

FACIAL EMOTION RECOGNITION USING KINECT

*Andreea Bejgu¹
Irina Mocanu²*

ABSTRACT

This paper aims to investigate different methods for recognition of human emotion based on facial expression. The facial emotions may be used to perform association with various elements from the environment to reflect the user's status. User interactions are achieved using a Kinect sensor. RGB images provided by Kinect are processed and significant face features are computed and represented as Action Units (AUs). These AUs are used for emotion recognition using different machine learning algorithms: Multi Layer Perceptron, RBF Networks (normalized Gaussian radial basis function network), J48 Tree (Class for generating a pruned or unpruned C4.5 decision tree), NNge Rule (Nearest-neighbor-like algorithm using non-nested generalized exemplars), Simple Logistic (linear logistic regression models), SMO (John Platt's sequential minimal optimization algorithm for training a support vector classifier) and Naïve Bayes. The application is tested using 5 users for 7 emotions: neutral, angry, sadness, joy, disgust, fear and surprise. We obtained an average accuracy of 62.5%. The best accuracy was obtained in case NNge of and J48 methods. Also these methods have a good initialization time (approximately 2s).

1. INTRODUCTION

People are able to communicate and interact both by gestures, speaking and facial expressions. It is shown that 43 face muscles can be combined in 10.000 facial expressions. These can be used to express intentions, emotions and desires about surroundings objects and people.

Automatic recognition of emotions implies analysis of all kinds of non verbal communication, from audio and video. Much of the work conducted on human non verbal behavior analysis has relied on individual modalities. Examples include analysis of facial expressions, body gestures and speech intonation and intensity. As latest psychological research studies suggest that different qualities of body movement are also indicators of some underlying affective phenomena, systems that recognize emotions from gestures combined with expressive information from other modalities are now under development

There are many applications from ambient intelligence that use emotion recognition. One such application can use emotions to adapt the environment to the users' preferences. For example, users from an intelligent environment can make different activities –some of these activities may deviate from his schedule. These may potentially affect the health of the supervised person. Some of them may be detected as being the result of negative psychic

¹ Master Student, University POLITEHNICA of Bucharest, Splaiul Independentei No. 313, Bucharest 060042, Romania; anda_9_1@yahoo.com

² Associate Professor, University POLITEHNICA of Bucharest, Splaiul Independentei No. 313, Bucharest 060042, Romania, irina.mocanu@cs.pub.ro

states. These states have the potential of being attenuated by changing some parameters of person's environment. For example, the changes may consist of images with predominant colour, projected onto a wall of the room. Ambient music and lighting are other ambient factors that can influence the psychological state of the supervised person. Another application refers to a vision-based system to supervise elderly people, tracking their physical daily activities using his / her smartphone. The system will extract a mobility pattern for each user based on its physical activity, generating reminders and alerts in case of low physical activity. Thus the system will advise the user to increase her/his level of physical exercise. The exercises will be performed interactively, as an adaptive game, specifically designed for elderly persons. The intensity level of the recommended physical exercises (hard, medium of low physical level) will be generated and adapted based on additional information: medical requirements, values of the current vital parameters of the user (for example heart beat measured with a BioHarness belt) and also based on his face emotions.

This paper analyses the main methods used for emotions recognition based on facial expressions extracted from images. The set of images are provided by a Kinect sensor.

The rest of the paper is organized as follows. Section 2 presents different some existing methods used for emotion recognition. The testing application is described in Section 3. Section 4 presents the current evaluation of the analyzed methods. Conclusions and future work are listed in Section 5.

2. EXISTING METHODS FOR EMOTIONS RECOGNITION

The paper analyses different methods for emotion recognition. These methods requires two steps: a) processing step – extract features from images or videos; b) recognition step – extracted features are used with different machine learning algorithms, such as: neural network, genetic algorithms or Bayesian networks in order to recognize the main emotions: sadness, anger, joy, fear, surprise and disgust. The main methods represent features as action units in order to a easier recognition and possibly linking with other elements extracted in the processing step (body or hand gestures, context, etc).

