

A Hybrid System Based on Wrinkles Shapes and Biometric Distances for Emotion Recognition

Rim Afdhal, Ridha Ejbali and Mourad Zaied

REsearch Groups on Intelligent Machines, University of Sfax
Sfax, Tunisia

Emails: {eng.afdhal.rim, ridha_ejbali, mourad.zaied}@ieee.org

Abstract— Communication has an important role in human interaction. It can be vocal, bodily, textual or emotional in order to help everyone to understand others and be understood. Generally, all these types of communication aim at expressing a thought or an idea via a set of selected words, gestures, sounds or emotions. Emotions are so important to express ourselves and understand the others even without words or sounds. This paper presents a hybrid emotion recognition system based on wavelet network using 1D fast wavelet transform. The proposed application is based on two approaches. The first one is based on the shapes of the wrinkles; the second is based on the biometric distances. We combine these two approaches in order to ameliorate the classification rates. The rates given by experimental results show the effectiveness of our proposed system.

Keywords- hybrid emotion recognition system; wavelet network; fast wavelet transform; shapes of the wrinkles; classification.

I. INTRODUCTION

Emotions play an important role in our existence [1]. It is an important field, which plays a relevant role in several studies. Let us take the example of Human-Computer Interaction (HCI), which has confined its researchers to the development of techniques founded on the use of screen-keyboard-mouse triplet.

Today, the user must make a progress without obstacles in its natural environment; the finger, the hand, the face or the familiar objects are regarded as an input / output device; the border between the electronic and physical worlds tends to be blurred. These new forms of interaction usually require the capture of the observable behavior of the user and his environment. They rely on artificial perception techniques, including computer vision.

Future generations of man-machine environment are going to be multimodal by integrating new information, from taking account of dynamic behavior, from speech and facial expressions, in order to make the use of the most intuitive and natural machinery. Through human-machine interaction we try to get an idea of the emotional state of the user for ergonomic interface conception and have a better feedback. Measuring the gaze direction of the user could be an effective way to perform certain tasks in graphical interface (such as selection of a window or a text box). Expressions (defined as muscle movements) along with the spoken language, joined both in terms of physical movement required to speech (movements

of lips), and in terms of emotional indicator accompanying the spoken language. They express a non-negligible part of meaning in oral communication.

This work presents a hybrid emotion recognition system based on the analysis of the shapes of the wrinkles as well as the biometric distances. To start with, we choose to analyze the facial expressions because the face is the most expressive and communicative part of a human being. In order to boost the classification rates gotten through the approach of the wrinkles, we combined the wrinkles approach with the approach of biometric distances inspired from an automatic emotion recognition approach [1] based on the facial expression.

The method based on the shapes of the wrinkles can be summarized in four steps. The first step consists of detecting the elements of the face using Viola and Jones method. The second step is used to locate the region of the wrinkles. The third step of our system consists of extracting information. The last one is the classification which is based on wavelet network using Fast Wavelet Transform (FWT) [1][2][3].

The approach of biometric distances, which is well explained in [5] can be summarized also in four main steps: detection of the elements of the face, localization of the characteristic points, tracking of features points and classification .

This paper is divided into three parts. The first part presents our proposed emotion recognition system. The second part focuses on the approach based on the shapes of the wrinkles and its different phases. In the third part, experiments are made to demonstrate the efficiency of the proposed approach by using the Chon-Kanade dataset. We finish the paper with the conclusion and the future work.

II. STATE OF THE ART

This section presents the different existing approaches in the literature corresponding to this area of research. There are several emotions recognition methods among which we cite emotion recognition by body gesture analysis [10] [11] [12].

The approach presented in [10] aims at analyzing the actions of a person in order to determine his emotional state. The analysis was based on the position and the movements of the upper part of the body (hands and head). It is restricted to the trajectory and velocity of points. These characteristic points make a triangle whose perimeter provides information about the qualitative aspects of their movement based on the

approach proposed by Camuri [13]. The analysis of a movement will provide a set of qualitative aspects. In other words, the transformation of the physical measures (the position, velocity and acceleration of the superior body parts) in a high level model such as righteousness, impulsiveness and fluency provides these qualitative aspects that allow their turn to recognize a particular emotion.

We can also mention the emotion recognition methods based on analyzing words [14] [15] [16]. Thanks to automatic speech emotion recognition systems, the machine becomes able to transform a signal into a sequence of words. But we must go further and learn the meaning of the word sequences and be aware of the context of the sentence pronunciation. It is at this level that the emotional dimension is involved. So we must take into consideration the intonation of the sentence in order to make the difference between a statement and a question. In addition to these approaches, many approaches were based on facial expression analysis [17][18][19][20].

