A Dynamically Weighted Multi-Modal Biometric Security System

Dane Brown*†, Karen Bradshaw †

*Modelling and Digital Science, Council for Scientific and Industrial Research, Pretoria, South Africa.

[1]dbrown@csir.co.za

†Department of Computer Science, Rhodes University, Grahamstown, South Africa.

[2]k.bradshaw@ru.ac.za

***Abstract*—The face, fingerprint and palmprint feature vectors are automatically extracted and dynamically selected for fusion at the feature-level, toward an improved human identification accuracy. The feature-level has a higher potential accuracy than the match score-level. However, leveraging this potential requires a new approach. This work demonstrates a novel dynamic weighting algorithm for improved image-based biometric feature-fusion. A comparison is performed on uni-modal, bi-modal, tri-modal and proposed dynamic approaches. The proposed dynamic approach yields a high genuine acceptance rate of 99.25% genuine acceptance rate at a false acceptance rate of 1% on challenging datasets and big impostor datasets.**

***Index Terms*—face, fingerprint, palmprint, feature-level, weighted**

I. Introduction

Biometric systems acquire and evaluate the unique biological and behavioural patterns from individuals. However, their susceptibility to forgery and noisy data have led to identity theft and system failure [1] [2]. This is especially a concern in non-invasive biometric modalities such as the fingerprint, face and palmprint, which are easier to acquire because they are generally external. Furthermore, with the pervasiveness of sensor data, due to its transfer through sensor networks, well-planned access control is vital in the real-world. Many real-world security problems change based on the application, but are especially evident in automatic or unsupervised security systems, in general.

A large number of security concerns were addressed by introducing multiple sources of biometric information into a single system [3] [4]. Moreover, these multi-modal biometric systems have the potential to improve recognition accuracy and application versatility. However, leveraging that potential by performing multi-modal biometric fusion is not appropriate to every application, for instance, achieving universality or improving user experience. Achieving universality while sustaining a high recognition accuracy in the case where one or more biometric modalities are inadequately acquired is a non-trivial problem [5] [3].

The development of multi-modal biometrics initially focused on fusing at the matching score-level – simply summing the matching scores achieved by individual modalities was optimal for most applications [4]. A significantly improved and relatively recent approach performs biometric fusion at the feature-level. Effective feature-level fusion combines the feature sets of corresponding biometric modalities, while taking into account their increase in dimensionality and compatibility before classifying the result. Feature-level fusion schemes are relatively understudied in the literature. This paper extends our recently published feature-fusion guidelines; see [6] for further details.

Furthermore, this paper provides a solution to the universality problem by constructing a dynamic feature-fusion scheme that intelligently selects one or more given modalities at feature-level. The fused feature sets are expected to yield an improved recognition performance compared with the individual modalities, in the majority of cases. However, where appropriate, the system uses only one modality or a reduced combination of modalities based on their computed quality. The key factor for excluding one or more modalities lies in the case where it cannot be enhanced to a recognizable result. The scope includes experimental analysis after varying the number of training samples from one to five. The experimental results are subsequently used to determine the system's adequacy in achieving universality.

II. Related Studies

Rattani *et al.* developed a multi-modal biometric identification system that fused the face and fingerprint modalities [4]. They found that feature-level fusion significantly outperformed the matching score-level's best approach. This was achieved by sufficiently reducing feature dimensionality and normalizing the features for compatibility. Furthermore, the extraction of image region of interests (ROIs) were used for both modalities. Two other face and fingerprint identification studies were performed by [7] and [8], both using the Curvelet transform, feature-fusion averaging followed by support vector machine (SVM) classification. The results of the two studies were not conclusive on the effectiveness of SVM classification at feature-level. They also did not investigate the effectiveness of feature-fusion concatenation – fusing features serially. Other face, fingerprint and palmprint studies are limited to human verification – a simple one-to-one database match.