In [1] emotions like: despair, interest, pleasure, sadness, irritation, joy, pride, are recognized based on both speech and gestures. Using both features, the accuracy of recognition is increased by 10%. The recognition was made with Hidden Markov Models, Bayesian networks and Dynamic Bayesian networks that are very well suited for fusing different sources of information in multimodal emotion recognition and can also handle noisy features and missing values of features by probabilistic inference.

Another method [2] combines face expressions with hand gestures. The recognized emotions are anger, disgust / contempt, fear, happiness, sadness and surprise.

Another method is developed by a group from *School of Mechatronics Engineering, University Malaysia Perlis* [3], They describe emotions using face description proposed by Ekman and Friesen [4] using 27 action units (AU) (for example lifting the chin, lowering elbow, etc). The identified AU are processed with neural networks and genetic algorithms. The method's accuracy is 87%.

Paper [5] performs emotion recognition from images and videos. A set of action units are associated to faces. The system extracts pyramid of histogram of gradients (PHOG) and local phase quantization (LPQ) features for encoding the shape and appearance information. For selecting the key frames, K-means clustering is applied to the normalized shape vectors derived from constraint local model (CLM) based face tracking on the image sequences. Shape vectors closest to the cluster centers are then used to extract the shape and appearance features. Table 1 and Table 2 describe the main characteristics of the analyzed methods.

Table 1: extracted features and recognition algorithms

Method	Extracted Features	Used algorithms	Recognized emotions
Recognition based on speech, gestures and face expressions [1]	-Face expression (eyebrows, eyes, nose, mouth); -Body movements (5 movement curves); -Speech, voice inflections	-image processing for noise reduction -pattern recognition using: -neural networks; -Bayes networks; -Dynamic Bayes networks; -Bayesian classifier (Weka).	Anger Desperation Interest Pleasure Sadness Irritation Joy Pride
Face and gestures analysis [2]	-Facial features (eyebrows, mouth, eyelids, eyes) represented as AU; -Body (Shoulders, arm movements: cover face with hands, slow or large movements)	-body recognition using image segmentation -noise reduction -face detection based on skin colour using Bayesian classifier -gestures recognition using Markov Models combined with Multi-Layer Neural Network	Anger Sadness Joy Disgust Fear Surprise
Facial expression analysis [3]	-15 points are extracted from the face, obtaining 27 AU (action units) ; -2 interests regions (eyes and lips)	-2 interests regions: lips and eyes; -grayscale filter; -color histogram equalization for contrast enhancing; -median filters for noise reduction; -Sobel filter for boundary detection; -pattern recognition using genetic algorithms and neural networks.	Anger Sadness Joy Disgust Fear Surprise Neutral
Facial expression analysis [5]	-face codification using AU	-face detection using Constraint Local Models; -clustering algorithms;	Anger Fear Joy Relief

		-pyramid histogram of oriented gradients;	Sadness
--	--	---	---------

Table 2: accuracy of emotion recognition

Method	Recognized emotions	Accuracy	Images / video	Image type
Recognition based on speech, gestures and face expressions [1]	Anger Desperation Interest Pleasure Sadness Irritation Joy Pride	-Fata: 48.3%-59.6% -Gesturi: 67.1%-83.2% -Vorbire: 57.1%-70.8% Cu 10% mai buna analiza corelata decat analiza separata a trasaturilor. Corelat: 78.3%-89.4%	Video: 25 fps	two cameras: one for body and one for face
Face and gestures analysis [2]	Anger Sadness Joy Disgust Fear Surprise	91.9% - 98.7% for gesture recognition	-	Images from a database – FABO [6] (face expression and gestures)
Facial expression analysis [3]	Anger Sadness Joy Disgust Fear Surprise Neutral	-87%	-	-frontal images
Facial expression analysis [5]	Anger Fear Joy Relief Sadness	-44% person independent -73% person dependent -56% total	60x60 for the detected face	-frontal images -10 persons: - 7 persons for the training phase; - 6 persons for the test phase;

