3D FACIAL EXPRESSION ANALYSIS BY USING 2D AND 3D WAVELET TRANSFORMS

S. C. D. Pinto[1], *J. P. Mena-Chalco*[1], *F. M. Lopes*[1,2], *L. Velho*[3] *and R. M. Cesar Junior*[1]

[1] Institute of Mathematics and Statistics, University of São Paulo, Brazil
[2] Federal University of Technology - Paraná, Brazil
[3] National Institute of Pure and Applied Mathematics, Brazil

ABSTRACT

This work presents a new approach for the 3D human facial expressions analysis. Our methodology is based on 2D and 3D wavelet transforms, which are used to estimate multi-scale features from real a face acquired by a 3D scanner. The proposed methodology starts by considering a dataset composed by faces displaying seven different facial expressions. An automatic pre-processing method, adopting an ellipsoidal cropping, is applied to the dataset. Thereafter, the 2D and 3D descriptors are extracted from different scales of wavelet transforms for the purpose of obtaining the facial expression features. The multi-scale features are represented in a multivariate feature space, which is analysed by the Sequential Forward Floating Selection algorithm using an entropy criterion function to select the subset of features that best represents each facial expression model. The obtained results corroborate the potential of multi-scale feature extraction for analysis of 3D facial expression.

Index Terms— Facial Expressions Analysis, Wavelet Transform, 3D Face.

1. INTRODUCTION

Although some aspects of the expressions are culturally determined, there are some universally recognized basic expressions. As proposed by Paul Ekman's Facial Action Coding System [1], there are six main facial expressions: joy, sadness, fear, anger, surprise and disgust. Facial expression analysis is a topic that has attracted much interest among researchers, being widely studied and having many applications, especially in the human-computer interaction and in the entertainment field. Therefore, it is of great importance to develop reliable and efficient methods of facial expression analysis.

Many techniques for 2D facial expression recognition have been proposed in the literature, but most of them suffer from limitations of 2D image acquisition, such as head pose variations and illumination changes. In order to circumvent these problems, some techniques have been proposed incorporating 3D information of facial surface. Tralakanidou *et al.* [2] combine 3D face geometry and 2D appearance data information to facial action unit detection and facial expression recognition in sequences of 2D and 3D images. Wang and Lien [3] use a 3D virtual facial model to obtain accurate estimates of the head rotation angle directions in order to improve in the 2D facial expression recognition process. In the case of 3D facial images acquired by a 3D scanner, great effort has been directed for face recognition task [4], but few studies have addressed the issue of recognition of 3D facial expressions. Mpiperis *et al.* [5] propose bilinear models for joint 3D identity-invariant facial expression recognition and an expression-invariant face recognition. Unlike the methodology applied to 3D face recognition that seeks for face regions that remain unaffected by facial expressions [6], our goal is to identify and to analyse the geometric deformations caused by facial expressions.

The main contributions of this work can be summarized as follows. First, we introduce the approach based on wavelet descriptors as a multi-scale tool to analyse 3D facial expressions. The potential of the wavelet transform arises from its capabilities for detecting and characterizing singularities, extracting instantaneous frequencies and producing multi-scale measures [7]. Second, we propose an automatic pre-processing method to normalize the faces, i.e., to detect and crop the facial region. This approach uses an ellipsoidal cropping through the detection of facial landmarks. Finally, aiming at classifying the facial expressions, a feature selection technique based on the Sequential Forward Floating Selection algorithm and an entropy criterion function was adopted [8]. To achieve success in this approach it is essential to work with a dataset that represents reliably the differences in facial expressions. Therefore, we created a 3D face dataset with different facial expressions.

2. METHODOLOGY

2.1. DATASET

The dataset was created for the analysis of facial expressions and reconstruction of 3D faces. This dataset was acquired using a non-contact 3D scanner Konica Minolta Vivid 910. The scanner is composed by a laser sensor and a digital camera video. The data is hence composed by registered texture and geometry data. The output mesh contains approximately 60K

The authors are grateful to CAPES, FAPESP, CNPq and FINEP.

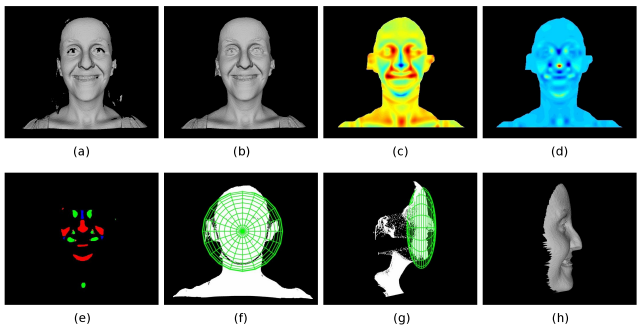


Fig. 1. 3D face normalization.

vertices. Texture images have been acquired with a resolution of 640×480 pixels (24 bits). The data was obtained from 10 subjects, for each subject there are samples of one neutral face and six main facial expressions, resulting in a total of 70 samples. The subjects are 5 women and 5 men, with most subjects aged between 19 and 27. Faces with occlusion were not considered.

