A Multi-Biometric Feature-Fusion Framework for Improved Uni-Modal and Multi-Modal Human Identification

Dane Brown*† and Karen Bradshaw*
*Department of Computer Science
Rhodes University
Grahamstown, South Africa
†Council for Scientific and Industrial Research
Modelling and Digital Sciences
Pretoria, South Africa
Email: dbrown@csir.co.za

***Abstract*—The lack of multi-biometric fusion guidelines at the feature-level are addressed in this work. A feature-fusion framework is geared toward improving human identification accuracy for both single and multiple biometrics. The foundation of the framework is the improvement over a state-of-the-art uni-modal biometric verification system, which is extended into a multi-modal identification system. A novel multi-biometric system is thus designed based on the framework, which serves as fusion guidelines for multi-biometric applications that fuse at the feature-level. This framework was applied to the face and fingerprint to achieve a 91.11% recognition accuracy when using only a single training sample. Furthermore, an accuracy of 99.69% was achieved when using five training samples.**

***Index Terms*—face, fingerprint, feature-level, multi-modal biometrics, state-of-the-art**

I. Introduction

Biometrics are measures of unique biological and behavioural patterns of an individual, which are evaluated using either verification or identification methods [1]. Biometrics are useful as they cannot be lost or forgotten in the traditional sense, unlike smart devices and passwords. However, their popularity has introduced security risks posed by forgers. Furthermore, real-world problems, such as damaged fingerprints and facial occlusions, pose the risk of unreliable biometric systems.

Multi-biometric fusion has proven effective in addressing real-world biometric applications. Multi-modal biometrics can also be used to solve non-universality and bad input data in well-planned applications by intelligently selecting an appropriate modality [2]. Therefore, uni-modal biometric recognition performance can be as important as multi-modal biometric recognition performance in certain multi-modal biometric applications.

Multi-modal biometric studies initially adopted the score-level fusion approach. However, the feature-level approach was shown to outperform the score-level in recent studies [3]. This was attributed to the rich discriminatory information available at the feature-level. Comprehensive reviews have been conducted at the matching score-level and the results were often used to construct fusion frameworks [4]. These frameworks provide important guidelines that enable the systematic implementation of multi-modal biometric systems for future research and applications. Feature-level fusion is in particular need of these guidelines because of the "curse of dimensionality" problem posed by the rich amount of data in features. Moreover, the classification of human identification systems result in a considerably higher dimensionality problem than human verification systems and thus the extracted features play a very important role during the classification process. This paper proposes a framework, based on preliminary experiments, which will be used to implement a state-of-the-art multi-modal biometric system with high human identification rates for both individual and fused modalities. Biometric modalities, represented by an image, are independent and complementary, thus feature-fusion guidelines are determined by applying the same feature transformation method on different modalities [4].

In this paper, different feature selection and transformation methods (feature processing modules) are applied to the face, palmprint and fingerprint. The fused feature sets are expected to improve recognition performance compared with the individual feature sets in the majority of cases. However, individual feature sets can outperform a fused feature sets containing one or more poor quality individual feature sets. This is expected to further improve the recognition performance. The scope includes the use of different sized datasets and varying the number of training samples used during data modelling.

The rest of the paper is organized as follows: Section II presents the related studies. Sections III and IV discuss the construction and application of the image-based feature-fusion guidelines. The experimental analyses and results are discussed in Section V. Section VI concludes the paper and discusses ongoing work.

II. Related Studies

The following related studies present multi-modal human identification systems that combine the face, fingerprint or palmprint.

Yao *et al.* [5] combined the face and palmprint and processed the fused dataset with four PCA-based feature-fusion algorithms. The best performing algorithm filters both input modalities with Gabor filters followed by weighted parallel fusion of the resulting feature set. The remaining three algorithms either omit the Gabor filtering or use non-weighted parallel fusion. The best performing algorithm produced a high accuracy with only a single training sample.

The AR face and PolyU palmprint datasets were used. Each dataset consisted of 20 images per 189 individuals with a resolution of 60×60. A 91% genuine acceptance rate (GAR) was achieved with a single training sample and 95% GAR was achieved using six training samples.

