Emotion Recognition by Modeling the Movement

Halima MHAMDI, Nessrine HAMROUNI, Mohamed Salim BOUHLEL
UR.SETIT
ENIS, Sfax University
Sfax, Tunisia
mhemdi.halima@gmail.com, neswebhamrouni@yahoo.fr, mbouhlel@enis.rnu.tn

ABSTRACT: *We are trying, through this paper, to expose emotion recognition approaches. In fact, the facial expressions analysis in computer vision is complicated by the variety of faces from one person to another, as well as problems related to the recording, such as lighting, the position of the face relative to the camera, and occlusions.*

In the context of the dynamic approaches, we can distinguish two different directions of research: the holistic approaches, where we analyze the face in its entirety, and the local approaches, where we concentrate on facial characteristics representing facial expressions. The relevant characteristic points are in particular eyes, eyebrows, mouth, and possibly the front. In general, the holistic methods are better to determine the expressions in their set, whereas the local methods can detect more subtle changes.

Among the various dynamic methods, we find dense Optical flow, applied holistically.

This method has also been applied locally by estimating the activity of several facial muscles. The calculation is performed in each window around the selected region of interest.

Keywords: Face Detection, Emotion Recognition, Facial Expressions, Optical Flow

Received: 1 March 2013, Revised 3 April 2013, Accepted 9 April 2013

1. Introduction

The Human Computer Interaction (HCI) is a discipline that is rapidly evolving. Environnement for future generations become multimodal human-machine incorporating new information from the consideration of the dynamic behavior, speech and / or facial expressions, so as to make the use of machinery and the most intuitive natural as possible.

The face is the most expressive and communicative human being, it is a major focus in current research on improving the UI for the establishment of a dialogue between the two entities.

Facial expression is a visible manifestation of a face of the mind (emotion, reflection), the cognitive activity of the physiological

activity (fatigue, pain), personality and psychopathology of person. Research in psychology has shown that facial expressions play an important role in the coordination of human conversation, and have a greater impact on the listener than the textual content of the message expressed. Therefore, the facial expression can be considered as an essential means of human communication.

Most of the facial expression information is contained in the main deformation permanent features of the face, characterized by a change visually perceptible. This distortion is due to intentional or unintentional activation of one or more of the 44 component face muscles. It continuously emits signs that the decoding, not only information about the emotional state of the person, but also sheds light on what is said

In this section, we will focus, first, on emotional communication channels. Second, we will define the difference between facial expression and emotion and finaly we will expose our approach to recognize emotion.

2. Emotional Communication Channels

2.1 Facial expressions

Facial expressions are a vector for transmission of emotion important, especially for conversational agents. There are three approaches to model and animate expressions. Parametric models [1] which directly manipulate the shape and geometry of the face by modifying an existing three-dimensional mesh.

The models based on recordings [2] rearrange pre-recorded images by applying morphing algorithms to obtain the desired expression.

Finally, physical models [3] faithfully reproduce human anatomy, often in several layers (bone structure, muscle tissue, skin) and simulate the contraction of muscles to produce speech.

The latter approach is more complex but also more flexible. It allows interpolating several facial expressions in order to obtain new or nuanced expressions [4].

Lip sync for animated conversational agents is done using procedural methods to synthesize movements directly from speech [5]. Finally, these movements are combined with emotional expression to produce the final animation. [6]

2.2 Synthesis of emotion in voice

The synthesis of emotions in speech is a fundamental issue for emotional agents. There are three main approaches [7] to generate emotions in speech; they are all based on traditional voice synthesis techniques.

First synthesis by forming [8] synthesizes a sound from acoustic parameters that are modified, to represent a given emotion. This approach provides flexibility but the voice is generated sounds robotic guards are unnatural.

Then the unit selection synthesis [9] uses a large corpus of speech (several hours) segmented into units of different sizes, which are assembled during the synthesis to produce the desired expression. The emotion that is transmitted is intrinsically linked to that present in the base, but the voice quality that is produced is practically equal to that of a human.

Finally concatenative synthesis [10] also uses a database, but it contains only the diphones for a certain language. To produce the desired emotion, a system of rules [11] modify the synthesis. This approach shows in practice as flexible as formant synthesis [12], but offers voice quality is much better.

It is also possible to obtain nuanced emotions, without degrading the quality of the voice, by spectral interpolation between two diphones recorded with a different emotion. [13]

2.3 Gesture

An important part of human communication is transmitted by the gestures. The system BEAT (Behavior Expression Animation Toolkit) [14] produces a non-verbal communication from a sentence and rules based on linguistic analysis and contextual obtained by the study of human behavior during conversations.