2.2. 3D face normalization

The purpose of this routine is to automatically determine which region of the 3D face is interesting to treat the facial expressions, eliminating the regions such as the neck, shoulder, hair, etc. From the facial geometry data (obtained through a 3D scanner), which is represented as range images, the 3D face (Figure 1(a)) can be represented as a surface function $z = f(x,y)$.

The normalization routine is developed as follows. First, we applied a procedure called Sorted Exact Distance Representation [9], in order to fill holes. This procedure uses the data structure of exact dilation for locating the k nearest points in the original data and the z value is set as the average calculated from these k points (Figure 1(b)).

Second, in order to detect the facial landmarks, the curvature of 3D faces is computed through of a surface S representation defined by a twice differentiable real valued function. For every $(x,y,z) \subseteq S$ we consider the mean (H) and the Gaussian (H) curvatures [10](Figure 1(c) and (d), respectively).

$$H = \frac{(1+{f_y}^2)f_{xx} - 2f_x f_y f_{xy} + (1+{f_x}^2)f_{yy}}{2(1+{f_x}^2+{f_y}^2)^{\frac{3}{2}}} \quad (1)$$

$$K = \frac{f_{xx}f_{yy} - {f_{xy}}^2}{(1+{f_x}^2+{f_y}^2)^2} \quad (2)$$

where f_x, f_y, f_{xy}, f_{xx}, f_{yy} are the first and second derivatives of f. Thus, the partial derivatives are numerically estimated for each point (i,j) on the grid, through a paraboloid approximation of the surface. Since the second derivative is sensitive to noise a Gaussian filter is applied to the range image so that it can smooth the surface. This filter is applied before computing the curvature. In order to obtain a description of the local behavior of the surface, the mean and the Gaussian curvature values were used into the HK classification [10] of the points of the surface (Figure 1(e)). The regions with high curvature were isolated using predefined threshold values. In [10] the considered threshold values were $Th = 0.04$, $Tk = 0.0005$. The nose position is obtained as the point with the maximum value of the Gaussian curvature among the elliptical convex regions. The eyes position was obtained as the points with maximum values of the mean curvature among the elliptical concave regions. Thus, the centroid of the 3D face is defined as the (x,y) value of the nose position. The z value is estimated with the mean of the eyes positions.

The last step, an ellipsoid of radius (R_x,R_y,R_z) centred at the face centroid (see Figure 1(f-g)) is used to crop the 3D face. R_x, R_y and R_z are dependent on the locations of characteristic points (inner corners of eyes and nose tip) of each individual. Thus, in general, the radius is not defined in absolute value but relative. Figure 1(h) shows an example of normalized 3D face.

2.3. Extraction of facial features using wavelets

The wavelet transform (WT) is a signal processing tool that has been successfully applied to many practical situations, from bioinformatics to the analysis of astronomical images. In this work, we have explored 2D and 3D WT. The continuous wavelet transform of a signal $f(t)$ is defined as [9]:

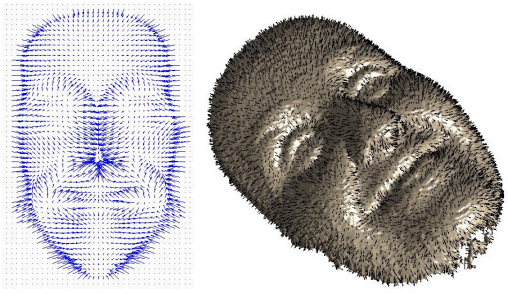
$$W_\psi(b,a) = \frac{1}{\sqrt{a}} \int \psi^*\left(\frac{t-b}{a}\right) f(t)dt \quad (3)$$

where ψ, b and a stand for the mother wavelet, the translation vector and the scale parameter, respectively. Furthermore ψ^* denotes the complex conjugate of the mother wavelet ψ. Different mother wavelets can be used, depending on the type of information expected to be extracted from the signal. Differential wavelets, like the derivatives of the Gaussian, are particularly suitable for analysing signal singularities and extracting differential information from the signal [11].