Vatsa *et al.* [5] proposed a context switching multi-modal biometric verification system consisting of the face, fingerprint and iris. An SVM decides between a uni-modal and multi-modal approach based on a non-linear threshold. However, this decision is limited to selecting only one of the three modalities or fusing all of them. Furthermore, fusion is performed at the matching score-level and only human verification is evaluated. This paper aims to extend their context switching method to any combination of modalities and to evaluate the human identification accuracy. Moreover, the multi-modal feature-fusion guidelines founded in [6] are applied in this paper in combination with context switching.

III. Proposed Methodology

This section explains the process of developing an improved multi-modal biometric system by using a novel dynamic weighting algorithm.

A. Datasets

The ARFace dataset [9], SDUMLA fingerprint dataset [10] and IITD palmprint dataset [11] were used in the experiments discussed in Section IV. The five samples per 100 individuals used for training and testing can be summarized as follows:

1) Face – The faces consisted of frontal poses with the following expressions and props: neutral; laugh; neutral dynamic lighting; sunglasses; and a scarf covering from the nose down. Imposter samples were similarly gathered from another 100 individuals.
2) Fingerprints – The left thumbprint images were used consisting of various qualities ranging from partials with absent core points and poorly-defined ridges to well-defined ridges. Imposter samples were similarly gathered from the left middle fingerprint of the same individuals.
3) Palmprint – The challenging IITD palmprint dataset was used instead of the popular, but near-ideal quality PolyU[1] [12] palmprint dataset. The originally tested PolyU samples produced a perfect accuracy and was thus excluded from this study. This result already shows promise toward improving image-based uni-modal biometric identification systems. The IITD palmprints consisted of touchless captured hands with uncontrolled poses. Imposter samples were similarly gathered from another 100 individuals.

B. Automated Image Registration

Image registration is an important first step in biometric recognition. At the local level, there are unique points within biometric data depending on the considered biometric modality. These points are determined for automated image registration based on the modality.

The fingerprint contains such points, located on ridge curvatures, that are either unique or sharper than those in other areas. The core point, also known as the reference point, is often defined as the sharpest concave ridge curvature based on the orientation field of the fingerprint image [13]. This point is especially useful as it serves as a guide during image registration, which is important for normalizing features. The core point is used to allocate a ROI, which minimizes the discrepancy of stretch and alignment differences among multiple fingerprint samples of the same finger. Many previous approaches to reference point determination critically rely on the local features such as Poincaré index or other properties of the orientation field [13]. Poincaré works well in good quality fingerprint images, but fails to correctly localize reference points in poor quality fingerprints such as partials or fingerprints with poor ridge and valley contrast. The solution used in this paper applies an edge preserving non-local means (NL-means) filter [14] before applying Poincaré.

A ROI can also be determined for facial images. The key points are the eyes, nose and mouth, which are used during facial image registration. These points are used to create a border around the face, centred at the nose, which helps to avoid typical changes to the face such as different hair, occluded ears and neck. A large number of Haar-like features are organized to create a classifier cascade. Haar cascading is a popular method for detecting features that are used as key points. Multiple Haar cascades are iterated when the selected one fails to detect a key point.

The ROI for palmprint images is determined by extracting the principal lines and applying the iterative closest point method, which estimates the translation and rotation parameters between the input and test image by minimizing the distance between the two sets of correspondence points [15]. This method ensures that the key points between fingers, used as a boundary for the ROI, correspond between the two images.

A fall-back mechanism is required in case of image registration failure. The fall-back mechanism, used during key point detection failure of the fingerprint, face or palmprint, is determined by using the confidence score of the versatile local binary pattern histogram (LBPH) classification method [6]. LBPH is a feature descriptor for the texture of images. A basic local binary pattern (LBP) operator assigns a label to every pixel of an image by thresholding the 3×3 pixel neighbourhood, based on a lower or higher value than the centre pixel. A special kind of LBP operator called extended LBP (ELBP) is used for the fall-mechanism. Instead of being limited to directly adjacent neighbours, the neighbourhood is extended to include interpolated pixels, based on a circular mask, that capture fine texture. This operator uses spatially enhanced histogram matching that enables partial matching and automatic pixel normalization on a pixel level, circular neighbourhood level and image level. This results in the distinct advantage of improved illumination, scale and rotation invariance compared with other methods [16]. This gives a confidence score that can determine various attributes of an image. Training the spatially enhanced histogram model has a very low time complexity. Furthermore, the training time is independent of the image resolution and it produces a small model size. Given m circular neighbourhoods, their corresponding spatially enhanced histograms are determined, with a size of $m \times n$, where n is the length of a single histogram.