The face and fingerprint feature-fusion studies [6] and [7] both use a Curvelet transform followed by SVM classification. The Curvelet transform makes the system resistant to misalignment and multiple capturing angles. However, the datasets used in both studies are not publicly available.

The limited studies, lack of metrics and the use of private datasets on face, fingeprint and palmprint feature-fusion systems make comparisons a non-trivial task. Furthermore, many of the studies only perform human verification.

III. Methodology

This section shows the steps taken to determine the multi-biometric feature-fusion framework.

A. Uni-modal Biometrics

The experimentation process was initialized by determining a face verification method that can outperform the current state-of-the-art, known as GaussianFace [8]. The goal of GaussianFace was to surpass the 97.53% face verification accuracy achieved by the human eye, on the cropped LFW dataset. GaussianFace surpassed this accuracy by 0.99% GAR with 0.15% FAR when training on five additional source domain datasets. GaussianFace achieved a face verification accuracy of 90% when only using the LFW dataset.

The face verification method proposed in this paper achieved a 95% face verification accuracy on the same LFW dataset. The methodology used a Laplacian of Gaussian (LOG) filter and a modified version of the extended Local Binary Pattern (ELBP) texture descriptor, followed by a transformation to the Eigen space. This algorithm significantly improves discrimination of between-class and within-class data. Therefore, this algorithm was extended from human verification to accurate human identification compatible with the face and other image-based modalities. Our methodology applies this image-based biometric identification algorithm to the face, fingerprint and palmprint modalities as follows.

B. Categories of Datasets

Two multi-modal datasets were formed by pairing the first 40 individuals of FVC2004 Fingerprint DB1 with ORL Face and SDUMLA Fingerprint right middle fingers with Fei Face, respectively. Fingerprint images consisted of the following three quality categories: partials with absent core points, poorly-defined ridges and well-defined ridges. Performing experiments on those fingerprint groups helped determine the feature selection and transformation algorithms most effective at dealing with all three groups. Face images were also organized into three categories, consisting of standard frontal faces, frontal faces that consisted of poses and props, and non-frontal faces.

More multi-modal datasets were formed by pairing PolyU palmprints with each of the above datasets, forming bi-modal and tri-modal datasets. The following subsection details the interactions of the most relevant feature processing modules and classifiers.

C. Preliminary Experiments

Uni-modal, bi-modal and tri-modal datasets were evaluated using various human identification systems based on combinations of feature processing modules and classifiers. All evaluation was performed after basic pre-processing, consisting of image alignment, pixel normalization and histogram equalization. The following results relate to the various face and fingerprint datasets only.

Local Binary Pattern Histogram (LBPH) proved to be a versatile classifier capable of achieving high recognition accuracy and only requiring basic pre-processing. LBPH was particularly robust to misalignment, dynamic lighting and scale. On the other hand, Eigen and Fisher classification performed poorly on the majority of datasets. However, Eigen and Fisher classification achieved recognition accuracies similar to that of LBPH for face and fingerprint images consisting of standard frontal faces and well-defined ridges, respectively. The improved performance on the latter datasets for both Eigen and Fisher can be attributed to the low variance in data across multiple samples of face and fingerprint images contained within those datasets.

An effort was made to reduce the variance in the majority of datasets by applying a LOG filter. The LOG filter significantly improved the recognition accuracy of all the datasets. It was particularly useful at lowering the data variance of multiple samples of face and fingerprint images consisting of poses and props and poorly-defined ridges, respectively. However, a side effect occurred when applying the LOG filter to a data class with varied lighting. This side effect occasionally caused a reduction in recognition performance.

The ELBP operator was modified in an effort to reduce the variation in data across multiple samples of an individual. The standard ELBP operator highlights high frequency data using interpolation, which introduces noise into an image. The modification to the ELBP operator reduced the noise captured by the feature descriptor by increasing the radius in correlation to the interpolated neighbours. The modified ELBP operator significantly outperformed the histogram equalization and pixel normalization under dynamic lighting conditions. This was used before the LOG filter to achieve the optimal feature set.

