Wavelet Domain Match Score Level Fusion for Multimodal Biometrics

Nicolas MORIZET, Graduate Student Member, IEEE
Department of Electronics, I.S.E.P., 21, rue d'Assas 75006 Paris
nicolas.morizet@isep.fr

Introduction

Multi-biometric systems have already proved to be able to outperform unibiometric systems by combining several modalities [1][2][3][4]. In those systems, the key issue is the fusion method. Among the different levels of fusion [5], match score level fusion has been very studied; most probably because fusing scores at that confidence level allows a parallel development of each unibiometric system and offers a good tradeoff between richness of information and ease of implementation. Besides, an advantage of fusion at this stage is that existing and proprietary biometric systems are not affected, allowing for a common middleware layer to handle the multimodal application but with a portion of information. In order to have a consistent fusion scheme, outputs of each individual matcher first have to be converted into a common format. Then, a score normalization process must be performed in order to have scores in the same range, such as MinMax, Z-Score, Tanh, adaptive normalizations [6]. Finally, a fusion rule is applied to normalized scores to obtain the fused ones. Many fusion techniques have been proposed so far: these methods include classification approaches such as neural networks, k-nearest neighbor, classification tree, SVM and some simple combination approaches such as min rule, max rule and product rules [7]. The purpose of this paper is to contribute a novel adaptive combination approach to match score level fusion, using wavelets and based on some key ideas of image processing and visual perception.

The remainder of this paper is organized as follows. In section 1, the basic theory of wavelet image fusion will be introduced. The use of chimeric users in multibiometrics will be discussed in section 2. An introduction to wavelet domain score fusion will be given in section 3 and the wavelet domain score fusion framework will be detailed in section 4. Then, some experiments and results will be described in sections 5 and 6 respectively. Finally, a conclusion will be drawn to show the advantages of the proposed method over several actual fusion techniques.

Wavelet image fusion

Wavelet Image Fusion (WIF) is an image fusion technique using the wavelet transform [8][9][10][11][12][13], a recent state of the art can be found in [14]. The main advantage of merging images in wavelet domain is that we can process various frequency ranges differently. To explain the WIF principle (Figure 1), let us consider two different images $I_1(x,y)$ and $I_2(x,y)$ to be fused into a new image. First, a two-dimensional discrete wavelet transform (2-D DWT, denoted as W), into L levels, is performed on both input images.

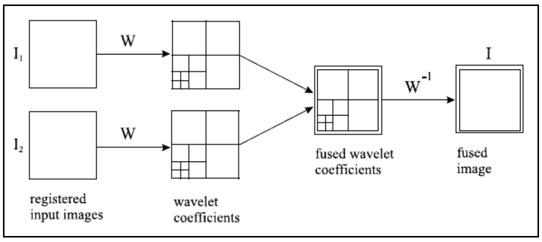

Figure 1. Wavelet Image Fusion (WIF) process

When an image is transformed into L levels, we get ($3L+1$) subbands, one approximation subband or baseband C_j of low frequency and $3L$ subbands D^h, D^v, D^d of high frequency. Let $f(x,y)$ be an original image, denoted by C_0, 2-D DWT can be performed as follows:

$$\begin{cases} C_{j+1} = HC_jH' \\ D^h_{j+1} = GC_jH' \\ D^v_{j+1} = HC_jG' \\ D^d_{j+1} = GC_jG' \end{cases}, \; (j = 0,1,\ldots,J-1) \qquad (1)$$

Reconstruction formula (2D-IDWT) is given by:

$$C_{j-1} = H'C_jH + G'D^h_jH + H'D^v_jG + G'D^d_jG$$
$$(j = J, J-1,\ldots,1) \qquad (2)$$

where h, v, d represents horizontal, vertical and diagonal respectively. H' and G' is conjugate transpose of H (low pass filter) and G (high pass filter) respectively. J is the decomposition level.
The process of decomposing an image with a 2-D DWT into L levels is called multiscale decomposition (MSD) or multiresolution analysis (MRA) [15].