Gesture Engine system [15] is particularly interested in the movements of the arms and hands. From key positions ("*key-frames*"), it automatically generates from the transcript of a conversation, gestures suited for the upper body. By combining non-verbal communication with changes in posture and generic animations ("*idle motion*") [16], it is possible to obtain an IVH (Interactive Virtual Human) particularly realistic.

3. Face Expressions and Emotions

3.1 Emotion

Expressions and emotions are closely related and sometimes confused; emotion is one of the generators of facial expressions. The emotion is expressed through many channels such as body position, voice and facial expressions.

An emotion usually involves a corresponding facial expression (whose intensity can be more or less controlled by individuals), but the reverse is not true: it is possible to mimic an expression representing an emotion without feeling that emotion. While expressions depend on individuals and cultures, there are usually a limited number of universally recognized emotions.

3.2 Facial expression

Facial expression is a facial expression full of meaning. Meaning may be the expression of an emotion, a semantic index or intonation in sign language.

The interpretation of a set of muscle movement's expression is dependent on the application context. In the case of an application in human-machine interaction where one wishes to know an indication of the emotional state of an individual, we seek to classify the measures in terms of emotions.

4. Emotion Recognition

Automatic analysis of facial expressions is done in two steps: extraction of facial structures (optical flow [17], deformable 3D motion model [18], and tracking point's structure [19]).

Then the classification of facial structures (Hidden Markov Model, recurrent neural network, Templates energy movement, Neural network feed-forward, Nearest neighbor with distance measurement ...) are implemented.

To detect emotions, we propose an approach whose main steps are shown in Figure 1. These steps are divided into two main phases:

✓ **Models construction:** it allows quantifying the motion from optical flow vectors to estimate the directional model and the magnitude for the entire sequence.

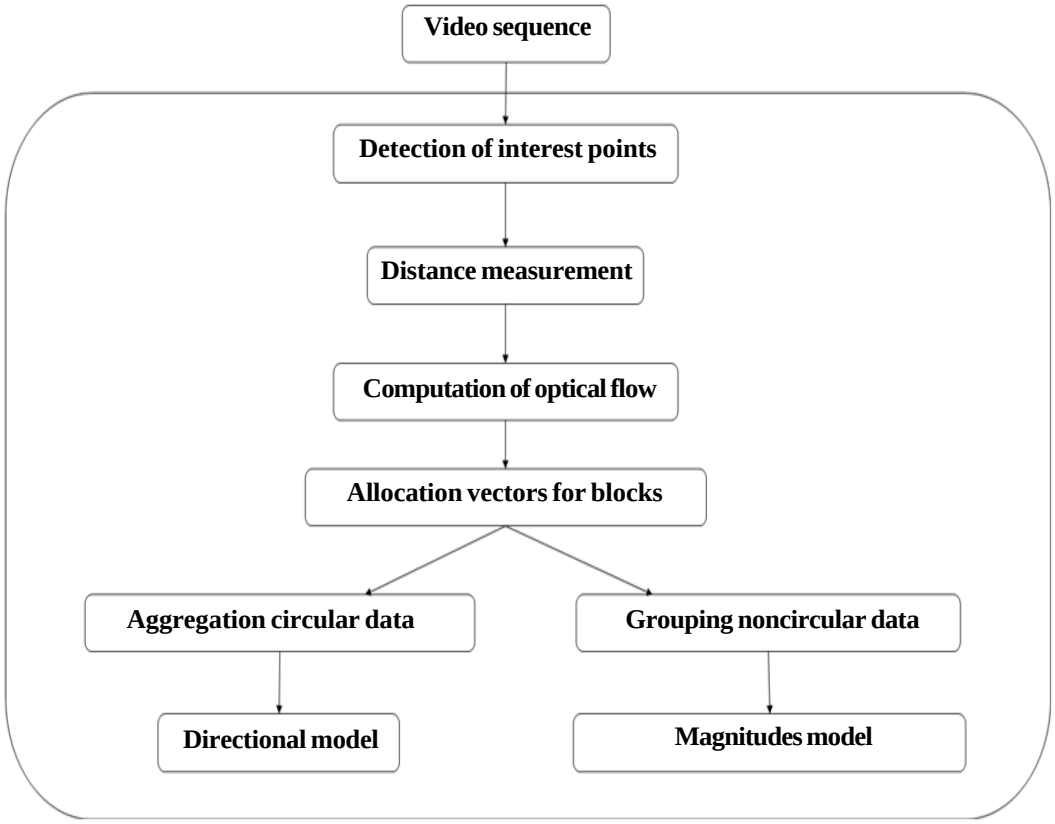
✓ **Emotion recognition:** it allows recognizing the action in a video comparing the model with models of video reference through a distance measurement.