This paper presents an emotion recognition system based on the analysis of the shapes of the wrinkles. We adopt this approach because the face is the most expressive and communicative part of a human being. Facial expression is a kind of visible manifestation of a spirit state, of cognitive activities, of physiological activities (tiredness, pain), of the character and the psychopathology of someone. Psychology researches have shown that facial expression plays an important role in the coordination of human conversation, and have a great influence on the listener than the textual content of the message expressed.

III. THE PROPOSED EMOTION RECOGNITION SYSTEM

This section presents our emotion recognition system. The system is the combination of two approaches the first one is the wrinkles approach [6][7], the second is an approach, which is based on the biometric distances [5].

We proposed this system in order to boost the classification rates of the emotion recognition system based on the shapes of the wrinkles, so, we decided to add the approach of the biometric distance to enhance the classification rates. Fig. 1 describes the proposed system.

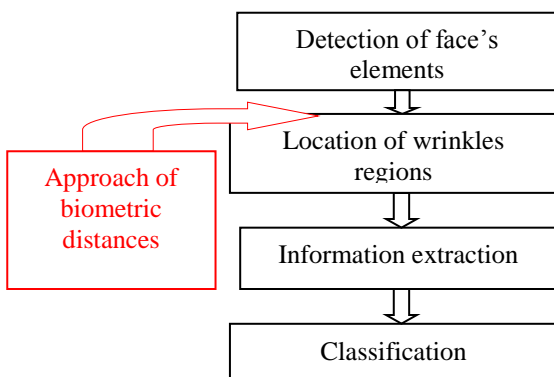


Figure1. Proposed emotion recognition system

A. The wrinkles approach

This approach has four stages: the detection of face's elements, the location of wrinkles regions, the information extraction and the classification.

The first phase aims at detecting the face as well as its different elements (eyes and mouth) using the viola and Jones detector. Figure 3 shows the detection of the face and its elements of the images of the Chon-Kanade databasis.

The second phase consists in locating 7 rectangles in the most important wrinkles in the face: a rectangle on the forehead, another on the chin, 2 rectangles on the corners of the eyes, 2 rectangles on the corners of the mouth and a rectangle on the upper of the nose.

The location of these regions will be at neutral state and will be relocated during the emotion. To achieve this step, we have developed an automatic method using the coordinates of the rectangles located by the Viola and Jones detector.

The third step of this approach is information extraction. The information will be extracted from the wrinkles regions by calculating the edge pixels number of each facial region expression and the neutral state. The difference between the two states will be calculated in order to prepare the training distances.

The last stage is the classification. We will use the wavelet networks [5] in order to recognize the basic emotions. This stage contains 4 steps. The first step is supposed to prepare the wavelet as well as the scaling functions. In the second step, we compute the weights by FWT [27] [28] then we compute the contributions from each library function. The last step's target is to choose the best features that best approximate the vector at the output of the network by setting a stopping criterion. At the end, we get the weight vector corresponding to the best contributions of every learning vector.

B. Wavelet network theory

“Wavelet networks” is a new theory which was introduced by Zhang and Benveniste in 1992 [14]. They used a combination of artificial neural networks based on radial basis, function and wavelet decomposition. Moreover, these researchers have explained how a wavelet network can be generated. It is defined by pondering a set of wavelets dilated and translated from one mother wavelet with weight values to approximate a given signal f. Eq. (1.1) represents the output of the network using a finite number of wavelets.

$$\hat{f} = \sum_{i=1}^n \omega_i \psi_i \quad 1.1$$

Fig. 2 shows an example of 1D neuron from the wavelet network. In order to extend the following architecture we can add dilated and translated scaling function's versions of the corresponding used wavelet in the hidden layer of the network.