[1]Captured in a controlled environment with user-pegs – restricting the hand-pose and wrinkling.

C. Feature Selection

The following feature selection algorithm, proven in previous work [6], reduces the differences among multiple same-class data while improving the discrimination among different classes of data.

The quality of the input image plays an important role in the performance of the feature extraction and matching algorithm [17]. A bandpass filter can be used to increase the amplitude of the mean component of the image. This has the effect of increasing the dominant spectral components while attenuating the weak components.

The Laplacian of Gaussian (LOG) filter can remove unwanted features on the high frequency spectrum before enhancing the remaining features, effectively increasing the mean component. However, a side effect can occur when applying the filter to badly aligned images. The increased feature discrimination of LOG further highlights the difference among multiple samples, leading to bad training and testing sets and consequently a lower recognition accuracy. The image registration procedure in Subsection III-B is thus imperative to the success of an image-based feature-fusion biometric system.

Another problem is achieving consistent lighting. In [6], a modified circular local binary pattern (ELBP) texture descriptor is shown to achieve superior performance over the standard pixel normalization and histogram equalization methods. The LOG filter is thus combined with modified ELBP for a highly discriminative and robust feature selection algorithm.

D. Dynamically Weighted Fusion

This subsection explains the dynamic weighting process used for context switching between various combinations of processed modalities. The LBPH confidence score used in Subsection III-B is employed for gauging the modalities to be used in the dynamic weighting process. The main advantage of this approach is the reliable elimination of unrecognizable samples – effecting class separation and accuracy during the classification process after modality fusion.

1) If the quality of a test image of any modality is below the threshold determined by the LBPH confidence score, it is excluded from the fusion and classification process.
2) The remaining modalities are ranked from the highest to lowest LBPH confidence score.
3) Application defined thresholds are applied depending on the rank of the modality. As the number of modalities decrease, the confidence score required for passing increases. In other words, if only one modality remains, the highest confidence score is required for it to pass the test.
4) The remaining modalities are fused using column-wise feature-fusion concatenation. This stitches horizontal pixels in the image space of each modality.

Fig. 1. Feature-fusion image selected by dynamic weighting algorithm.

E. Feature Transformation and Eigen Classification

Image classification algorithms aim to express the most relevant image properties. Feature transformation is used to express a feature vector in an alternate space to improve discrimination. This often allows for a reduction in dimensionality and intra-class variation – resulting in near-identical relevant information across multiple samples. The following Eigen classification method is applied to the dynamically weighted feature-fusion image illustrated in Fig. 1 – in this example only the face and the fingerprint are selected.

Eigen classification can be used to maximize inter-class separation to discriminate effectively between different individuals. The Eigen classifier uses Principal Component Analysis to represent statistically key features that define a given feature set. An efficient model can be constructed from principal components, retaining key features of samples in one class. The distances among Eigenvalues are compared between the trained model and the model to be tested during matching.

Given N sample images x, the total scatter matrix is defined as [18]:

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

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

IV. Experimental Analysis and Results

The impostor samples were included in the experiments to comprehensively test the proposed methodology against false recognition.

The receiver operating characteristic (ROC) curve in Fig. 2 shows that the face significantly outperforms the fingerprint and palmprint using one training sample. The face and fingerprint multi-modal system achieves the best bi-modal recognition accuracy. The tri-modal system achieves a slight improvement over the best bi-modal system. The dynamically weighted multi-modal biometric system significantly outperforms all of the systems with a 93.75% genuine acceptance rate (GAR) at 1% false acceptance rate (FAR).

