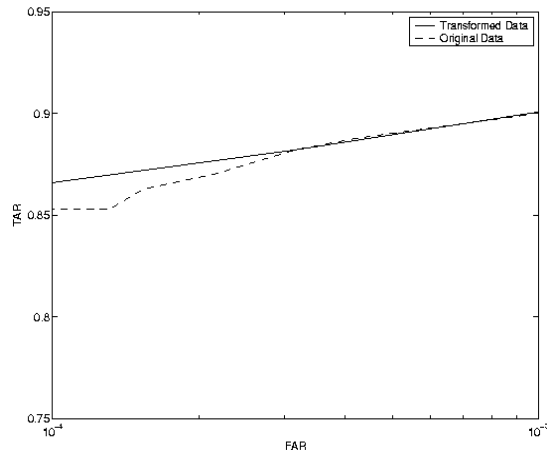
For the second set of experiments, a larger training dataset was generated for the four modalities by combining the 3000 NIST-Face scores with the first 3000 NIST-Fingerprint scores (We refer to this dataset as NIST-3000). The test dataset used was the 517 sample NIST-Multimodal dataset. We first show that the quantile transformation improves the Receiver Operating Characteristic (ROC) curves even on single modality. For example Fig. 6 displays the ROC on the right index fingerprint distribution for both the original data and the transformed data. We transformed the distribution using a quantile bin of size 4.

After transformation, the scores take an approximate uniform distribution. The imposter samples get concentrated in  $[0, 0.5]$  and the genuine samples get concentrated in  $[0.5, 1]$  (as illustrated in Fig. 3).

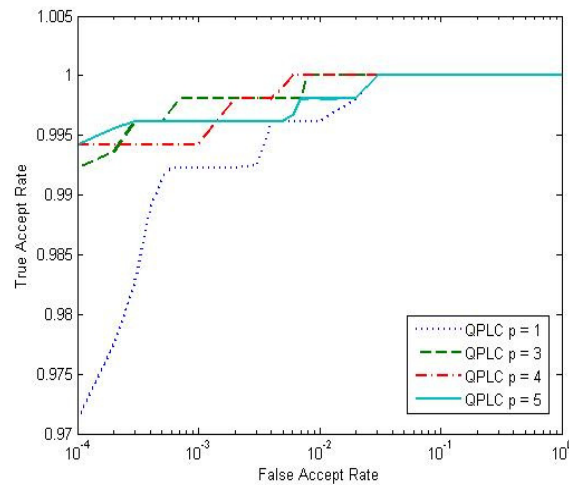
As discussed in section 3.1, raising the normalized transformed scores to a positive integer power changes the genuine and imposter distributions and we show that this helps the classification. The ROC plots in Fig. 7 and Fig. 8 compare the performance of QPLC with different values of powers for the NIST-multimodal and NIST-3000 datasets respectively. It can be seen that the classification performance improves with the higher values of power for lower values of FAR and then the curves coincide for the higher values of FAR as expected.

In Fig. 9 we compare the ROC of the linear SVM, QPT based SVM and Logistic Regression on the NIST-517 dataset. The results indicate that the QPLC achieves significant improvement in the TAR values for low values of FAR as compared to the other techniques. Further, we note that the QPT based SVM performs better than the SVM trained on the original dataset. Fig. 10 shows the

results of these classifiers on the larger NIST-3000 dataset. These results also show a similar trend as in Fig. 9. The QPLC outperforms other techniques and the QPT based SVM significantly outperforms the SVM trained on the original dataset. The improvement in the SVM performance as a result of the QPT is important since SVM is a widely used scalable classifier.



**FIGURE 6:** Comparison of ROC performance on original data and transformed data for right index finger print recognition.

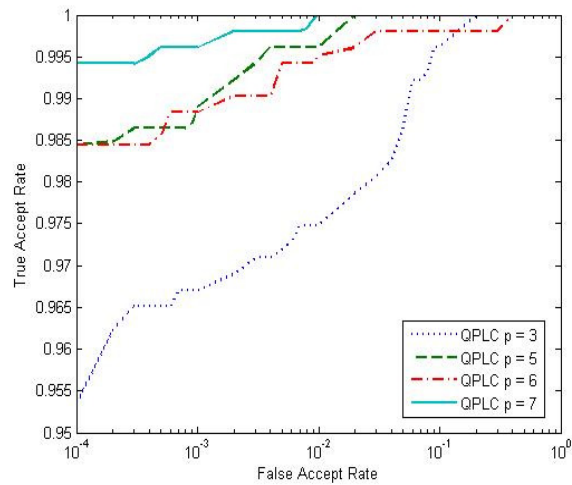


**FIGURE 7:** Comparison of ROC performance on quantile transformed data raised to different values of powers for NIST-517 dataset.