Then, the transformed images are combined in wavelet domain using a defined fusion rule ϕ then transformed back to the spatial domain using a two-dimensional inverse discrete wavelet transform (2-D IDWT, denoted as W^{-1}) to give the resulting fused image $I(x,y)$ (3).

$$I(x,y) = W^{-1}\left\{\phi\left[W(I_1(x,y)), W(I_2(x,y))\right]\right\} \qquad (3)$$

Besides the multiscale decomposition analysis derived from the 2-D DWT, a key issue in MSD-based image fusion is how to form the fused MSD representation from the MSD representations of the source images.

The processing to achieve this goal is what we called a fusion rule. Some general alternatives for constructing a fusion rule are illustrated in Figure 2.

These include several choices:

- The choice of an activity level measurement: the activity level of an MSD coefficient reflects the local energy in the space spanned by the term in this expansion corresponding to this coefficient,

- A coefficient grouping method: the method to associate MSD coefficients with each other,

- A coefficient combining method: the method to produce the composite MSD representation, starting from the source MSD representations,

- A consistency verification method: consistency verification attempts to exploit the idea that it is very likely that a good fusion method will compute neighboring coefficients in the composite MSD in a similar manner. For Choose Max (CM) combining, consistency verification is especially simple; it ensures that a composite MSD coefficient does not come from a different source image from or all of its neighbors.

When making fusion decisions, one common method is to select the MSD coefficient with the larger activity level. This makes the assumption that a larger activity implies more information, which must be taken with care since it is not always true.

To explain the different alternatives available in forming a fusion rule, we make the assumption that there are just two source images, *X* and *Y*, and the fused image is *Z*. We note that all the methods described can also be extended to cases with more than two source images (which could correspond to more than two biometric modalities in our case). Generally, for an image *I* we denote the MSD representation as D_I and the activity level A_I.

Thus, we shall encounter D_X, D_Y, D_Z, A_X and A_Y. Let $\vec{p}=(m,n,k,l)$ indicates the index corresponding to a particular MSD coefficient, where *(m,n)* is the spatial position in a given frequency band, *k* the decomposition level and *l* the frequency band of the MSD representation. Thus, $D_I(\vec{p})$ and $A_I(\vec{p})$ are the MSD value and activity level of the corresponding coefficient respectively.

The *coefficient-based activity* (*CBA*) measures consider each coefficient separately. The activity level is described by the absolute value (4) or square of corresponding coefficient in the MSD representation.

$$A_I(\vec{p})=\left|D_I(\vec{p})\right|, \tag{4}$$

The following presents some selection schemes in wavelet domain:

- Minimum selection (*MinS*) scheme: This simple scheme just picks the coefficient in each subband with the *smallest magnitude*:

$$D_Z(\vec{p})=\min\left(\left|D_X(\vec{p})\right|,\left|D_Y(\vec{p})\right|\right), \tag{5}$$

- Maximum selection (*MaxS*) scheme: This simple scheme just picks the coefficient in each subband with the *largest magnitude:*

$$D_Z(\vec{p})=\max\left(\left|D_X(\vec{p})\right|,\left|D_Y(\vec{p})\right|\right), \tag{6}$$

- Average selection (*AveS*) scheme: This simple scheme just computes the mean of each pair of coefficients in each subband:

$$D_Z(\vec{p})=\frac{D_X(\vec{p})+D_Y(\vec{p})}{2}, \tag{7}$$

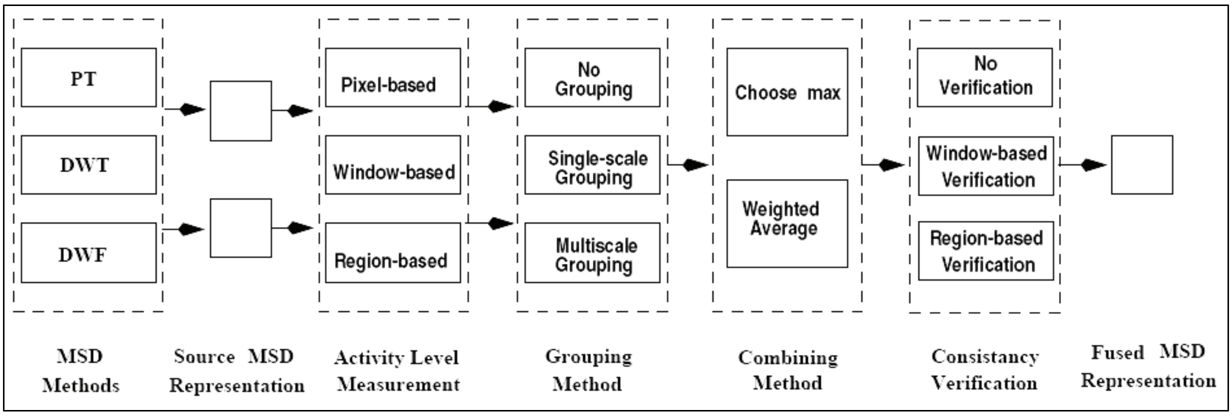

Figure 2. The generic framework of image fusion schemes

- Weighted average selection (*WaS*) scheme: This scheme developed by Burt and Kolczynski [21] uses a normalized correlation between two images' subbands over a small local area. The resultant coefficient for reconstruction is calculated from this measure via a weighted average of the two images' coefficients,

- Window-based verification (*WBV*) scheme: This scheme developed by Li et al. [9] creates a binary decision map to choose between each pair of coefficients using a majority filter.

On the use of chimeric users in multibiometrics

Combining biometric modalities of a database with biometric modalities of another database results in the creation of chimeric users (i.e. virtual subjects created with biometric traits of different users) [16][17]. This commonly used practice in the multimodal literature was questioned during the 2003 Workshop on Multimodal User Authentification [18].

At least two arguments (first one is technical, the second one is ethical) may justify the use of chimeric users:

- Modality independence assumption: two or more biometric traits (in our case, face and iris biometric modalities) of a single person are often assumed independent of each other (Definition 1, [19]). A demonstration using a correlation matrix with face and speech classifiers can be found in [1].

Definition 1. Assume *X* and *Y* are two mean-centered random variables. *X* and *Y* are said to be statistically independent if $E(XY) = E(X)E(Y)$.

- Privacy issue: participants in the multimodal biometric experiments are often not ready to let institutes keep record of too much of their personal information at the same place.

The use of chimeric users may overcome problems due to lack of real-user databases and it has been shown that generating multiple chimeric databases does not degrade nor improve the performance of a fusion operator when tested on a real-user database with respect to using only a real-user database [20].

Introduction to wavelet domain score fusion

In match score level, a similarity matrix contains scores, deriving from a biometric matcher, which express the similarity between a target set (persons that are known to the system, i.e. the database) and a query set (subjects that are to be compared against the gallery set) (Figure 3).

For a similarity matrix *S* with a generic element $A_{i,j}$, the value at (*i,j*) expresses the similarity between the i^{th} query subject and the j^{th} target subject.

The same subjects are in both sets, but with separate instances of their biometric signatures. From here, we can define two types of scores: genuine scores result from comparing elements in the target and query sets of the same subjects. Imposter scores are those resulting from comparisons of different subjects.

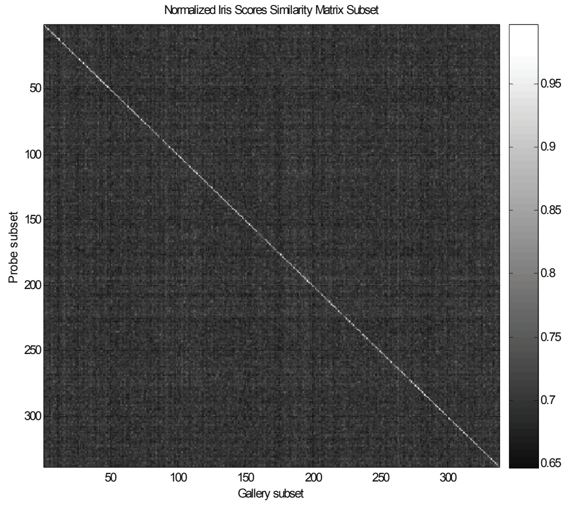

Figure 3. An example of a similarity matrix. Darker values represent weak similarity, brighter values indicate stronger similarity