3. DESCRIPTION OF THE TESTING APPLICATION

The paper adapts method described in [3] in order to perform emotion recognition from images provided by a Kinect sensor. Face features are detected as in [7]. Eyebrows, eyes, nose and mouth are used to describe face features as action units. Figure 1 presents an example of the detected face features [7]. The recognized emotions are neutral, angry, sadness, joy, disgust, fear and surprise.

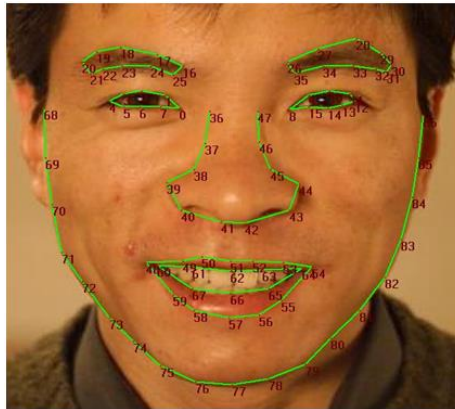


Figure 1. Extracted face features, as in [7]

The architecture of the application is provided in Figure 2. Collected images from a Kinect sensor are stored in the Image Database. Face features are extracted from images and codified by action units. The action units are stored in the AU Database. AU are used for emotion recognition. Multi Layer Perceptron, RBF Networks, J48 Tree, NNge Rule, Simple Logistic, SMO and Naïve Bayes. are used for emotion recognition.

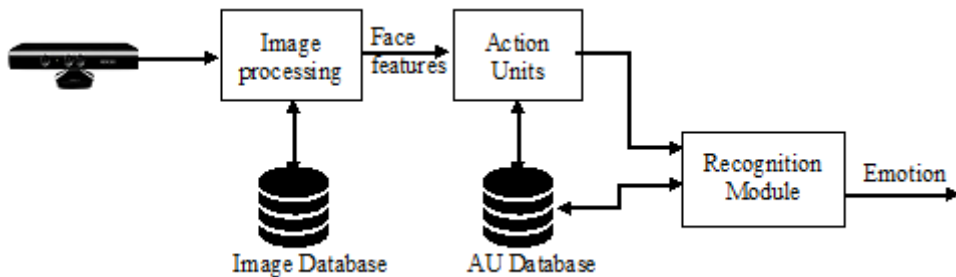


Figure 2. System architecture

A set of action units AU0-AU11, described in Figure 3 a)-i) are correlated with the corresponding values of the neutral state. The following AUs are computed based on the detected face features:

- AU0 - vertical distance of the eye
- AU1 - the inner brow angle with the horizontal
- AU2 - the outside of the brow angle with the horizontal
- AU3 - distance between the eyebrows
- AU4 - the distance between the left eye and left eyebrow
- AU5 - the distance between the right eye and right eyebrow
- AU6 - tight nose ("wrinkled")
- AU7 - enlarged nostrils / narrowed (nose width at nostrils)

- AU8 - vertical distance of the mouth (height)
- AU9 - horizontal distance of the mouth (width)
- AU10 and AU11 - mouth corners left / high (AU11 - the average distance of the points above the line that connects the corners of the mouth; AU10 - average distance of points below the line joining corners of the mouth)