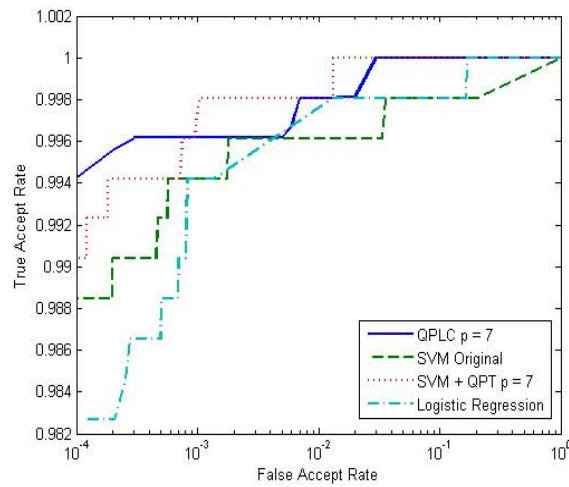
Table.1 summarizes the TAR values for the FAR of 0.01 percent for all these fusion techniques. In [5] the authors have proposed a likelihood ratio test (LRT) based biometric score fusion. As their results are also based on the 517 sample dataset, we directly compare their LRT based result with the proposed technique in the Table 1. We also directly report the results of a linear classifier based fusion technique [4] on the same dataset. We also compare our transformation to the well known tanh score normalization [11, 15]. We use SVM to classify the tanh transformed scores and the result is reported in Table.1. This result and the results reported in [15] for tanh normalization in combination with different fusion methods on NIST-517 dataset indicate that QPT



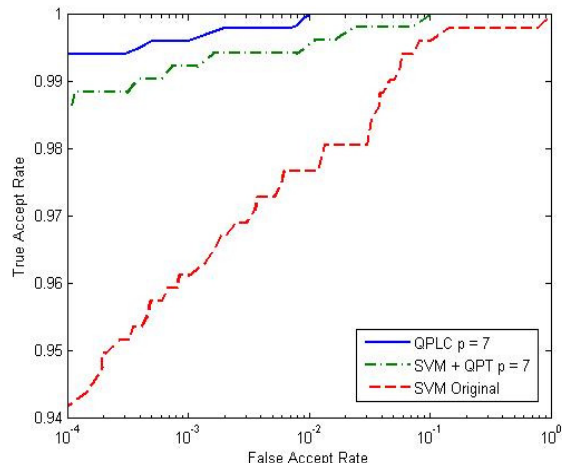
performs better than tanh normalization. From all the results, we can observe that the performance of QPLC is better than all the other techniques compared.



**FIGURE 8:** Comparison of ROC performance on quantile transformed data raised to different values of powers for NIST-3000 dataset.



**FIGURE 9:** Comparison of ROC performance of QPLC with SVM, SVM + QPT  $p=7$ , and Logistic Regression on the NIST-517 dataset.



**FIGURE 10:** Comparison of ROC performance of QPLC with SVM and SVM + QPT  $p = 7$  on the NIST-3000 dataset.

Technique	NIST-Multimodal	NIST-3000
LC [4]	99.00	-
GMM [5]	99.10	-
Logistic Regression	98.26	-
SVM + tanh	90.56	-
SVM	98.84	94.19
SVM + QPT	99.03	98.65
QPLC	<b>99.42</b>	<b>99.42</b>

**TABLE 1:** TAR (%) values for different methods at 0.01% FAR

## 5. CONCLUSIONS

In a multimodal score fusion problem, often one has to deal with the scores from the various modalities whose dynamic ranges and probability distributions vary a lot. As a solution to this problem, we have proposed a quantile transformation which is independent of the dynamic ranges of each modality and is not highly susceptible to the outliers. Further we show that raising the normalized quantile values to a power greater than one results in a lower FAR and a higher TAR.

Finally, a linear classifier (QPLC) and a SVM trained on the QPT scores significantly outperformed the other classifiers (LRT [5], linear classifier [4] and SVM) that were trained on the original scores confirming the utility of the QPT. We also compared it with other score normalization methods like tanh [11, 15] and found that QPT performs better.