Genuine scores should be as high as possible and imposter scores should be as low as possible. Thus, highest scores must appear along the diagonal of the similarity matrix and constitute what we called the genuine scores line (GSL). A perfect biometric system (100% recognition rate) should entirely separate genuine scores from imposter scores.

In our approach, scores row vectors of similarity matrices are transformed into images (with a simple 2-D mapping, see next section) to be fused in wavelet domain. Since higher scores are given to more similarity, a contrast enhancement operation must improve the recognition rate of the multi-biometric system, improving the separation between genuine and imposter scores.

A common issue in image processing when considering finite length signals such as images is known as board distortions or edge-effects. In our method, those signal-end effects are managed by half-point 2-D symmetric padding (Figure 4).

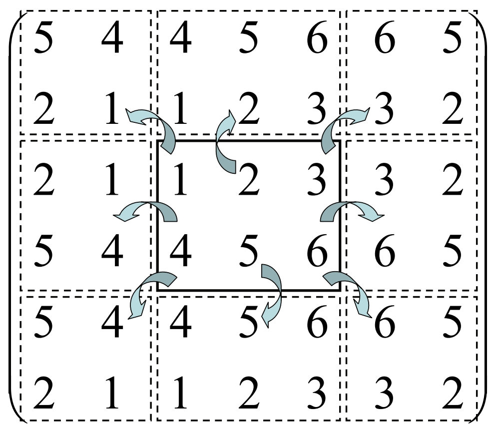

Figure 4. Illustration of an half-point 2-D symmetric padding on a (2x3) matrix

Since we are going to perform a MRA on images, time-scale wavelets have been focused on. We retained the Daubechies' wavelet which is an orthogonal compactly supported wavelet with highest number of vanishing moments for a given support width.
Since the more vanishing moments, the more efficient the contrast enhancement, a Daubechies' wavelet having a high order might be suitable for our application (Figure 5).

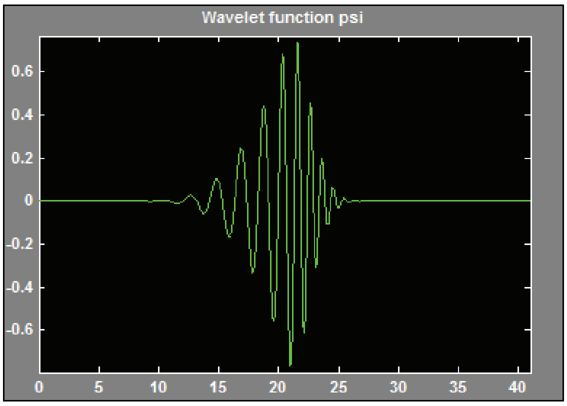

Figure 5. Daubechies' wavelet of order 19

Wavelet domain score fusion framework

The entire framework that is going to be described is depicted in Figure 6.
Let $V_{i,1}$ be the i^{th} row vector of scores (row vector of scores for the i^{th} probe subject) deriving from the first similarity matrix. A simple 2-D mapping is performed on $V_{i,1}$, being reshaped into a square matrix: assume M is the length of $V_{i,1}$, the side c of the square is defined such as:

$$\begin{cases} c = \sqrt{M} & \text{, if } M \text{ is a perfect square} \\ c = \left\lceil \sqrt{M} \right\rceil & \text{, otherwise} \end{cases}$$

That operation minimizes the area of the square matrix for computation purpose. When M is not a perfect square, a zero-padding is applied when filling the end of the matrix. Zero-padding is suitable here, since we work with similarity scores; no additional information is added or subtracted. The number N_z of zeros in the zero-padding operation is simply $N_z = c^2 - M$.