Right eye

$$AU0 = (d1 + d2) / 2 - AU0_{neutral_state}$$

Figure 3. a) AU0



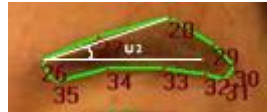
Left eye



Right eyebrow

$$AU1 = AU1_{neutral_state} - \text{Max}(u1, u2)$$

Figure 3. b) AU1



Left eyebrow



Right eyebrow

$$AU2 = AU2_{neutral_state} - \text{Max}(u1, u2)$$

Figure 3. c) AU2

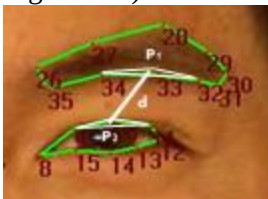


Left eyebrow



$$AU3 = d - AU3_{neutral_state}$$

Figure 3. d) AU3



Left side

$$AU4 = d - AU4_{neutral_state}$$

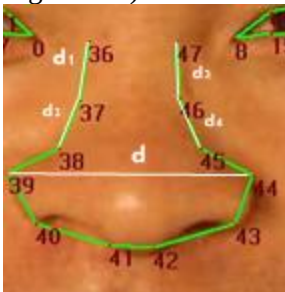
Figure 3. e) AU4



Right side

$$AU5 = d - AU5_{neutral_state}$$

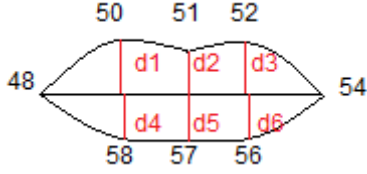
Figure 3. f) AU5



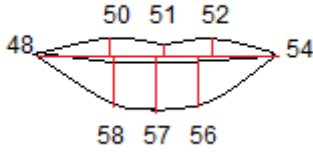
$$AU6 = (d1 + d2 + d3 + d4) / 4 - AU6_{neutral_state}$$

$$AU7 = d - AU7_{neutral_state}$$

Figure 3. g) AU6, AU7



neutral state

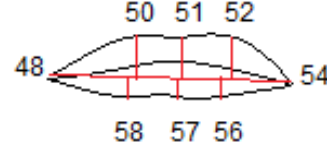


Raised mouth corners

$$AU10 = (d4 + d5 + d6) / 3 - AU10_{neutral_state}$$

$$AU11 = AU11_{neutral_state} - (d1 + d2 + d3) / 3$$

Figure 3. i) AU10, AU11



Left mouth corners

4. SYSTEM EVALUATION

Images are collected with a Kinect sensor [8] and they are processed with FaceTrackingVisualization [7] from Microsoft Kinect SDK [9]. Emotion recognition is performed using the Weka library [10] (Multi Layer Perceptron, RBF Network, J48 Tree, NNge Rule, Simple Logistic, SMO, Naïve Bayes). We use two methods for features computing:

- **M1:** AUs are computed as in Figure 3. a) – i).
- **M2:** AUs are computed as in M1; averages values for neutral states from training database are saved; in the testing phase, each person is calibrated regarding the known persons.

For training and testing phase we use the following datasets:

- **Set A** – sitting person at approximately 1m distance from Kinect; 800 instances;
- **Set B** – sitting person at approximately 1-1.5m distance from Kinect; 950 instances;
- **Set C** – sitting person at approximately 1m distance from Kinect; 650 instances;
- **Set D** – sitting person at approximately 1m distance from Kinect; 500 instances;
- **Set E** – sitting person at approximately 2m distance from Kinect; 580 instances;
- **Set F** – sitting person at approximately 1.5m distance from Kinect; 730 instances;
- **Set G** – sitting person at approximately 1m distance from Kinect; 400 instances;
- **Set H** – standing person at approximately 2m distance from Kinect; 900 instances;
- **Set I** – sitting person at approximately 1m distance from; 1000 instances;
- **Set J** – standing person at approximately 2m distance from Kinect; 900 instances;

- **Set K** – sitting person at approximately 1.5m distance from Kinect; the Kinect sensor is situated above the user.

The datasets were processed with both M1 and M2 methods in the following cases:

1. **Case1:** features are computed as in method M1; the training datasets are Set D to I and the testing datasets are Set A to F; the accuracy of the emotion recognition for the testing set is given in Figure 4.