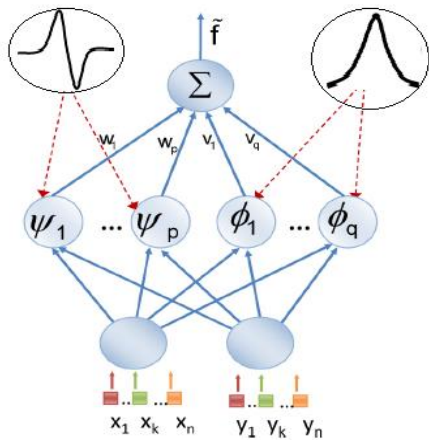


Figure 2. Wavelet network using scaling function

Many researchers [25][26] have used the projection technique of the signal f on the dual basis of the wavelets and the hidden layer's scaling functions in order to calculate the output weight connections of the wavelet network. This type of technique offers precise weights' values, but it has a main defect when it comes to determining the hidden layer's weights to the output layer, because it leads to calculating the matrix's inversion Φ which requires an intensive calculation as the matrix is so large. Our technique is based on the FWT. The wavelet networks are not only simple but also rapid and strong. The FWT uses a simple and a fast technique in order to facilitate the calculation of the approximation and the details. The classification phase aims at creating a wavelet modeling every vector of pixels of learning.

In order to create the network of every vector, we have, first, to prepare the wavelet and the scaling functions. Then, we calculate the weights by FWT [4][21][22][23][24] and the contributions from the library function. After that, we choose the best features by setting a stopping criterion. Then, we will get the vector of weights of every learning vector. After the training, the test step determines the appropriate class of the test vector. Consequently, every test vector will be projected on the network of all the training vectors in order to get its weight. Moreover, we calculate the distance between the vector of weight of the training and the test. In addition, the distances will be sorted.

At the end, the algorithm recognizes the suitable class of the vector of the test which has the smallest distance.



Figure 3. Detection of the face and its elements

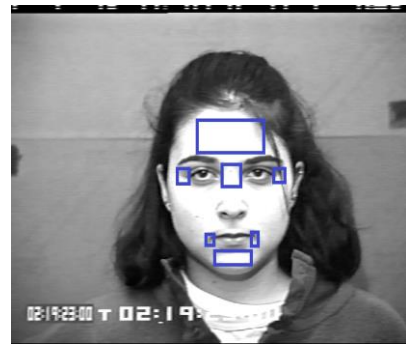


Figure 4. Location of the wrinkles regions at the neutral state

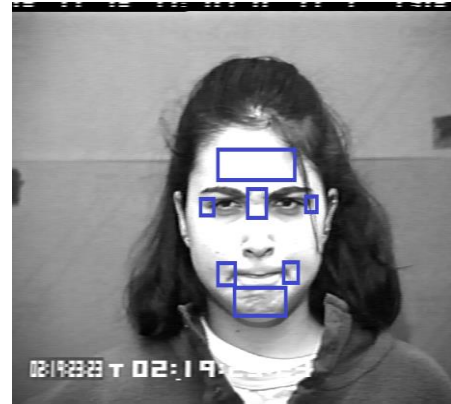


Figure 5. Location of the wrinkles regions during the emotion

IV. RESULTS

This section presents the different classification rates of our proposed emotion recognition system.

The Chon-Kanade database [9] contains a set of facial expressions images in grayscale of men and women of different ethnicities. The size of each image is 640 by 490 pixels. The orientation of the camera is front and the small movements of the head are present. This data set is very useful for facial expression recognition. Fig. 6 presents some examples of images of this data set.



Figure 6. Examples of images from Chon-Kanade dataset

TABLE I. PRESENTATION OF DIFFERENT PARAMETERS FOR EVALUATING OUR EMOTION RECOGNITION SYSTEM

	Chon-Kanade
Number of facial expressions	7
Total number of images	1070
Number of learning images	714
Number of test images	356

The classification rates of the wrinkles approach are presented in Table 1. It shows that the FWT has correctly classified the neutral emotion with a rate equal to 100%. The wrinkles regions of this emotion are not modified so the vector of difference of pixels is null. So, there are no doubts concerning this emotion.

The system classifies the emotion of joy with a classification rate equal to 72.73% and classifies the disgust with a classification rate equal to 63.44%. Finally, it classifies the fear emotion with a classification rate equal to 36.36%. It ranks sadness, anger and surprise with a classification rate equal to 54.55%. The reason for the low rates is due to the difficulty of detecting the wrinkles regions as shown in Fig. 7. Our dataset contains image of persons who did not express these emotions with the same manner. Let us take the emotion of sadness as shown in Fig. 8. There are persons who express the same emotion when their eyebrows are curved and their mouth are tight. However, there are other persons who express this emotion with released eyebrows and tight mouths. That is why we propose a hybrid system which will better improve these rates.