Thus, a square matrix is constructed by vertically concatenating scores blocks of size of c.

The same process is performed on $V_{i,2}$ which is the i^{th} row vector of scores deriving from the second similarity matrix. Then, following the wavelet image fusion scheme (Figure 1), a 2-D DWT is performed on each image matrix with a Daubechies' wavelet, into L levels, where L takes value between [1, L_{max}].
The maximum decomposition level L_{max} of an image of size of (S x S) is computed as follows:

$$L_{\max} = \max_{L} \left\{ 2^L \leq S \right\}, \quad L \in N^{*+} \qquad (8)$$

Once the decomposition in wavelet domain has been done, wavelet coefficients in each subband are combined according to a fusion rule ϕ to obtain a fused scores matrix.

Concerning the fusion rule, *MinS* scheme is chosen for approximation coefficients (representing low frequencies in the 2-D DWT decomposition, assimilated to imposter scores). Besides, highest scores from each row vector of scores are thought to be at the same location (thanks to the structure of similarity matrices) and wavelet coefficients having large absolute values contain the information about the salient features of the images such as edges and lines. Therefore, *MaxS* scheme is chosen for details coefficients (representing high frequencies in the 2-D DWT decomposition, assimilated to genuine scores), in order to reinforce those maximum values.

Then, a 2-D IDWT is performed to recover our fused scores matrix in the spatial (score) domain. Finally, that matrix is reshaped into a fused row vector of scores (inverse 2-D mapping) and a vector cropping is performed in order to recover the length of each initial row vector of scores (we simply remove the N_z final values).

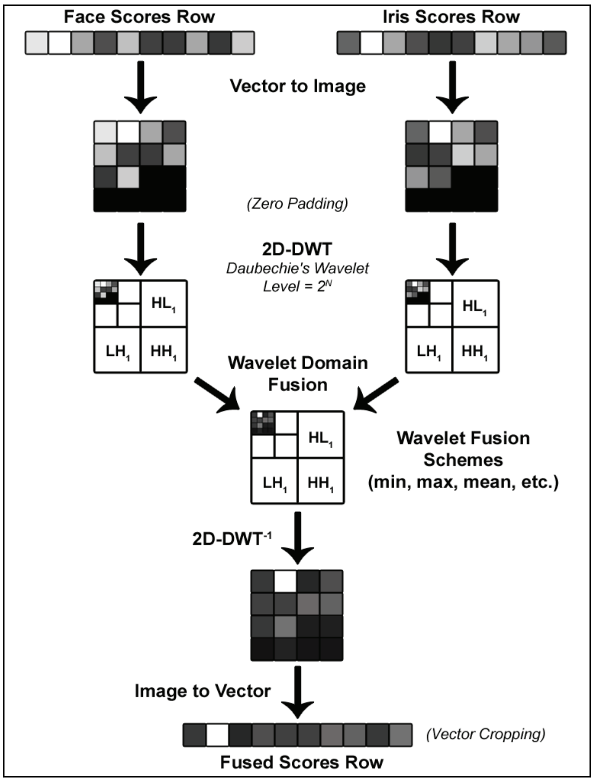

Figure 6. Wavelet domain score fusion framework. (e.g. 2^{nd} row of scores of similarity matrices)

The result is a fused similarity matrix where genuine scores have been increased and imposter scores have been decreased at the same time, leading to a better separation between genuine and imposter scores, hence a better recognition rate of the multimodal system. This process can be seen as a contrast enhancement operation (Figure 6).

