# MULTI LIBRARY WAVELET NEURAL NETWORKS FOR 3D FACE RECOGNITION USING 3D FACIAL SHAPE REPRESENTATION

*Wael BEN SOLTANA, Wajdi BELLIL, Chokri BEN AMAR and Adel M.ALIMI*

REsearch Group on Intelligent Machines (REGIM)
University of Sfax National Engineering School of Sfax,
B.P. W, 3038, Sfax, Tunisia
phone: + (216) 74274088 (#527), fax: + (216) 74275595, email: wael.bensoltana@ieee.org, wajdi.bellil@ieee.org, Chokri.benamar@ieee.org and adel.alimi@ieee.org, web: http://www.regim.org

## ABSTRACT

*This paper presents a new approach for 3D face modeling and recognition. Motivated by finding a representation that embodies a high power of discrimination between face classes, a new type of 3D shape descriptors is suggested. We have developed a fully automatic system which uses an alignment algorithm to register 3D facial scans. In addition, scalability in both time and space is achieved by converting 3D facial scans into compact wavelet metadata. Our system consists in two phases. The first phase is called enrolment composed of 3 steps: data processing, alignment and metadata generating. The metadata generating step is powered by the use of Multi Library Wavelet Neural Networks (MLWNN). The second phase is called Authentication it starts with the calculation of depth distances between a probe and gallery 3D face. A K-Nearest Neighbors (K-NN) technique is used for 3D face classification. The results of this contribution are more interesting, in comparison with some others works, in term of recognition rate using the GavabDB 3D facial database.*

## 1. INTRODUCTION

Automated Face Recognition is the process of determining a subject's identity from digital imagery of their face without user intervention. The term in fact encompasses two distinct tasks; Face Verification is the process of verifying a subject's claimed identity while Face Identification involves selected the most likely identity from a database from subjects. This dissertation focuses on the task of face verification, which has a myriad of applications in security ranging from border control to personal banking.

Recently, the use of 3D facial imagery has found favor in the research community due to its inherent robustness to the pose and illumination variations which plague the 2D modality. The field of 3D face recognition is, however, yet to fully mature and there remain many unanswered research questions particular to the modality. The relative expense and specialty of 3D acquisition devices also means that the availability of databases of 3D face imagery lags significantly behind that of standard 2D face images. Human recognition of faces is rooted an inherently 2D visual system and much is known regarding the use of 2D image information in the recognition of the individuals. The corresponding knowledge of how discriminative information is distributed in the 3D modality is much less well defined.

Various algorithms have been proposed in literature based on wavelets to address at least some of these problems in 3D face recognition. Wang and Chua [5] present a system based on both 3D range data as well as the corresponding 2D gray-level facial images. To extract view-invariant features, a rotation invariant 3D spherical Gabor filter (3D SGF) is proposed. Furthermore, a 2D Gabor histogram is employed to represent the Gabor responses of the 3D SGF. To match a given test face with each model face, the Least Trimmed Square Hausdorff Distance is employed. Moreover, Kakadiaris and al. [6] describe facial data using a deformed facial model. The deformed model captures the details of an individual's face and represents this 3D geometry information in an efficient 2D structure by utilizing the model's UV. This structure is analyzed in the wavelet domain and the spectral coefficients define the final metadata that are used for comparison among different subjects. In this section, we present here a small sample of relevant work that is not meant to be exhaustive.

As is well known traditional methods for 3D face recognition are marked with their robust less to facial expression. In this paper, we propose a new method based on MLWNN, in order to well describe 3D faces and guarantee the invariance of the generated model. In the first step, a preprocessing phase consists to search a template face mask in order to localize the face region in the range data. This process is shown in figure 1. In the next step, we compute 4 feature vectors from the frontal face region: Points clouds, surface normal, MLWNN model for Z-information (Z_MLWNN) and MLWNN model for central profile representation (P_MLWNN).

For recognition, we use a hierarchical architecture to filter out the most similar K classes (Persons) using K-NN on Depth Euclidean Distance and then to feed these K classes into a second classifier based on parallel architecture using feature vectors. Compared to other approaches for 3D face recognition, such as [9, 10], our approach leads to a better recognition rate. Rather than matching the entire surfaces of two faces (the probe image and a gallery image), we match both the global feature : Points cloud and Surface normal, and the local feature : Z_MLWNN and P_MLWNN.