The Gabor filter is a popular means of improving feature discrimination, however, it did not improve the accuracies of the fused datasets. Moreover, it reduced the accuracy of the Eigen and Fisher classifiers. However, it achieved an improved recognition accuracy on non-partial fingerprint datasets and slightly improved the recognition accuracy of all face datasets when using the LBPH classifier. On the other hand, the LOG filter lowered the accuracy of the LBPH classifier. The LBPH classifier did not respond positively to the majority of feature processing modules. Reducing the data variance by 1%, using principal component analysis (PCA) produced the best improvement to LBPH recognition accuracy across all datasets.

The results on the individual palmprint dataset corresponded to the interactions described above. However, contrary to the face and fingerprint datasets, the palmprint dataset was of consistent and high quality. Fused datasets that consisted of the palmprint thus achieved perfect accuracies in all cases when using the modified ELBP and LOG filter combination. The interactions observed in these preliminary experiments were used to construct a framework consisting of fusion guidelines. The framework was applied to the proposed systems in the following section.

IV. Implementing the Proposed Systems

The proposed systems are implemented based on the feature-fusion framework determined in Section III. The face and fingerprint are combined using feature-level fusion in the final experiments as follows.

A. Dataset

The face and fingerprint datasets were used from the SDUMLA multi-modal database [9]. Since it is a true multi-modal database, all individuals, totalling 106, were used in the final experiments discussed in Section V. Eight samples of the left thumbprint were selected from the fingerprint images. The frontal faces were selected consisting of various poses and props – normal, smile, frown, surprise, looking down, eyes shut, hat and glasses. The training samples were sequentially chosen from one to five and the rest were used for testing. To the best of our knowledge there are no studies that fuse face and fingerprint data acquired from the SDUMLA multi-modal database.

B. Pre-processing

Pixel normalization sets the pixel values of an image to a constant mean and variance for improved consistency of lighting and contrast. Histogram equalization is another method for improving the consistency of lighting and contrast across a dataset [10]. Histogram equalization is often more effective than pixel normalization, but the greyscale range is distributed uniformly by applying a non-linear transformation, which can cause a slight side effect on the histogram shape. Therefore, histogram equalization should be avoided in most histogram-based matching methods.

C. Feature Selection

The LOG filter enhances the discrimination of the face and fingerprint images by removing unwanted features on the very high frequency spectrum, while effectively increasing the mean image component. However, increased feature discrimination can further highlight the difference among multiple samples of badly aligned images, consequently reducing the recognition accuracy.

A cropping method based on multiple Haar cascades and Poincaré index was used to automatically prune noise and resize the face and fingerprint datasets to a 75×75 region of interest (ROI), respectively. Multiple Haar cascades were iterated to detect the face, eyes, nose and mouth. Many Haar cascades were in place for redundancy purposes in the case of detection failure. The outlining as illustrated in Fig. 1 was determined based on the detected face, eyes, nose and mouth and was used to eliminate or reduce partial occlusions that affect face recognition. Non-local means filtering [11] was used to enhance the fingerprint before applying the Poincaré index algorithm. The inconsistent pressure applied during fingerprint capturing was dealt with by cropping the fingerprint around the core point.

Feature discrimination was further improved by applying our proposed novel algorithm, consisting of the modified ELBP operator and LOG filter, to the face and fingerprint ROIs.

D. Feature-Fusion Transformation

The highly discriminative face and fingerprint feature vectors are combined using serial vector fusion. Eigen and Fisher classifiers require the feature vectors to be transformed to the Eigen space. This Eigen space representation is only a reconstruction for visual purposes and is not used during classification. The LBPH classifier makes use of a spatial histogram representing standard image space and will be elaborated further in the next subsection. Instead of showing histograms of data, the more descriptive ELBP representation is shown in the figure. The fused feature vectors representing the Eigen and Fisher methods as well as the LBPH method is illustrated in Fig. 1 on the left and right side, respectively.

E. Multi-modal Classification

The key element of the Eigen classifier is represented by the following total scatter matrix, when given N sample images x [12]:

$$S_t = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \quad (1)$$

where $m \in \mathbb{R}^n$ is the mean image obtained from the samples.