Experiments

For our experiments, we used two very well-known databases: FERET face database [22] and CASIA V3 iris database [27]. We precisely followed the testing framework described in [6], allowing the conduction of normalization and fusion technique evaluations. This framework consists of the following steps:

1. Assemble two sets of biometric signatures: a target set and a query set. For practical tests, the intersection of these two sets should not be null.
2. Obtain a match-score for each pair of query and target signatures and store in a similarity matrix, whose size is query set size by target set size.
3. Extract gallery set (any arbitrary subset of the target set) and probe set (any arbitrary subset of the query set) to perform "virtual" experiments on a subset of the population.
4. Repeat steps 1-3 for each biometric mode.
5. Assemble and align the similarity matrices from step 2; this includes converting data to a common format (distance or similarity measure), forming subsets to obtain matrices of the same size, and data mating to create real or virtual subjects, relying upon the assumption that the individual modalities concerned are statistically independent of one another and could thus be assigned arbitrarily (though consistently) to form a set of mated virtual subjects for the purpose of testing.
6. Score normalization. Assemble similarity matrices to a common number range.
7. Score fusion. Fuse the set of normalized similarity matrices into a single fusion similarity matrix.
8. Compute performance statistics (e.g. ROC curve) for verification from the genuine and imposter scores. Use each fusion score as a threshold and compute the false-accept rate (FAR) and false-reject rate (FRR) by selecting those imposter scores and genuine scores, respectively, on the wrong side of this threshold and divide by the total number of scores used in the test. A mapping table of the threshold values and the corresponding error rates (FAR and FRR) are stored. The complement of the FRR (1 – FRR) is the genuine accept-rate (GAR). The GAR and the FAR are plotted against each other to yield a ROC curve, a common system performance measure.

The face similarity matrix (FSM) has been constructed from the FERET face database [22], using the FA Gallery Set and the FB Probe Set. Face preprocessing has been performed according to the CSU System 5.0 [23]. Then, normalized face images have been compressed with a 2-D DWT into two levels, using a biorthogonal wavelet ('bior1.3'). Finally, Log-Gabor PCA (LG-PCA) [24] face recognition algorithm has been implemented, which is based on basic PCA [25]. To generate match-scores, the Mahalanobis distance metric [26] (turned into a similarity distance) has been computed. Finally, the FSM is (*1193x1193*).

The iris similarity matrix (ISM) has been generated from the CASIA V3 iris database [27]. The iris preprocessing and signatures generation algorithms that have been used are based on [28]. Iris preprocessing consists of iris segmentation, unwrapping and histogram equalization (occlusions have not been taken into account). Iris signatures are two vectors of size of (*16x32*), deriving from a 2-D wavelet packets analysis, keeping the second and the tenth wavelet packets. We have generated 338 (out of 396) different classes, five iris signatures per subject (three for the gallery, two for the probe).

In order to compensate eye rotation effect (from -15° to +15°), 31 circular shifts (almost one shift per angle unit) are performed on each probe iris signature before computing a combined cosine similarity measure [28] with all gallery signatures. Thus, for each probe and for all the computed similarity measures, the best similarity measure is kept as the match-score. This leads to an iris similarity matrix of size of (676x1014). In order to be consistent with a real system, we randomly kept one row out of two (random probe subject) and one column out of three (the best corresponding gallery match), leading to a final ISM that has been reduced to (338*x*338).

Since the FSM is bigger than the ISM, a sliding method, along the genuine scores line (Figure 3), of three different (338x338) submatrices deriving from the FSM has been achieved to consistently fuse scores, row by row (Figure 6).

Assume X is the mean-centered face similarity matrix and Y is the mean-centered iris similarity matrix.

According to *Definition 1*, we found $E(XY) \approx E(X)\,E(Y)$, with an error $\varepsilon=10^{-4}$. This result confirms the assumption that the individual modalities concerned are statistically independent of one another and thus can be used for creating "virtual users".

Results

We present here some mean EERs, since we have three face similarity submatrices to be fused tree times with a single one iris similarity matrix. We compare our proposed fusion technique with some actual fusion rules, for several normalization methods. The LG-PCA face recognition algorithm has been used for face modality.