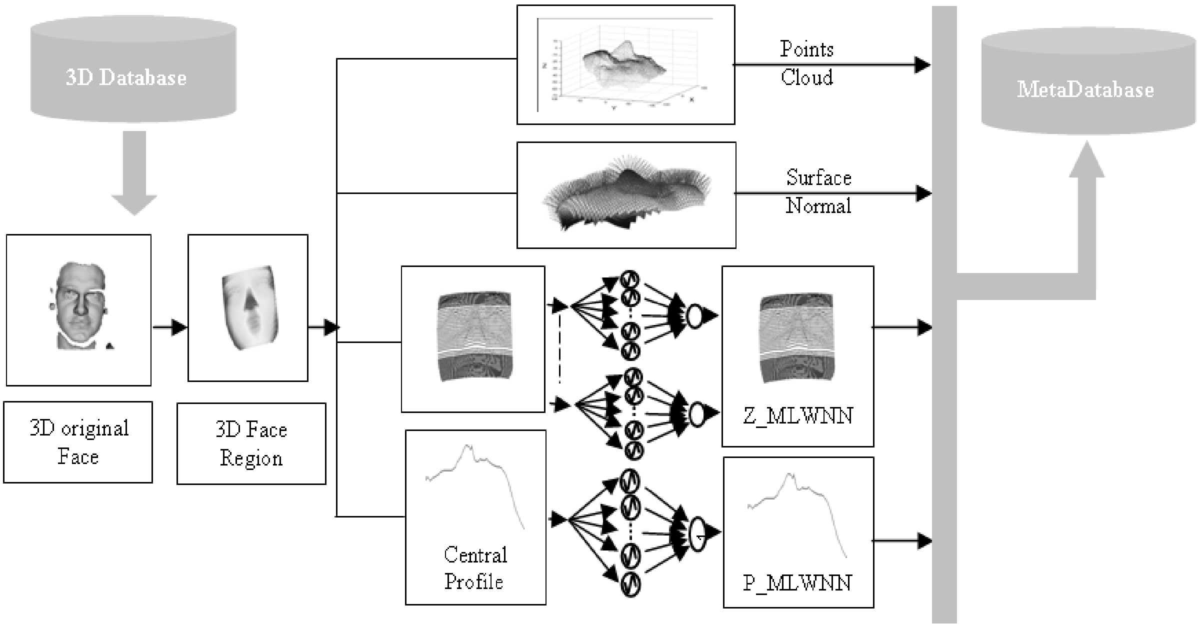


Figure 1- Block diagram of our system for 3D face recognition from range data

This paper is organized as follow: Section 2 discusses the optimization of the proposed approach. Section 3 gives tutorials examples and we compare the performance of our models with the performance of previously published examples. Finally, section 4 gives conclusions for the present research work and others possibilities of future research directions.

## 2. OPTIMIZING APPROACH

The aim is to represent facial data using shape feature extraction and a wavelet neural network. The wavelet neural network captures the details of an individual's face and represents this 3D geometry information in an efficient 2D architecture by utilizing the Multi Library Wavelet Neural Networks parameterization. Our face recognition procedure can be divided in two phases, enrollment and authentication.

### 2.1 Enrollment

*2.2.1 Data Preprocessing*

The purpose of pre-processing is to minimize the impact of the quality of input data. This stage consists to search a template face mask in order to localize the 3D face region in the range data. The template face mask is created according to all faces in the database.

*2.2.2 Alignment*

Before the rigid registration in performed on the faces the centre of mass of all faces is moved to the origin of the coordinate system. This compensates for large differences in the distance between subjects.

*2.1.3 Metadata Generating*

The metadata vector is composed of 4 features: Points clouds, surface normal, MLWNN model for Z-information (Z_MLWNN) and MLWNN model for central profile representation (P_MLWNN). Concerning the new model based on MLWNN is presented in section 2.3. The reader may consult the paper [7] for further details concerning MLWNN. The relevant information of 3D face is encoded on a 2D grid and focused through the Z-axis. In fact, this information is subdivided in horizontal lines. Each line is considered as an entity which will be analyzed (approximated) using the MLWNN. For each 3D face a specific wavelet neural network parameters are generated. This analyze is applied also in the central profile information. The structural parameters of the network (weights, dilations, and translations), Points clouds and surface normal define the metadata that are used for comparison among different subjects.