The Fisher classifier performs extra class-specific dimensionality reduction by clustering same class data tightly and

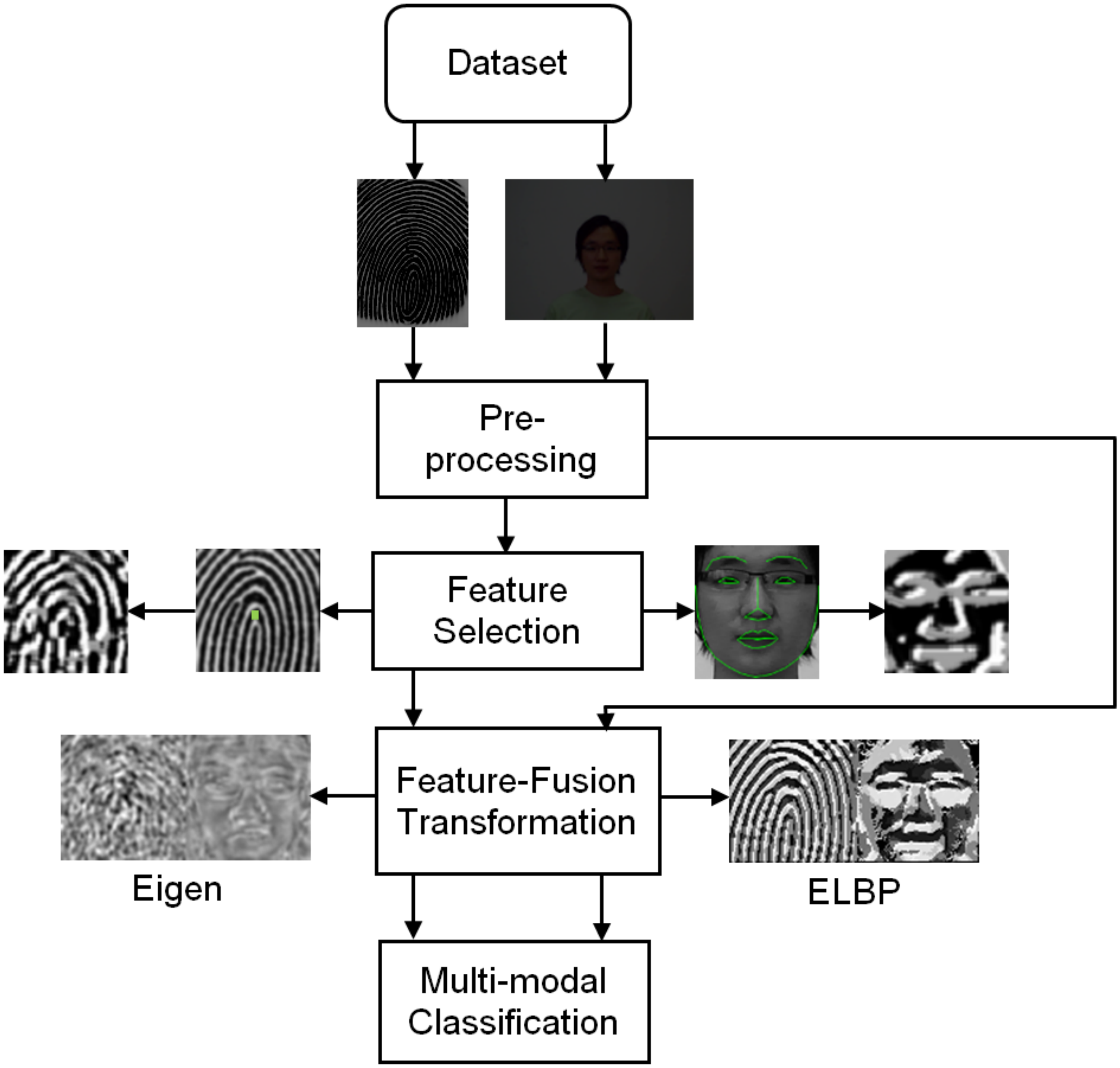

Fig. 1: Overview of Proposed Methodology.

maximizing the separation of different classes in a lower-dimensional representation. The key element of the Fisher classifier is thus represented by the following between-class and within-class scatter matrix.