For our wavelet fusion rule, according to Figure 2, we performed a DWT as the MSD method with a Daubechies' wavelet of order 19 ('db19') and two decomposition levels. The absolute value of wavelet coefficients as been used as the activity level measurement (pixel-based level) combined with a multiscale grouping method. *MinS* scheme has been chosen for approximation coefficients and *MaxS* scheme has been selected for details coefficients. Nevertheless, no verification method has been implemented. Tests were conducted on a multimodal database of 338 subjects with a sliding method for the face similarity matrix.

Table 1. Mean EER (%), LG-PCA algorithm

Normalization Method	**Fusion Rule**			
	Min	Max	Sum	Proposed
MinMax	2.96	0.28	0.04	**0.02**
Z-Score	1.43	0.22	**0.03**	0.09
QLQ	1.88	0.19	0.11	**0.08**

We can first notice that our wavelet fusion method is the best fusion rule using MinMax normalization. However, Z-Score normalization appears not to be very suitable for our method. An interesting result is that proposed wavelet fusion rule outperforms “QLQ–Sum” combination, which is considered as one of the most efficient technique in multimodal fusion [29].

Another interesting result is to observe the Mean ROC curves, comparing face modality only, iris modality only and our fusion method combining face and iris modalities (Figure 7). We can clearly see that fusion really outperforms both face and iris modalities.

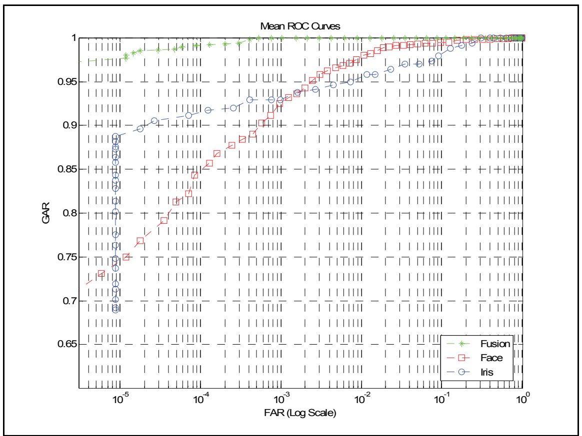

Figure 7. Mean ROC Curves. Face modality only (squares), iris modality only (circles) and fusion of iris and face using MinMax normalization and wavelet domain scores fusion (asterisks).

For a FAR=10^{-3}% (a common chosen operational point), face and iris modalities give nearly 92% recognition rate whereas our fusion method reaches 100% recognition rate.

Conclusions

First, recent score normalization and fusion rules have been introduced and wavelet image fusion theory has been explained. Then, we proposed a novel combination technique called wavelet domain scores fusion, trough a precise framework. This adaptive technique is based on a visual perception principle (contrast enhancement) and takes advantages of merging images in wavelet domain to process various frequency ranges differently. Finally, a testing framework for multimodal databases has been used to present some encouraging results. Our proposed method does not need any threshold and can be easily implemented on an embedded hardware since it mainly uses discrete wavelet transforms. Wavelet domain scores fusion appears as a flourishing candidate over actual combination approaches, as underlined by the promising results. Some objectives of the future work should be to test some local activity level measurements, implement a verification method, construct a bigger multimodal database and study the robustness of our system to noise.

Acknowledgements

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. Portions of the research in this paper use the CASIA-IrisV3 collected by the Chinese Academy of Sciences Institute of Automation (CASIA). The author would like to thank B. Giraud for interesting technical discussions and Y. Meyer for invaluable theoretical advices in numerous work sessions.

References

[1] Kittler, J., Hatef, M., P.W. Duin, R., and Matas, J. (1998). “On combining classifiers”. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 20(3):226-239.
[2] Wang, Y., Tan, T., Jain, A.K., (2003). “Combining face and Iris biometrics for identity verification”. In AVBPA’03.
[3] Wang, Y., Wang, Y., Tan, T., (2004). “Combining fingerprint and voiceprint biometrics for identity verification: an experimental comparison”. In Proc. ICBA’04.
[4] Shu, C., and Ding, X., (2006). “Multi-biometrics fusion for identity verification”. In ICPR’06.
[5] Nandakumar, K., (2005). “Integration of multiple cues in biometric systems”. M. Sc. Thesis, p.29.
[6] Snelick, R., Indovina, M., Yen, J., Mink, A., (2003). “Multimodal biometrics: issues in design and testing”. In ICMI’03, pp.68-72, New York, USA.
[7] Ma, Y., Cukic, B., Singh, H., (2005). “A classification approach to multi-biometric score fusion”. In AVBPA’05.
[8] Nikolov, S., Hill, P., Bull, D., Canagarajah, N., (2001). “Wavelets for image fusion”. Wavelets in signal and image analysis, from theory to practice”, chapter 1, Kluwer Academic Publishers.