### 2.2 Authentication

The first process computes Depth Euclidean Distance DED:

$$\mathrm{DED} = \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\overline{I_{i,j}} - I_{i,j}\right)^2}}{\sqrt{\sum_{i=1}^{N}\sum_{J=1}^{M}\left(\overline{I_{i,j}}\right)^2}} \qquad (1)$$

With $\overline{I_{i,j}}$ $I_{i,j}$ are image depth of gallery and probe 3D face image. This distance function (eq. 1) can be viewed as a discrete approximation of the volumetric difference between two 3D facial surfaces (gallery and probe face). Potential 3D face candidate is obtained. The second phase consists in calculating the Euclidean distance between the query 3D face and others 3D faces in the gallery for Points clouds, surface normal, Z_MLWNN and P_MLWNN. Each feature provides a nearest a 3D face from the gallery. Then, the face having the highest vote is declared as the final opinion.

### 2.3 MLWNN for 3D face modeling

*2.3.1 Classical wavelet neural network architecture*

Wavelets occur in family of functions and each is defined by dilation $a_i$ which controls the scaling parameter and translation $t_i$ which controls the position of a single function, named the mother wavelet $\psi(x)$. Mapping functions to a time-frequency phase space, WNN can reflect the time-frequency properties of function. Given an n-element training set, the overall response of a WNN is (eq. 2):

$$\hat{y}(x) = \sum_{i=1}^{Np} w_i \psi_i \quad where\ \psi_i = \Psi\left(\frac{x-t_i}{a_i}\right) \qquad (2)$$

Where $N_p$ is the number of wavelet nodes in the hidden layer and $w_i$ is the synaptic weight of WNN. This can also be considered as the decomposition of a function in a weighted sum of wavelets, where each weight $w_i$ is proportional to the wavelet coefficient scaled and shifted by $a_i$ and $t_i$.

*2.3.2 Multi Library Wavelet Neural Network Model*

A MLWNN can be regarded as a function approximator which estimates an unknown functional mapping (eq. 3):

$$y = f(x) + \varepsilon \qquad (3)$$

Where f is the regression function and the error term ε is a zero-mean random variable of disturbance. Constructing a MLWNN involves two stages: First, we should construct a wavelet library W= {W1, W2,…, Wn}of discretely dilated and translated versions of some mothers wavelets function Ψ1, Ψ2,…, Ψn. Then select the best M wavelets based on the training data from multi wavelet library W, in order to build the regression (eq. 4).

$$\hat{y}(x) = \sum_{i\in I} w_i \Psi_i^1(x) + \sum_{i\in I} w_i \Psi_i^2(x) + \ldots + \sum_{i\in I} w_i \Psi_i^n(x) \qquad (4)$$

This network can be considered composed of three layers: a layer with $N_i$ inputs, a hidden layer with $N_p$ wavelets and an output linear neuron receiving the weighted outputs of wavelets. Both input and output layers are fully connected to the hidden layer. To make use of MLWNN we used a selection method to initialize the translation and dilation parameters of wavelet networks trained using gradient-based techniques.

## 3. PERFORMANCE AND EVALUATIONS

We used the GavabDB [8] 3D face database for automatic facial recognition experiments. The database GavabDB contains 427 images of 3D meshes of the facial surface. These meshes correspond to 61 different individuals (45 male and 16 female), and 9 three dimensional images are provided for each person. The total of the database individuals are Caucasian and their age is between 18 and 40 years old. Each image is a mesh of connected 3D points of the facial surface without the texture information for the points. The database provides systematic variations in the pose and the facial expressions of the individuals. In particular, there are 2 frontal views and 4 images with small rotations and without facial expressions and 3 frontal images that present different facial expressions.

In our experiments, we used the two neutral frontal images and the three frontal images with smile, accentuated laugh and random gesture expressions. One neutral frontal image is used as gallery, and the other four images, the second frontal neutral and the three frontal images with different expressions are used as probe images.

So, we have 61 frontal faces used as gallery model and the 244 faces different from the gallery faces are used for the evaluation of the proposed approach. We used two simulations examples using the GavabDB.