Given C classes the between-class scatter matrix is defined as [12]:

$$S_b = \sum_{i=1}^{C} N_k(\mu_k - \mu)(x_k - \mu)^T \tag{2}$$

while the within-class scatter matrix is defined as:

$$S_w = \sum_{i=1}^{C} \sum_{x_k \in \mathbb{X}_i} (x_k - \mu)(x_k - \mu)^T \tag{3}$$

The LBPH classifier is a feature descriptor based on the local binary pattern (LBP) operator. We use a special kind of LBP operator known as ELBP. This operator uses spatially enhanced histogram matching, which can perform partial matching and automatic pixel normalization on a pixel level, circular neighbourhood level and image level. The advantages of this are illumination, scale and rotation invariant texture classification as opposed to Eigen and Fisher classification [13]. The disadvantages are the limited or negative effects on recognition performance when applying image pre-processing and feature selection techniques before LBPH classification.

Three baseline systems are created by dividing the fused dataset into classes processed by the three classifiers.

V. Experimental Analysis and Results

The various proposed systems are identified by the following combinations of feature selection and transformation schemes with the Eigen, Fisher or LBPH classifier: Pixel normalization is used for the LBPH baseline system, Histogram equalization is used for the Eigen and Fisher baseline systems, referred to as Eh; LOG is referred to as L; Modified ELBP is referred to as LBP; Modified ELBP followed by LOG is referred to as LBPL; LOG followed by modified ELBP is referred to as LLBP; and PCA reduction is referred to as PCA.

The Eigen baseline fusion system always outperforms the face and fingerprint as illustrated in Fig. 2. The LBPL feature vector outperforms the other system when using a single training sample with an accuracy of 90.84%. The LLBP feature vector achieves the best accuracy when using five training samples with an accuracy of 99.69%.

In Fig. 3 the Fisher classifier performs noticeably weaker than the Eigen classifier when using two training samples. This is attributed to the fact that the second sample of the dataset is generally of a different category (as defined in Subsection III-B) compared with the first sample. This affects Fisher in particular due to its high sensitivity to differences in training samples.

The LBPH baseline fusion system produces a lower accuracy than the face when using three training samples as illustrated in Fig. 4. This is of huge significance because the fusion versions of all other systems perform better than their face and fingerprint counterparts. This also shows that while LBPH is a good general texture classifier, it excels at face classification. PCA reduction improves the recognition accuracy by 3% on average. Other LBPH methods that use feature selection before classification performed poorly, in general, and were omitted from the results.

The receiver operating characteristic (ROC) curve in Fig 5 shows that all three systems have a low FAR in general. The EigenLLBP system provides a highly significant improvement over the FisherLLBP and LBPHPCA. FisherLLBP and LBPHPCA achieved the same maximum GAR at about 2.4% FAR. However, FisherLLBP achieved a better overall FAR than LBPHPCA across thresholds.

It is clear based on the four figures that the ELBP operator was successfully combined with the LOG filter to significantly improve feature discrimination in the Eigen space. EigenLLBP in particular showed the strength of a good feature selection algorithm by achieving the highest recognition accuracy of all the fusion schemes, at 99.69%. The individual results of EhL and LBP improved the accuracies over the baseline systems on average, but only their combinations are evaluated in this paper. LBPL achieved the best average recognition accuracy and shows great promise for single sample human identification. The results observed in this paper demonstrate the importance of a feature-fusion framework.