As presented in the preprocessing section, the 3D face can be represented as a surface given by the function $z = f(x,y)$. Hence, we have worked with the first partial derivatives with respect to x and y of the a 2D Gaussian function $g(x,y) = e^{-(x^2+y^2)}$ denoted as ψ_x and ψ_y [9]. These functions are adopted as the mother wavelets to extract differential shape measures from the 3D surface, and their wavelet transforms are denoted as W_{ψ_x} and W_{ψ_y} respectively. So, we applied 2D WT to the extrapolated surface to calculate the wavelet gradient ∇_W, i.e.,

$$\nabla_W = \left(W_{\psi_x}, W_{\psi_y}\right) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right). \quad (4)$$

Figure 2(a) shows the gradient field obtained for a given expression. Two particularly relevant features invariant to translation and rotation are then extracted from the gradient: the gradient mean magnitude $\overline{\nabla}_W$; and the dispersion of the gradient orientation, i.e., in order to quantify this dispersion, we have calculated the entropy of the orientation distribution $\mathbb{E}_{\nabla_W}$ [9].



(a) (b)

Fig. 2. The gradient field in (a) and the normal field in (b) of the 3D face.

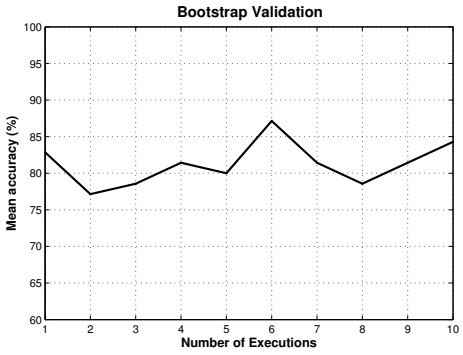


Fig. 3. Bootstrap validation by using 10 executions.

However, since we are dealing with 3D faces, not all facial expressions characteristics can be described by 2D features. Therefore, an important alternative is to treat the faces as a volumetric shape represented by the expression $w = f(x,y,z)$ [7]. We explore, again, the first partial derivatives of the Gaussian function as mother wavelets, but using the 3D Gaussian function $g(x,y,z) = e^{-(x^2+y^2+z^2)}$. Therefore, we have ψ_x, ψ_y and ψ_z and consequently, the definition of the wavelet transforms W_{ψ_x}, W_{ψ_y} and W_{ψ_z}. From the capabilities of the 3D WT to perform numerical differentiation, we calculate the normal field of the analysed 3D shape. Figure 2(b) shows the normal field of the 3D face.

$$N(x,y,z) \simeq [W_{\psi_x}, W_{\psi_y}, W_{\psi_z}] \tag{5}$$

Once the normal field is obtained, it is important to get their most important geometrical properties in terms of a set of meaningful features. Our goal is to characterize differences between the geometrical properties of facial deformation of interest across individuals.

Therefore, by taking into account that the selected features should be invariant to rotation and translation, we worked with the analysis of the local orientation distributions, i.e., density histograms were obtained for the orientation of the normal vectors along the 26-neighborhood around each surface element. More specifically, for each point (x,y,z) of the surface, the mean, maximum and minimum values of the angles θ defined between the normal at each surface point (x,y,z) and the normals at each surface neighbor point are obtained, as well as the respective histograms. Global features are extracted from three histograms in order to reduce the dimensionality of the classification space. These 15 features include the mean value and the central moments from second to fifth order [9]. Therefore, by combining the 2D and 3D features, we have a total of 17 features for each analysed scale.

2.4. Feature Selection

The feature selection methods attempt to select a subset of features, which produces a good description of the entire dataset and provides the classification or prediction of the classes present in the dataset.

In general, feature selection methods are based on a search algorithm and a criterion function. The optimal search methods return the best set of features, but their computational cost are excessive for many problems, such as the present work. In order to reach a good cost-benefit between computational cost and quality of the solution, the Sequential Forward Floating Selection (SFFS) algorithm [12] was adopted and the conditional entropy was used as a criterion function to assign a quality value to the subset of features [8].

In the interest of achieving unbiased results avoiding highly variability and due the small number of samples available it was adopted the bootstrap validation method [13].

3. RESULTS

The results presented were obtained by using the feature selection software [8] and its default parameter values. The software was applied to select the subset of features that best classify each facial expressions by considering just 10 different subjects. In other words, the training set is composed by 10 instances of each of the 7 facial expression and 102 features, i.e., 17 features by 6 different scales of 2D and 3D WT.

The SFFS algorithm with the entropy criterion function was explored. Figure 3 shows the bootstrap validation results using 10 executions, which presents good accuracy rates in view of the small number of facial expressions samples available. The most frequent selected features were the mean of the medium θ ($a = 0.001$) and mean of the minimum θ ($a = 0.002$) of the normal field . These features achieved the best classification results during the validation process.