In the first example we try to examine and compare the capacity of the MLWNN and traditional one for 3D face modeling. The second example is to evaluate the performance of the proposed approach for 3D face recognition. The measurement for the quality of reconstructed faces is the common use of the normalized mean-square error (NMSE) given by the equation 5:

$$\text{NMSE} = \frac{\frac{1}{N*M} * \sum_{i=1}^{N}\sum_{j=1}^{M}\left(Z_{i,j} - \overline{Z_{i,j}}\right)^2}{\sum_{i=1}^{N}\sum_{J=1}^{M}\left(Z_{i,j}\right)^2} \qquad (5)$$

Where $\overline{Z_{i,j}}$ *and* $Z_{i,j}$ are the values of Z points in the reconstructed face and the original face, respectively. N and M are the row and column of the Z information respectively.

### 3.1 MLWNN vs Traditional WNN

We used 61 frontal faces with two different dimensions: 20*20 and 50*50. The mean NMSE are given in Table 1 for 5, 20, 40 and 61 faces. For example, the mean NMSE is equal to $6.35*10^{-6}$, when using MLWNN on 5 faces of 20*20 pixels, much smaller than $3.52*10^{-5}$ using Traditional WNN (TWNN). However, for 61 faces of 50*50 pixels the mean NMSE is $7.91*10^{-6}$ using MLWNN and $1.31*10^{-5}$ via TWNN.

Table 1-Variation of NMSE for 3D face modeling

| | **Face 20*20** | | **Face 50*50** | |
|---|---|---|---|---|
| | **TWNN** | **MLWNN** | **TWNN** | **MLWNN** |
| **5 faces** | $3.52*10^{-5}$ | $6.35*10^{-6}$ | $0.90*10^{-5}$ | $5.69*10^{-6}$ |
| **20 faces** | $4.55*10^{-5}$ | $9.02*10^{-6}$ | $1.13*10^{-5}$ | $6.83*10^{-6}$ |
| **40 faces** | $4.94*10^{-5}$ | $9.15*10^{-6}$ | $1.25*10^{-5}$ | $7.46*10^{-6}$ |
| **61 faces** | $5.42*10^{-5}$ | $9.17*10^{-6}$ | $1.31*10^{-5}$ | $7.91*10^{-6}$ |

From this table, we can see clearly that the proposed MLWNN leads to the best model in term of NMSE because using multi-regressors introduces a best approximation capacity.

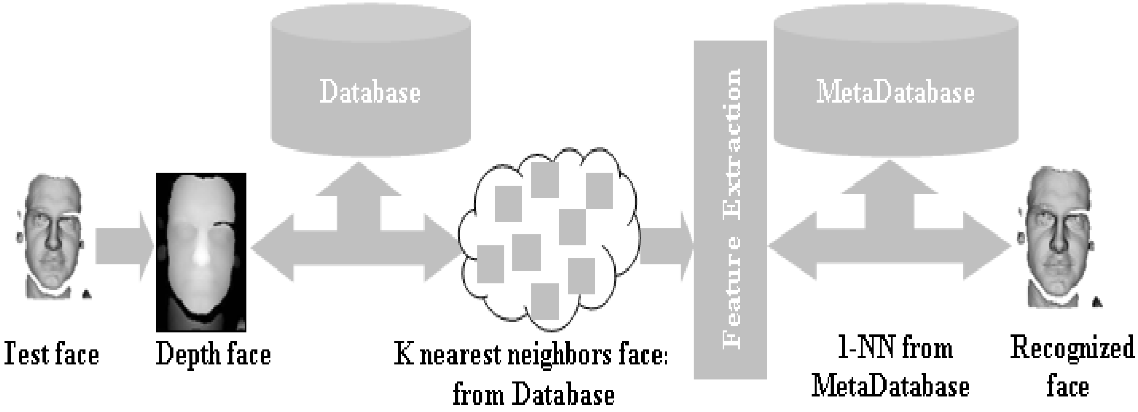


Figure 2-3D face Recognition System

### 3.2 Experimental Result of 3D Face Recognition System

Two different approaches for 3D face recognition were presented by Moreno et al. in [9] and Mahoor et al. in [10] and both were evaluated using GavabDB. In [9], they selected a set of 30 features and using Support Vector Machines (SVM). In [10] they generated a 3D binary image which shows the locations of the ridge lines in the range facial image and using LTS-Hausdorff distance.
We were used the same conditions for simulations and applying our approach presented in figure 2. In our approach, after calculating the depth distances we choose the 10 nearest 3D face from the database. Then, we extract the feature vectors. Afterwards, we use the consensus voting [3] based on K-NN technique in order to identify the corresponding 3D face. The performance is generally measured in term of Recognition Rate. The recognition rate is showed in Table2. It may be observed from this table that the recognition rate obtained on neutral 3D faces (Frontal) is 68.97% using the proposed approach without MLWNN; it is equal to 94.83% with MLWNN, 93.5% for Mahoor approach [10] and 90.16% for Moreno approach [9]. Someone can see the superiority in term of rate recognition of the proposed method including MLWNN in feature vector.