The hybrid system presents an enhancement of the classification rates. It classifies the emotion of joy with a classification rate equal to 90%. However, the system of the approach of the wrinkles classifies it with a classification rate equal to 72.73%. It classifies the anger with a classification rate equal to 100% but the wrinkles approach with a classification rate equal to 63.44%. Finally, it classifies the fear emotion with a classification rate equal to 80%. However, the first approach classifies it with a classification rate equal to 36.36%. We have become aware that our system made an improvement in the emotions of joy, anger, fear and neutral. We have also noticed that it is more robust than the first approach.

Figure 7. False wrinkles regions detection



Figure 8. Different expressions in the emotion of sadness

TABLE II. CLASSIFICATION OF THE CHON-KANADE DATA BASIS WITH FWT BY THE WRINKLES APPROACH AND THE HYBRID SYSTEM

	Wrinkles Approach	Hybrid System
joy	72.73 %	90 %
anger	54.55 %	100 %
disgust	63.44 %	63.44 %
sadness	54.55 %	54.55 %
fear	36.36 %	80 %
surprise	54.55 %	54.55 %
neutral	100 %	100

TABLE III. CLASSIFICATION RATES WITH THE APPROACH DESCRIBED IN [6]

joy	37.5%
anger	62 %
disgust	62.5 %
sadness	75 %
fear	50 %
surprise	25 %

Table 3 presents the classification rates of the system described in [6]. It has classified the class joy with a rate equal to 37.5% and the class anger with a rate equal to 62%. Besides, it has classified the class disgust with a rate equal to 62.5%, the class sadness with a rate equal to 75%, the class fear with a rate equal to 50% and the class surprise with a rate equal to 25%. However, our system has classified the first class with a rate equal to 90%, the second class with a rate equal to 100%, the third one with a rate equal to 63.44%, the fourth with a rate equal to 54.55% and the fear with a rate equal to 80%.



V. CONCLUSION

The two approaches used in this work to recognize emotions are based on the wavelets networks. An algorithm of training of these networks based on the 1D Fast Wavelet Transform has been proposed and has been implemented.

Our method of emotion recognition is the combination of two approaches. The first one contains four main stages: the detection of face's elements, the location of wrinkles regions in the face, the information extraction finally the classification. The second one contains also four main stages: detection of the elements of the face, localization of the characteristic points, tracking of features points and classification. Experiments on the dataset (Chon-Kanade) are made to evaluate the efficiency of our proposed approach. The performances of the Fast wavelets networks used for emotion recognition are clear and the results obtained are encouraging. The robustness and the rapidity of the proposed training algorithm that are based on 1D Fast Wavelet Transform theory increases these performances.

Our contributions are at the level of the second stage by locating automatically the wrinkles regions on the face and at the level of classification that we used the Fast Wavelet Transform as a classifier. We are looking actually to extend this work by recognizing the secondary emotions.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

REFERENCES

[1] C. Ben Amar, M. Zaied, and A. M. Alimi, "Beta wavelets. Synthesis and application to lossy image compression," *Advances in Engineering SoftWare*, Elsevier, special issue, *Advanced Algorithms and Architectures for Signal Processing*, vol. 36, no. 7, pp. 459-474, 2005.

[2] R. Afdhal, A. Bahar, R. Ejbali, and M. Zaied, "Face detection using beta wavelet filter and cascade classifier entrained with Adaboost". 8th International Conference on Machine Vision (ICMV'8) November 19-21, 2015 | Barcelona, Spain.

[3] R. Ejbali, Y. Benayed, M. Zaied, and A. M. Alimi, "Wavelet networks for phonemes Recognition", *International conference on systems and information processing* 2009.

[4] A. El Adel, M. Zaied, and C. Ben Amar, "Learning wavelet networks based on Multiresolution analysis: Application to images copy detection Communications", *Computing and Control Applications (CCCA)*, 2011.

[5] M. Zaied, R. Mohamed, and C. Ben Amar, "A power tool for Content-based image retrieval using multiresolution wavelet network modeling and Dynamic histograms". *International Review on Computers and Software (IRECOS)* 2012.

[6] K. Ghanem and A. Caplier, "Towards a full emotional system, Behaviour & Information Technology" 2013, 32:8, 783-799.

[7] R. Afdhal, R. Ejbali, M. Zaied, and C. Ben Amar. December 14-16. Emotion recognition using features distances classified by wavelets network and trained by fast wavelets transform., 14th International Conference on Hybrid Intelligent Systems (HIS'14) proceedings, , Gulf University for Science and Technologie , Kuwait., pp.238,241.

[8] M. Frank, P. Perona, and Y. Yacoob, "Automatic extraction of facial action codes" ,2001.