QPLC is also designed to particularly handle imbalanced data. We observe that for NIST-3000 dataset, QPLC outperforms other linear classifiers. Since we consider the maximization of AUC explicitly under the constraint of achieving a certain minimum TAR, it is not affected by the imbalance in the samples. The linear classifier of QPLC is constrained by the dimensionality. We

have four modalities and it was possible to design an explicit search mechanism. However, if the dimensionality increases, it may not be possible to perform the explicit search. Overall for relatively low-dimensional dataset and highly imbalanced class samples, QPLC has the potential to outperform the existing classifiers. In this paper, we tested with NIST-BSSR1 dataset, and as a future study we expand the experiments with other multi-modal biometric datasets as well.

## 6. REFERENCES

- [1] National Institute of Standards and Technology, NIST Biometric Scores Set – Release 1, <http://www.itl.nist.gov/iad/894.03/biometricscores/>, 2004.
- [2] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [3] C. Cortes and V. Vapnik, Support-vector networks, *Machine learning*, 20(3):273-297, 1995.
- [4] M. C. Zoepfl and H. J. Korves, Improving identity discovery through fusion, *ITPro Jan/Feb 2009*.
- [5] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, Likelihood ratio based biometric score fusion, *IEEE Trans on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, Feb 2008.
- [6] S. Garcia-Salicetti, M.A. Mellakh, L. Allano, and B. Dorizzi, Multimodal biometric score fusion: the mean rule vs. support vector classifiers, *Proc. EUSIPCO*, 2005.
- [7] M. Villegas and R. Paredes, Score Fusion by Maximizing the Area under the ROC Curve, *Pattern Recognition and Image Analysis: 4th Iberian Conference, IbPRIA*, June 2009.
- [8] A. K. Jain, A. Ross, Multibiometric systems, *Communications of the ACM, Special Issue on Multimodal Interfaces* 47 (1), 34–40, 2004.
- [9] J. Kittler, M. Hatef, R. P. Duin, J. G. Matas, On Combining Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (3), 226–239, 1998.
- [10] S. Prabhakar, A. K. Jain, Decision-level Fusion in Fingerprint Verification, *Pattern Recognition* 35 (4), 861–874, 2002.
- [11] A. K. Jain, K. Nandakumar, and A. Ross, Score Normalization in Multimodal Biometric Systems, *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, December 2005.
- [12] B. Ulery, A. R. Hicklin, C. Watson, W. Fellner, and P. Hallinan, *Studies of Biometric Fusion*, NIST, Tech. Rep. IR 7346, September 2006.
- [13] C. Dwork, R. Kumar, M. Naor and D. Sivakumar, Rank aggregation methods for the Web, *WWW '01: Proceedings of the 10th international conference on World Wide Web*, 613-622, 2001.
- [14] M. L. Gavrilova and M. M. Monwar, Fusing multiple matcher's outputs for secure human identification, *Int. J. Biometrics*, 1(3), 329-348, 2009.
- [15] N. Poh, J.V. Kittler, and T. Bourlai, Quality-based score normalization with device qualitative information for multimodal biometric fusion, *IEEE Transactions Systems, Man, and Cybernetics – Part A*, 40(3), 539-554, 2010.

- [16] M. Vatsa, R. Singh, A. Noore, and M. Houck, Quality-augmented fusion of level-2 and level-3 fingerprint information using DS<sub>m</sub> theory, *International Journal of Approximate Reasoning*, 50(1), 2009.
- [17] M. Vatsa, R. Singh, A. Ross, and A. Noore, Likelihood ratio in a SVM framework: fusing linear and non-linear face classifiers, *IEEE Computer Vision and Pattern Recognition Workshop*, Anchorage, AK, 1-6, 2008.
- [18] R. Singh, M. Vatsa, and A. Noore, "Intelligent Biometric Information Fusion using Support Vector Machine," *Soft Computing in Image Processing: Recent Advances*, M. Nachtgael, D. Weken, E. Kerre, and W. Philips (editors), Springer-Verlag Publishers, 2008.
- [19] M. Vatsa, R. Singh, and A. Noore, Improving Iris Recognition Performance using Segmentation, Quality Enhancement, Match Score Fusion and Indexing, *IEEE Transactions on Systems, Man, and Cybernetics - B*, 38(3), 2008.
- [20] Andrade, C. and von Solms, S. H., Investigating and comparing multimodal biometric techniques, *Policies and Research in Identity Management*, IFIP International Federation for Information Processing, vol. 261, pp. 79–90, 2008.