Fig. 3 illustrates the same comparison as the previous graph, but with an extra training sample. The GAR for

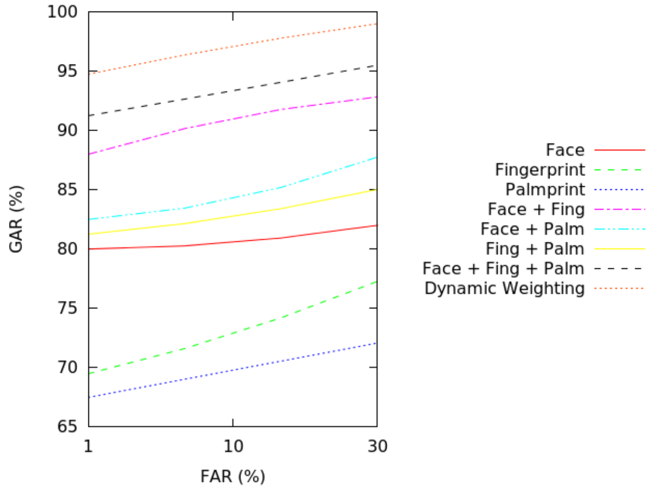

Fig. 2. Comparison of uni-modal, multi-modal and weighted systems using one training sample.

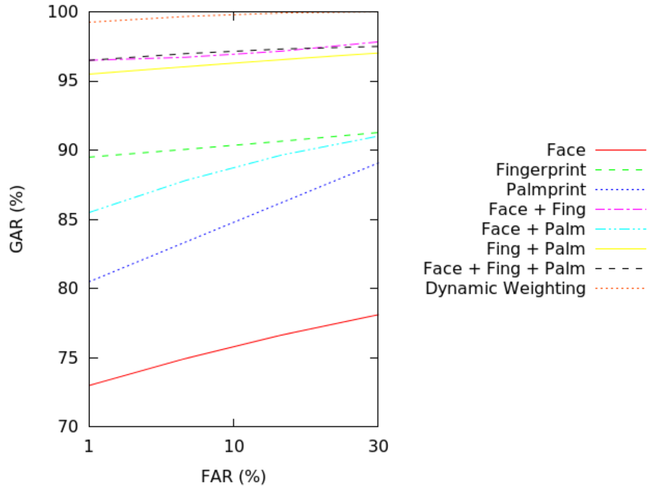

Fig. 4. Comparison of uni-modal, multi-modal and weighted systems using three training samples.

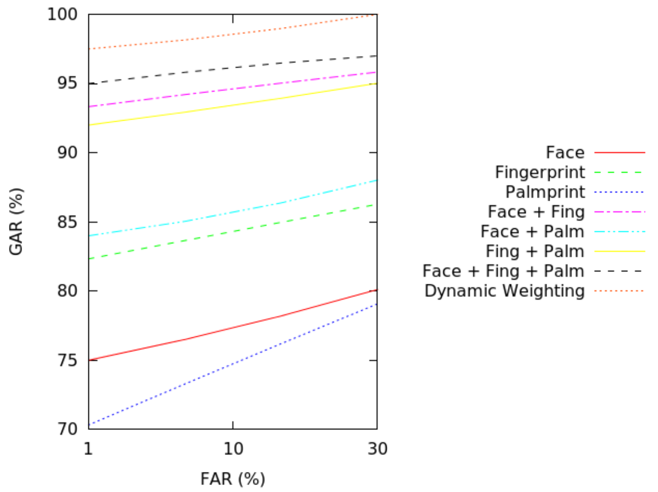

Fig. 3. Comparison of uni-modal, multi-modal and weighted systems using two training samples.

the face drops significantly, but is still higher than that for the palmprint. Upon manual inspection it was revealed that face samples four and five across all individuals were often falsely recognized. This was attributed to the severity of the occlusions caused by the props on many faces. The single fingerprint and the face and palm bi-modal system achieve similar accuracies. The weighted system's GAR improves by 3.75%.

The observed face accuracy drop is the most pronounced in Fig. 4, using three training samples. The single fingerprint surpasses the face and palmprint due to the poor performance of the face. The palmprint also improves significantly on its own. The remaining bi- and tri-modal systems all improve and achieve very similar accuracies. The weighted system clearly outperforms all other systems with 99.25% GAR and 1% FAR.