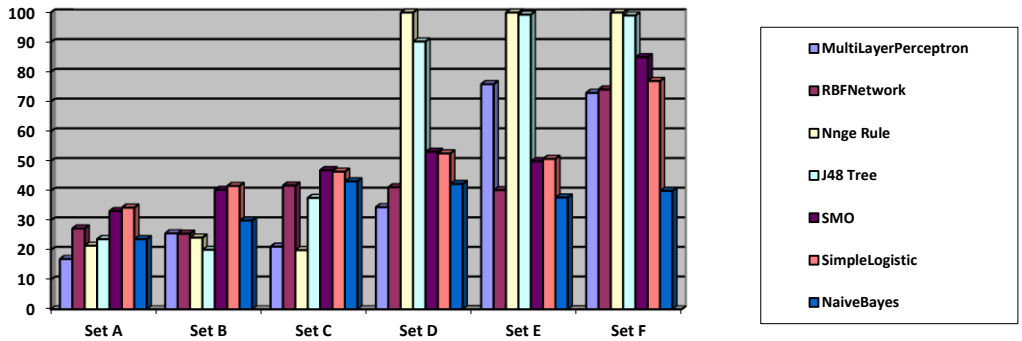


Figure 4. Accuracy for emotion recognition (**Case 1**)

2. **Case 2:** features are computed as in method M1; the training set is chosen of 70% images from Set A to I; the testing set is composed of 30% images from Set A to I; the accuracy of emotion recognition for the testing set is given in Figure 5.

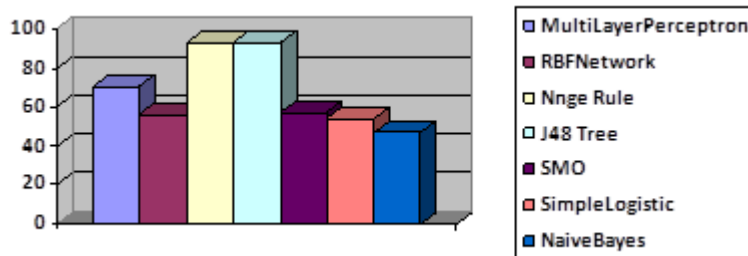


Figure 5. Accuracy for emotion recognition (**Case 2**)

3. **Case 3:** features are computed as in method M2; the training datasets are Set D to I and the testing datasets are Set A to F; the accuracy of the emotion recognition for the testing set is given in Figure 6.

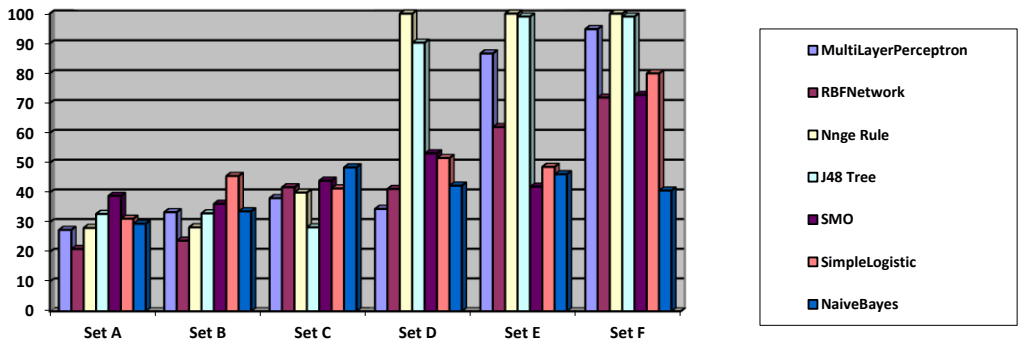


Figure 6. Accuracy for emotion recognition (*Case 3*)

4. **Case 4:** features are computed as in method M2; the training set is chosen of 70% images from Set A to I; the testing set is composed of 30% images from Set A to I; the accuracy of emotion recognition for the testing set is given in Figure 7.