4.1 Models construction

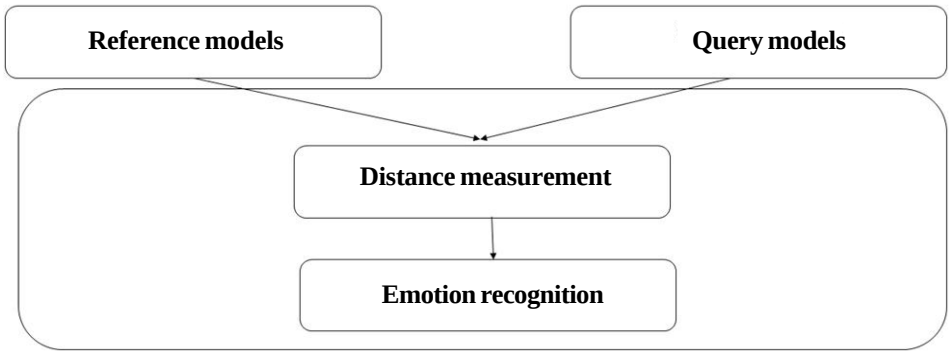
To construct the model of a video sequence, we start by extracting a set of points of interest in each image. We used the sensor points of interest Shi and Tomasi [20] for finding the corners with a high natural value. We also believe that, in the videos discussed, the position of the camera and the lighting conditions allow to obtain a large number of points of interest can be easily detected and tracked.

Having defined the set of points of interest, we follow their movements on the following images through optical flow vectors (figure 2). For this, we use the implementation of Bouguet [21] KLT tracking algorithm [22], which turns quickly and efficiently to manage the points lying near the edge of the image. The result is a set of motion vectors, where a vector is defined by an origin, direction and magnitude.

The next step is to divide the scene into a grid of $M \times N$ blocks. Then each motion vector is associated with the block that corresponds to it, depending on its origin.

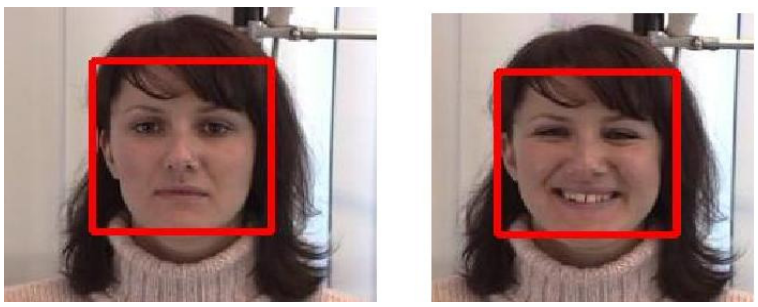


(a)

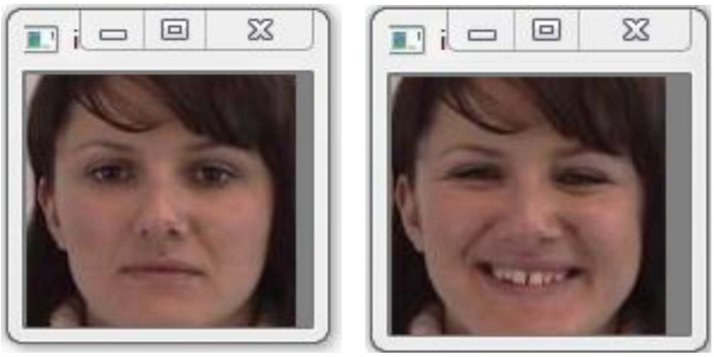


(b)

Figure 1. Approach's stages. (a) Construction models phase. (b) Emotion recognition phase



(a)



(b)

(c)

Figure 2. Face dection. (a) Source image, (b) Face detected, (c) Optical flow vectors

4.2 Emotion Recognition

Once the model is calculated from the video sequence, we detect the emotion corresponding to the query sequence based on the reference video. Emotions are detected by comparing the model of a query sequence patterns associated with reference sequences using a distance measurement. The model associated with the smallest distance from the emotion of a model query sequence is retained.