It is clear based on the three graphs that the weighted system improves the human identification accuracy by intelligently selecting the weight of one or more biometric modalities. This shows great promise for solving the universality problem and improving human identification in general.

V. Conclusion and Future Work

Security applications that are automatic or unsupervised in the real-world are still threatened by the continuous improvement in sensor penetration methods. Improving the security using the proposed dynamically weighted multi-modal biometric system effectively mitigates the pervasiveness of sensor data on networks in access control. Regardless of the application, user experience and security is significantly improved. Moreover, the proposed system provides a solution to the universality problem posed by many external image-based biometric modalities. A comparison was performed on uni-, bi- and tri-modal combinations of the face, fingerprint and palmprint. The popular PolyU palmprint dataset was first tested, resulting in perfect accuracy. The challenging IITD dataset was thus used for the experimental analysis. The systems in the comparison made use of a new uni-modal and multi-modal biometric algorithm designed based on a novel framework in previous work. The comparison was extended by including the novel dynamically weighted multi-modal biometric system proposed in this paper. The hard to identify samples in the face dataset showed the importance of the dynamic weighting of modalities. The challenging palmprint dataset further emphasized its importance. The fingerprint uni-modal system and face and fingerprint bi-modal system were the best combinations in their respective classes. The proposed weighted system significantly outperformed the rest of the systems with a 99.25% GAR and 1% FAR.

In future, more combinations of image-based biometric modalities will be investigated with additional experimentation. Live blood vessels can be detected from a distance for all external image-based modalities. This can be used in combination with the proposed system as an extra security mechanism in applications that require maximum security.

Acknowledgement

The authors would like to thank the CSIR, Information Security department for their financial support. Thank you to the authors of the publicly available datasets used in this paper. This work was undertaken in the Distributed Multimedia CoE at Rhodes University, with financial support from Telkom SA, Tellabs/CORIANT, Easttel, Bright Ideas 39, THRIP and NRF SA (UID 90243). The authors acknowledge that opinions, findings and conclusions or recommendations expressed here are those of the authors and that none of the above mentioned sponsors accept liability whatsoever in this regard.

References

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

[2] D. Kaur and G. Kaur, "Level of fusion in multimodal biometrics: A review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 2, pp. 242–246, 2013.

[3] 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.

[4] 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.

[5] 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.

[6] D. Brown and K. Bradshaw, "An investigation of face and fingerprint feature-fusion guidelines," in *Beyond Databases, Architectures and Structures. Advanced Technologies for Data Mining and Knowledge Discovery*. Springer, 2015, pp. 585–599.

[7] 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.

[8] 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.

[9] A. M. Martinez, "The ar face database," *CVC Technical Report*, vol. 24, 1998.

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

[11] A. Kumar, "Incorporating cohort information for reliable palmprint authentication," in *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*. IEEE, 2008, pp. 583–590.

[12] D. Han, Z. Guo, and D. Zhang, "Multispectral palmprint recognition using wavelet-based image fusion," in *Signal Processing, 2008. ICSP 2008. 9th International Conference on*. IEEE, 2008, pp. 2074–2077.

[13] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *Image Processing, IEEE Transactions on*, vol. 9, no. 5, pp. 846–859, 2000.

[14] 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.

[15] W. Li, B. Zhang, L. Zhang, and J. Yan, "Principal line-based alignment refinement for palmprint recognition," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 1491–1499, 2012.

[16] 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.

[17] S. Chikkerur, A. N. Cartwright, and V. Govindaraju, "Fingerprint enhancement using STFT analysis," *Pattern Recognition*, vol. 40, no. 1, pp. 198–211, 2007.

[18] 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.

Dane Brown is currently pursuing his PhD. in Computer Science at Rhodes University. The focus of his research is computer vision, information security and biometrics.

Karen Bradshaw obtained her PhD from Cambridge University and is an Associate Professor at Rhodes University, Department of Computer Science. Her research interests are computer vision, distributed computing and robotics frameworks.