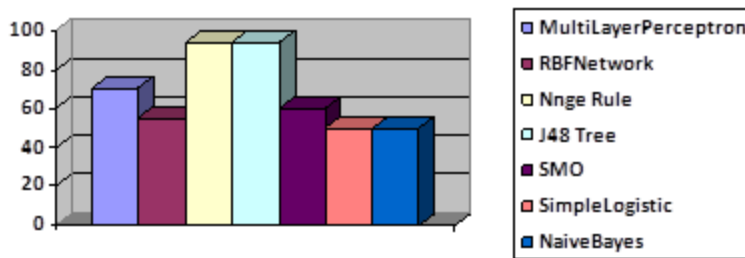


Figure 7. Accuracy for emotion recognition (*Case 4*)

Case 3 and 4 show an increasing accuracy than in Case 1 and 2. The execution time is 0.2s (for Naïve Bayes), 0.38s (for J48) and over 25s (for Multi Layer Perceptron). The execution times were measured for approximately 9000 analyzed instances. The average accuracy for Naïve Bayes, Simple Logistic and SMO is 50% and the initialization time is 1.3s. NNge and J48 have an average accuracy of 85% and an initialization time of 2s. The Multi Layer Perceptron was used with 3 hidden layers and 500 neurons on each layer. The initialization time varies from 2-3s to 15s and the accuracy of emotion recognition is 50%. Hence this classifier can be used for offline testing.

5. CONCLUSION AND FUTURE WORK

The paper presents a study for emotion recognition based on face features extracted from images collected from a Kinect sensor. On average the accuracy of emotion recognition is 62.5%. In future we'll use the NNge and J48 methods to recognize face emotions in an interactive game, adapted to elderly people in order to increase their physical activity.

REFERENCES

- [1] G. Caridakis, G. Castellano, L. Kessous, A. Raouzaoui, L. Malatesta, S. Asteriadis, K. Karpouzis, "Artificial Intelligence and Innovations 2007: From Theory to Applications", in IFIP International Federation for Information Processing, vol. 247, 2007, pp. 375-388
- [2] P. Metri, J. Ghorpade, A. Butalia (Department of Computer Engineering, MIT-COE, Pune), "Facial Emotion Recognition Using Context Based Multimodal Approach", 2012.
- [3] M. Karthigayan, M. Rizon, R. Nagarajan, S. Yaacob (*School of Mechatronics Engineering, University Malaysia Perlis (UNIMAP)*), "Genetic Algorithm and Neural Network for Face Emotion Recognition", <http://cdn.intechopen.com/pdfs-wm/5178.pdf>, accessed November 2014.
- [4] Ekman P, Friesen WV. Unmasking the face: a guide to recognizing emotions from facial clues. Englewood Cliffs, NJ: Prentice-Hall; 1975.
- [5] A. Dhall, A. Asthana, R. Goecke, T. Gedeon (School of Computer Science and School of Engineering, CECS, Australian National University, Australia and Vision & Sensing Group, Faculty of Information Sciences and Engineering, University of Canberra, Australia), "Emotion Recognition Using PHOG and LPQ features", IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), pp. 878 – 883, 2011
- [6] H. Gunes and M. Piccardi, "A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior", in Proc. of ICPR 2006 the 18th International Conference on Pattern Recognition, Vol. 1, pp. 1148-1153, Aug. 2006, Hong Kong
- [7] Microsoft SDK for FaceTracking: <http://msdn.microsoft.com/en-us/library/jj130970.aspx>,
accessed November 2014
- [8] Kinect Sensor Xbox 360: <http://www.xbox.com/en-US/xbox360/accessories/kinect/KinectForXbox360>, accessed November 2014
- [9] Microsoft SDK for FaceTracking: <http://msdn.microsoft.com/en-us/library/jj130970.aspx>
- [10] E. Frank, "RBFNetwork: Classes that implement radial basis function networks", WEKA documentation:
<http://weka.sourceforge.net/packageMetaData/RBFNetwork/index.html>, accessed November 2014