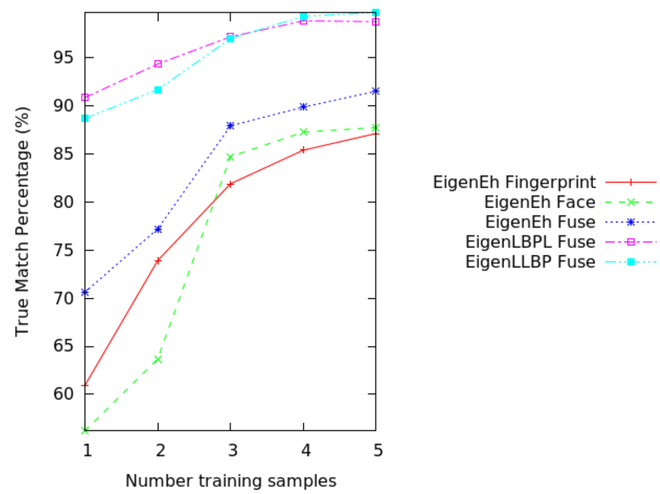

Fig. 2: Comparison of Eigen Methods

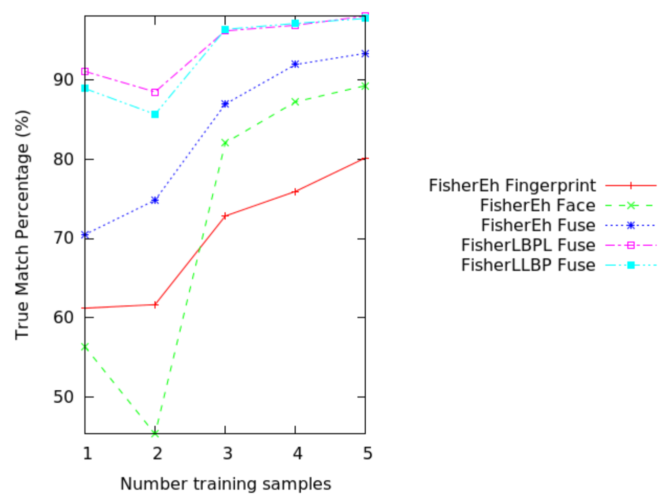

Fig. 3: Comparison of Fisher Methods

VI. Conclusion and Ongoing Work

An accurate multi-modal biometric identification system was created based on our accurate uni-modal verification system that outperformed the state-of-the-art. Preliminary experiments were conducted on the face, fingerprint and palmprint to discover a framework that provides feature-fusion guidelines. A comparison was performed on fingerprints, faces and their fused dataset using three baseline classifiers as a performance reference. The comparison was extended by combining a modified ELBP operator and a LOG filter and producing a highly accurate multi-modal biometric system based on the framework. The LBPH classifier achieved the best accuracy in the baseline fusion systems and for the face biometric. LBPH proved to be robust to misalignment, dynamic lighting and scaling. The Eigen and Fisher classifiers produced the highest accuracies after combining the highly complementary ELBP operator and LOG filter. The lack of metrics and the use of private datasets in many multi-modal

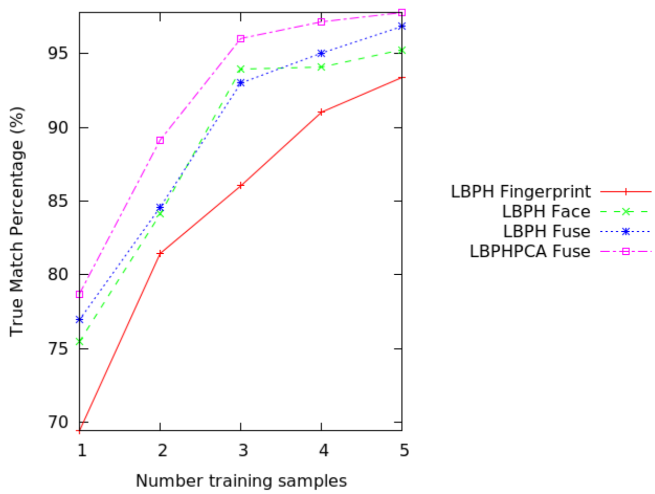

Fig. 4: Comparison of LBPH Methods.