Table 2 -Comparison of the proposed approach with other works (Recognition Rate)

| | **Our algorithm** | | **Mahoor and al.** | **Moreno and al.** |
|---|---|---|---|---|
| | **With MLWNN** | **Without MLWNN** | | |
| **Frontal** | 94.83 | 68.97 | 93.5 | 90.16 |
| **Smile** | 80 | 21.81 | 82 | 77.86 |
| **Accentuated laugh** | 67.86 | 35.71 | - | - |
| **Random gesture** | 50 | 15.91 | - | - |

In other cases, the recognition rate decrease with the presence of 3D face expression. The best rate is obtained using Mahoor approach 82% our proposed approach gives a rate of 80% which is better than Moreno 77.86%. With Accentuated laugh the recognition rate is equal to 67.86% and it is equal to 50% with Random gesture.

## 4. CONCLUSION AND FUTURE WORKS

We presented a model based on 3D Facial Shape Face Representation and MLWNN for 3D Face Recognition. A fully automatic alignment algorithm and the advanced wavelet analysis, the Multi Library Wavelet Neural Network, resulted in robust state-of-the-art. We demonstrated by simulations that MLWNN introduces precisions in the field of 3D face recognition because Wavelet Networks present immunity on geometric transformations such as rotations, translations and dilatations. The result of this contribution is a solution for 3D face modeling and recognition that achieves the highest accuracy on 3D face recognition rate.
The system consists in two phases: enrolment and authentication. The first phase is composed of 3 steps: data processing, alignment and metadata generating. The last step computes 4 feature vectors from the face region: Points clouds, surface normal, MLWNN model for Z-information and MLWNN model for central profile representation. The second phase starts with the calculation of depth distances between a probe and gallery 3D face. A consensus voting scheme based on K-Nearest Neighbors technique is used in order to identify the corresponding 3D face.
As future work, we propose to introduce genetic algorithm in the decision phase instead of KNN technique. Moreover, we try to improve the recognition rate under expression and pose variations. Also, we will use other databases such as Face Recognition Grand Challenge (FRGC).

## 5. REFERENCES

[1] V. Blanz, “Face recognition based on a 3D morphable model,” *7th International Conference on Automatic Face and Gesture Recognition,* Page(s):617– 624, 10-12 April 2006.
[2] A. Rama, F. Tarres, D. Onofrio, and S. Tubaro, “Mixed 2D-3D Information for Pose Estimation and Face Recognition,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, Volume 2, Page(s):II-361 - II-364, 14-19 May 2006.

[3] B. Gökberk, M.O. İrfanoğlu and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing*, Volume 24, Issue 8, Pages 857-869, 1 August 2006.
[4] L. Chao, A. Barreto, J. Zhai and C. Chin, "Exploring face recognition by combining 3D profiles and contours," *Proceedings IEEE SoutheastCon*, Page(s):576 – 579, 8-10 April 2005.
[5] Y. Wang and C.S. Chua, "Robust face recognition from 2D and 3D images using structural Hausdorff distance," *Image and Vision Computing*, Volume 24, Issue 2, Pages176-185, 1 February 2006.
[6] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis, "Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* VOL. 29, NO. 4, pp: 640-649, APRIL 2007.
[7] W. Bellil, M. Othmani, C. Ben Amar and M.A. Alimi, "*A New Algorithm for Initialization and Training of Beta Multi-Library Wavelet Neural Networks,*" Advances in Robotics, Automation and Control, ISBN: 78-953-7619-16-9, Edition in Tech October 2008.
[8] A.B. Moreno and A. Sánchez, "GavabDB: a 3D Face Database," *Workshop on Biometrics on the Internet COST275,* Vigo, 2004, 77-85, March 25-26.
[9] A.B. Moreno, A. Sanchez, J. Velez, J. Diaz, "Face recognition using 3D local geometrical features: PCA vs. SVM," *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis,* Page(s):185, 15-17 Sept. 2005.
[10] M.H. Mahoor, M. Abdel-Mottaleb, "3D Face Recognition based on 3D rigide lines in range data,"*ICIP*,2007.