[9] Li, H., Manjunath, B.S., Mitra, S.K., (1995). "Multi-sensor image fusion using the wavelet transform". Graphical Models and Image Processing, vol.57, pp. 235-245.
[10] Zhang, Z., and Blum, R.S., (1999). "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application". In Proc IEEE, 87(8):1315-1326.
[11] Santos, M., Pajares, G., Portela, M., de la Cruz, J.M., (2003). "A new wavelets image fusion strategy". In IbPRIA 2003, pp. 919-926, Puerto de Andratx, Mallorca, Spain.
[12] Chao, W., and Zhong-Fu, Y., (2007). "Perceptual contrast-based image fusion: a variational approach". Acta Automatica Sinica, 33(2):132-137.
[13] Hill, P., Canagarajah, N., Bull, D., (2002). "Image fusion using complex wavelets". In BMVC'02, Cardiff, UK.
[14] Dasarathy, B.V., (2007). "Image fusion: advances in the state of the art". Guest editorial, 8(2):114-118.
[15] Mallat, S.G., (1999). "A Wavelet Tour of Signal processing". Academic Press.
[16] Dorizzi, B., Garcia-Salicetti, S., Allano, L., (2006). "Multimodality in biosecure: evaluation on real vs. virtual subjects". In ICASSP'06: IEEE International Conference on Acoustics, Speech and Signal Processing, vol.V.
[17] Garcia-Salicetti, S., Mellakh, M.A., Allano, L., Dorizzi, B., (2005). "A generic protocol for multibiometric systems evaluation on virtual and real subjects". In AVBPA'05.
[18] Dugelay, J.-L., et al., (2003). In MMUA'03: Workshop on Multimodal User Authentification, Santa Barbara, USA.
[19] Feller, W., (1971). "An introduction to probability theory and its applications", vol.2, John Wiley & Sons.
[20] Poh, N., and Bengio, S., (2006). "Using chimeric users to construct fusion classifiers in biometric authentification tasks: an investigation". In ICASSP'06: IEEE International Conference on Acoustics, Speech and Signal Processing, vol.V, pp.1077-1080.
[21] Burt, P.J. and Kolczynski, R.J., (1993). "Enhanced image capture through fusion". In Proc. ICCV'93: the 4th International Conference on Computer Vision, pp.173-182.
[22] Phillips, P. J., et al., (1996). "The FERET september 1996 database and evaluation procedure".
[23] The CSU System 5.0 user's guide (2003): http://www.cs.colostate.edu/evalfacerec/algorithms/version5/faceIdUsersGuide.pdf.
[24] Perlibakas, V., (2005). "Face recognition using principal component analysis and log-Gabor filters". Unpublished manuscript.
[25] Turk, M., and Pentland, A., (1991). "Eigenfaces for Recognition". Journal of Cognitive Neuroscience, 1(1):71-86.
[26] The Mahalanobis distance, internet address: http://en.wikipedia.org/wiki/Mahalanobis_distance.
[27] The CASIA Database (2003), internet address: http://www.cbsr.ia.ac.cn/irisdatabase.htm.
[28] Rydgren, E., Ea, T., Amiel, F., Rossant, F., Amara, A., (2004). "Iris features extraction using wavelet packets". In ICIP'04, pp.861–864.
[29] Ross., A. A., et al., (2006). "Handbook of multibiometrics". Springer-Verlag, New York, USA.