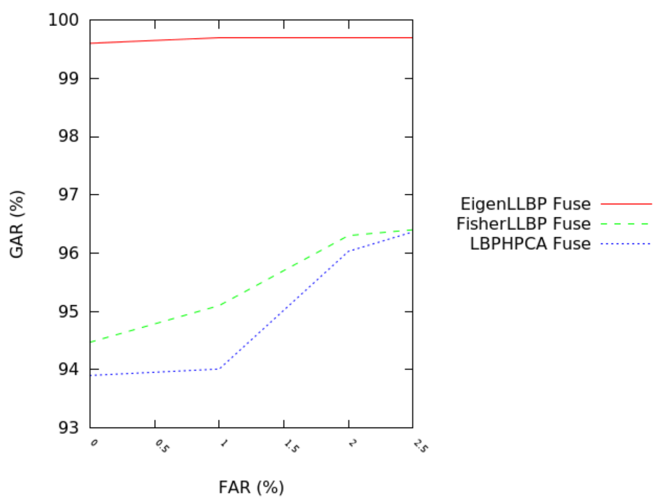

Fig. 5: ROC curve of the best performing Eigen, Fisher and LBPH methods for five training samples.

feature-level fusion studies make isolating the contributing feature processing modules a non-trivial task. The research introduced in this paper serves as a foundation for selecting appropriate features for image-based multi-modal fusion and solving the generalized feature-fusion framework problem.

In future, more combinations of biometric modalities and feature processing modules will be investigated with additional experimentation in an effort to extend this multi-biometric feature-fusion framework. An intelligent modality context switching algorithm will also be added. This context switching algorithm will automatically determine an appropriate modality and whether the system should operate uni-modal or multi-modal biometrics, based on LBPH quality scoring.

A comparison was performed on fingerprints, faces and their fused dataset using three baseline classifiers. The comparison was extended by combining a modified ELBP operator and a LOG filter. Additionally, principal components were removed from the LBPH training and testing images. The LBPH classifier achieved the best accuracy in the baseline systems and was robust to misalignment, dynamic lighting and scaling. The Eigen and Fisher classifiers yielded the best accuracies when combining the strengths of ELBP and LOG. Feature-level fusion research often makes use of well-known image processing and classification techniques without reasoning. Analyzing and testing many of these techniques to measure progress in the state-of-the-art is a non-trivial problem. Therefore, the guidelines introduced in this paper are the first step to solving the generalized feature-fusion framework problem.

An extended list of guidelines for designing optimal feature-fusion schemes is being investigated. This requires a comprehensive review of the important factors covered in this paper as well as additional experimentation on more biometric modalities, feature processing modules and more datasets. Additionally, an intelligent system that can dynamically switch between uni-modal and multi-modal operations is being investigated for solving the case of non-universality and bad input data in some modalities during data acquisition.

References

[1] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric quality: from assessment to multibiometrics," 2015.

[2] M. Vatsa, R. Singh, and A. Noore, "Context switching algorithm for selective multibiometric fusion," in *Pattern Recognition and Machine Intelligence*. Springer, 2009, pp. 452–457.

[3] A. Rattani, D. R. Kisku, M. Bicego, and M. Tistarelli, "Feature level fusion of face and fingerprint biometrics," in *Biometrics: Theory, Applications, and Systems*, 2011, pp. 1–5.

[4] R. Raghavendra, B. Dorizzi, A. Rao, and G. H. Kumar, "Designing efficient fusion schemes for multimodal biometric systems using face and palmprint," *Pattern Recognition*, vol. 44, no. 5, pp. 1076–1088, 2011.

[5] Y.-F. Yao, X.-Y. Jing, and H.-S. Wong, "Face and palmprint feature level fusion for single sample biometrics recognition," *Neurocomputing*, vol. 70, no. 7, pp. 1582–1586, 2007.

[6] M. V. Karki and S. S. Selvi, "Multimodal biometrics at feature level fusion using texture features," *International Journal of Biometrics and Bioinformatics*, vol. 7, no. 1, pp. 58–73, 2013.

[7] P. Sharma and M. Kaur, "Multimodal classification using feature level fusion and SVM," *International Journal of Computer Applications*, vol. 76, no. 4, pp. 26–32, 2013.

[8] C. Lu and X. Tang, "Surpassing human-level face verification performance on lfw with gaussianface," *arXiv preprint arXiv:1404.3840*, 2014.

[9] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: a multimodal biometric database," in *Biometric Recognition*. Springer, 2011, pp. 260–268.

[10] A. C. Bovik, *Handbook of image and video processing*. Academic Press, 2010.

[11] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 60–65.

[12] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.