In order to investigate the accuracy rates, Table 1 presents the frequencies of classification by considering each facial expression individually and its classification results obtained from bootstrap validation with 10 executions. Note that, Joy and Anger expressions can be recognized with high accuracy, i.e., 90%, but Surprise and Fear expressions are easily confused with others ($73\% - 74\%$). The case where Sur-

Table 1. Frequencies of classification errors by considering each facial expression individually, obtained from validation results.

	Classification Results Frequency						
	neutral	joy	sadness	surprise	anger	disgust	fear
neutral	**79**	1	5	3	0	10	2
joy	2	**90**	0	0	2	3	3
sadness	5	0	**84**	2	3	0	6
surprise	6	13	2	**73**	0	1	5
anger	0	5	5	0	**90**	0	0
disgust	6	13	0	1	0	**79**	1
fear	2	11	7	4	0	2	**74**

prise, Disgust and Fear expressions are confused with Joy (13%, 13% and 11% respectively) can be explained because these expressions have a higher variance as expressed by 10 subjects, explaining a greater similarity in the visual appearance of these expressions.

4. CONCLUSIONS

Despite the proposed algorithm has been tested on a small database, only ten samples of each facial expression, the experimental results achieve 81%, in average, of correct facial expression detection. While this result seems only reasonable we believe that our methodology can achieve a suitable classification accuracy, i.e., we believe that our methodology can be successfully applied to other datasets as BU-3DFE [14] and Bosphorus [15], and also in the current dataset witch we are working on increasing the samples (subjects).

Regarding the multi-scale features, the subset that presented the best classification of facial expressions was composed by the mean of the medium θ ($a = 0.0001$) and the mean of the minimum θ ($a = 0.002$) of the normal field. It is worth noting that the features order 2 central moment of the medium θ ($a = 0.00025$) and order 2 central moment of the maximum θ ($a = 0.00025$) of the normal field, and entropy of the gradient orientation distribution ($a = 0.0005$) also appeared among the selected features, but with lower frequency. This shows that the combination of the extracted features at different scales led to better identification and characterization of the facial expressions.

A future extension in this work is to include local measures of the faces, i.e. applying the methodology for features extraction using wavelet transform in interest regions where larger deformation caused by certain expressions appears.

5. REFERENCES

[1] P. Ekman and W. V. Friesen, "The facial action coding system: A technique for the measurement of facial movement," in *Consulting Psychologists*, 1978.

[2] F. Tsalakanidou and S. Malassiotis, "Real-time 2D+3D facial action and expression recognition," *Pattern Recognition*, vol. 43, no. 5, pp. 1763–1775, 2010.

[3] T. Wang and J. J. Lien, "Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation," *Pattern Recognition*, vol. 42, no. 5, pp. 962–977, 2009.

[4] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3-d face recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 38, no. 1, pp. 155–173, Feb. 2008.

[5] I. Mpiperis, S. Malassiotis, and M.G. Strintzis, "Bilinear models for 3D face and facial expression recognition," *Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 498–511, Sep. 2008.

[6] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," in *CVPR*, Miami, 2009, IEEE, pp. 2575–2582.

[7] S. C. D. Pinto, R. M. Cesar-Jr, D. Gokcay, and L. da F. Costa, "Characterization of neuroanatomic structures using 3D wavelet-based normal fields," in *ISSPA*, Paris, 2003, IEEE, pp. 479–482.

[8] F. M. Lopes, D. C. Martins-Jr, and R. M. Cesar-Jr, "Feature selection environment for genomic applications," *BMC Bioinformatics*, vol. 9, no. 1, pp. 451, Oct. 2008.

[9] L. F. Costa, "Robust skeletonization through exact euclidean distance transform and its application to neuromorphometry," *Real-Time Imaging*, vol. 6, no. 6, pp. 415–431, Dec. 2000.

[10] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognition*, vol. 39, no. 3, pp. 444–455, Mar. 2006.

[11] A. Grossmann, "Wavelet transforms and edge detection," in *Stochastic Processes in Physics and Engineering*, pp. 149–157. D. Reidel Publishing Company, 1988.

[12] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature-selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[13] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, 1995, vol. 14, pp. 1137–1145.

[14] L.J. Yin, X.Z. Wei, Y. Sun, J. Wang, and M.J. Rosato, "A 3D facial expression database for facial behavior research," in *FGR*, 2006, pp. 211–216.

[15] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Celiktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," *Biometrics and Identity Management*, pp. 47–56, 2008.