5. Conclusion

In this paper we presented the concept of emotion recognition. We have proposed an approach based on movement direction models and models magnitude.

We first extract the optical flow vectors to acquire statistical models on the direction and magnitude of movement.

Our future work will be directed towards a model that estimates the video main directions and magnitudes in all blocks of the scene. We will use a distance measurement to detect emotion by comparing the model of a query sequence in the reference models.

References

[1] Parke, F. (1974). A Parametric model for Human Faces. PhD thesis, University of Utah, Salt Lake City, UT.

[2] Ezzat, T., Geiger, G., Poggio, T. (2002). Trainable Videorealistic Speech Animation. *In*: Proc. SIGGRAPH'02, p. 388–398.

[3] Kahler, K., Haber, J., Seidel, H. -P. (2001). Geometry-based Muscle Modeling for Facial Animation, *In*: Proc. Graphics Interface, p. 37–46, June.

[4] Albrecht, I., Schröder, M., Haber, J., Seidel, H. -P. (2005). Mixed feelings: Expression of non-basic emotions in a muscle-based talking head, *Virtual Reality*, 8 (4) 201-212.

[5] Cohen, M., Massaro, D. (1993). Modeling Coarticulation in Synthetic Visual Speech. *In*: N. Magnenat-Thalmann and D. Thalmann,editors. *Models and Techniques in Computer Animation*, p. 139–156.

[6] Bui, T., Heylen, D., Nijholt, A. (2004). Combination of facial movements on a 3D talking head, *In*: Proceedings Computer Graphics International 2004 (CGI 2004) Crete, Greece June, *IEEE Computer Society*, p. 284-291

[7] Schroder, M. (2001). Emotional speech synthesis - a review, *In*: Proc.Eurospeech 2001, Aalborg, p. 561–564.

[8] Cahn, J. E. (1990). The Generation of Affect in Synthesized Speech, *Journal of the American Voice* I/O, 8, 11-18.

[9] Black, A. (2003). Unit Selection and Emotional Speech, *In*: Proceedings EUROSPEECH 2003.

[10] Rank, E., Pirker, H. (1998). Generating emotional speech with a concatenative synthesiser, *In*: Proceedings of ICSLP'98, 3, p. 671-674.

[11] Murray, I. R., Edgington, M. D., Campion, D., Lynn, J. (2000). Rule-based Emotion Synthesis Using Concatenated Speech, *In*: Proceedings of the ISCA Workshop on Speech and Emotion, p. 173-177.

[12] Oudeyer, P. Y. (2002). The production and recognition of emotions in speech: features and algorithms, *International Journal in Human-Computer Studies*, 59/1-2, p. 157-183, special issue on Affective Computing.

[13] Turk, O., Schröder, M., Bozkurt, B., Arslan, L. (2005). Voice Quality Interpolation for Emotional Text-to-Speech Synthesis, *In*: Proceedings Interspeech 2005, p. 797-800.

[14] Cassell, J., Vilhjalmsson, H., Bickmore, T. (2001). BEAT: the Behavior Expression Animation Toolkit, SIGGRAPH 2001, *In*: Proceedings of the Computer Graphics.

[15] Hartmann, B., Mancini, M., Pelachaud, C. (2002). Formational parameters and adaptive prototype instantiation for mpeg-4 compliant gesture synthesis. *In*: Computer Animation 2002, p. 111–119.

[16] Egges, A., Magnenat-Thalmann, N. (2005). Emotional communicative body animation for multiple characters, First International Workshop on Crowd Simulation (V-Crowds).

[17] Cohn, J. F., Zlochower, A. J., Lien, J. J., Kanade, T. (1998). Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression, *In*: Proc. Int'l Conf. Automatic Face and Gesture Recognition, p. 396-401.

[18] DeCarlo, D., Metaxas, D., Stone, M. (1998). An Anthropometric Face Model Using Variational Techniques, *In*: Proc. SIGGRAPH, p. 67-74.

[19] WANG, M., IWAI, Y., YACHIDA, M. (1998). Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model, *In*: Proc. Int'l Conf. Automatic Face and Gesture Recognition, p. 324-329.

[20] FASEL, B., Shi, J., et, Tomasi, C. (1994). Good features to track. *In*: Internatioal Conference on Computer,Vision and Pattern Recognition (CVPR), p. 593–600.

[21] Bouguet, J.-Y. (2000). Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. *Intel Corporation Microprocessor Research Labs*.

[22] Lucas, B., et, Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *In*: International Joint Conference on Artificial Intelligence (IJCAI), p. 674–679.