[9] R. Lanjewar, S. Mathurkar, and N. Patel, 2015, "Impelementation and comparaision of speech emotion recognition system using Guassian Mixture Model (GMM) and K-nearest Neighbor (K-NN) techniques".

[10] D. Glowinski and A. Camurri, 2008. "Technique for automatic emotion recognition by body gesture analysis".

[11] G. Castellano, S. D Villalba and A. Camurri, (2007). "Recognising Human Emotions from Body Movement and Gesture Dynamics". In A. Paiva, R. Prada & R.W. Picard (Eds.), *Lecture Notes in Computer Science*: vol. 4738. *Affective Computing and Intelligent Interaction*, Second International Conference, ACII 2007, Lisbon.

[12] A. Camurri, I. Lagerlöf and G. Volpe,. (2003). "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques". *International Journal of Human-Computer Studies*, 59(1-2), 213-225.

[13] M. Zaied, C. Ben Amar, and M. A. Alimi. 14(1-2)(2005), "Beta wavelet networks for face recognition", *J. Decision Syst.*, Lavoisier 2005 Edition, 109-122.

[14] R. Lanjewar, S. Mathurkar, and N. Patel, "Impelementation and comparaision of speech emotion recognition system using Guassian Mixture Model (GMM) and K-nearest Neighbor (K-NN) techniques", 2015.

[15] S. Mariooryad and C. Busso, 2015, "Facial Expression Recognition in the Presence of Speech using Blind Lexical Compensation".

[16] X. Xu, Y. Li, X. Xu, Z. Wen, H. Che , S. Liu , and J. Tao, 2014. "Survey on discriminative feature selection for speech emotion recognition".

[17] C. Wu, W. Wei, J. Lin, and W. Lee. "Speaking Effect Removal on Emotion Recognition From Facial Expressions Based on Eigenface Conversion", 2013.

[18] A. Halder1, R. Mandal1, A. Konar1, A. Chakraborty, and R. Janarthanam, 2011. "Emotion Recognition from Facial Expression using General Type-2 Fuzzy Set".

[19] J. Hakura, R. Domon, and H. Fujita, 2013. "Emotion Recognition Method Using Facial Expressions and Situation".

[20] C. Chang, J. Tsai, C. Wang, and P. Chung, 2009. "Emotion Recognition with Consideration of Facial Expression and Physiological Signals".

[21] S. Hassairi, R. Ejbali, and M. Zaied, "Sparse Wavelet Auto-Encoders for Image Classification", *International Conference on Conférence Digital Image Computing: Techniques and Applications (DICTA)*, 2016.

[22] M. Zaied, C. Ben Amar, and M. A Alimi. "Award a new wavelet based beta function", *International conference on signal, system and design*, SSD031, 185-191, 2003.

[23] R. Ejbali, Y. Benayed, M. Zaied, and A. M. Alimi, "Wavelet networks for phonemes recognition", *International conference on systems and information processing*, 2009.

[24] B. Guedri, M. Zaied, and C. Ben Amar. "Indexing and images retrieval by content High Performance Computing and Simulation", (HPCS), 2011.

[25] A. El Adel, M. Zaied, and C. Ben Amar, "Learning wavelet networks based on Multiresolution analysis: Application to images copy detection Communications", *Computing and Control Applications (CCCA)*, 2011.

[26] W. Bellil, C. Ben Amar, M. Zaied, and M. A. Alimi, " La fonction Beta et ses dérivées: vers une nouvelle famille d'ondelettes", *First International Conference on Signal, System and Design, SCS'04*, 201-207, 2004.

[27] O. Jemai, R. Ejbali, M. Zaied, and C. Ben Amar, "A speech recognition system based on hybrid wavelet network including a fuzzy decision support system", *Seventh International Conference on Machine Vision*, 2014.

[28] M. Zaied, R. Mohamed, and C. Ben Amar, "A power tool for Content-based image retrieval using multiresolution wavelet network modeling and Dynamic histograms", *International Review on Computers and Software (IRECOS)*, 2012.

[29] R. Ejbali, M. Zaied, and C. Ben Amar, "Multi-input Multi-output Beta Wavelet Network: Modeling of Acoustic Units for Speech Recognition", *International Journal of Advanced Computer Science and Applications*, 2012.

- [30] R. Ejbali, M. Zaied, and C. Ben Amar, "Intelligent approach to train wavelet networks for Recognition System of Arabic Words", International Conference on Knowledge Discovery and Information Retrieval, 2010.