



HAL
open science

Facial Expressions Recognition: Development and Application to HMI

Zekhnine Chérifa, Berrached Nasr Eddine

► **To cite this version:**

Zekhnine Chérifa, Berrached Nasr Eddine. Facial Expressions Recognition: Development and Application to HMI. 6th IFIP International Conference on Computational Intelligence and Its Applications (CIIA), May 2018, Oran, Algeria. pp.573-584, 10.1007/978-3-319-89743-1_49 . hal-01913915

HAL Id: hal-01913915

<https://inria.hal.science/hal-01913915>

Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Facial expressions recognition: Development and application to HMI

Zekhnine Chérifa¹ and Berrached Nasr Eddine²

¹ University Of Science and Technology of Oran Mohamed Boudiaf BP. 1505 El M'naouer -31 000, Oran, Algeria

² University Of Science and Technology of Oran Mohamed Boudiaf BP. 1505 El M'naouer -31 000, Oran, Algeria

elnfi@yahoo.fr, laresi.usto.2015@gmail.com

Abstract. We present in this paper, a facial expressions recognition system to command a mobile robot (Pioneer-3DX). The proposed system mainly consists of two modules: facial expression recognition and robot command. The first module aims to recognize the facial expressions like happiness, sadness, surprise, anger, fear, disgust and neutral using Gradient Vector Flow (GVF) snake to find ROI (Region Of Interest like: mouth, eyes, eyebrow) segmentation from FEEDTUM database (video file). While the second module, analyses the segmented ROI to recognize with Euclidian distance calculation (compatible with the MPEG-4 description of the six universal emotions) and Time Delay Neural Network classifier. Finally, the recognized facial expressions were used as control commands for the mobile robot displacement (forward; backward; turn left; turn right) in ROS (Robot Operating System).

Keywords: Facial Expression Recognition, GVF Snake, MPEG-4 Description, Time Delay Neural Network, Mobile Robot, Robot Operating System.

1 Introduction

Automatic analysis of facial expressions constitutes an important tool for research in human machine interaction (HMI). Autonomous robot control systems are complex systems that consist of a sensor, a decision-making control system and a motor drive system.

The sensor can be either a visual system, a speech system, or a manual control system [1]. In [2], [3], numerous systems are proposed to control a robot or a wheelchair using head or face movement. Such systems involve body movement and are not suitable for people with extreme physical disabilities where head or face movement is difficult. Speech controlled systems [4] are also not suitable for people with speech disability. Thus, current research has been focused on design of systems, which can be a good solution to these problems. The best alternative is to design a system where command is derived from recognizing the user's facial expressions like happiness, sadness, surprise, anger and neutral.

Recently, many works were about the automatic localization of the face and its characteristic features. Indeed, the objective is to develop some interactive systems capable to analyze and to interpret the user's behavior. Several works have been made in this sense. The methods described in [5], [6] use models of facial movements, and those of [7] and [8] the methods of classification by neural networks.

We already developed a FER (Facial Expression Recognition) system in the static case (JAFFE database) with MLP neural network, trained by the backpropagation of gradient algorithm. We used data extracted by a local method to model the facial expressions. Thus introducing the geometric coordinates of 19 points characterizing the ROI manually as made in [9]. These extracted data are normalized, so that the facial expressions remain invariant for either the change of scale or slant of the individual head. The performances of our recognition system, have been established on 4 facial expressions "neutral, joy, surprise, fear", among the seven universal expressions, because there is no very big difference (in geometric side) between neutral and fear expressions, and between joy and surprise expressions [10].

We have developed another FER system in the dynamic case (FEDTUM database [11]) with Time Delay Neural Network classifier, trained by the backpropagation of gradient algorithm. In this system, we use GVF snake and the Euclidian distance calculation (compatible with the MPEG-4 description of the six universal emotions) manually and we interpret this analysis in articular movements of an arm manipulator robot "Mentor 5-dof" in the setting development of HMI application [12].

In this article, we present our FER system (the dynamic case) with some improvements in the modeling, the analysis and the interpretation of the facial expressions from image sequences, to command the mobile robot displacement (forward; backward; turn left; turn right) in ROS (Robot Operating System).

2 Facial Expression Modeling

Features extraction is the crucial and complex part in any shapes recognition system. It often uses the results of the statistics theory, of evaluation, to get a transformation from the representation space toward the space of interpretation. The major problem of this part can be seen as the resolution of a two sub-problems:

- What measures must be made?
- What features from row data to use as input?

In our case, input data are features of facial expressions modeling frames. The facial expression is measured by the temporary non-rigid distortion (0.25-5s) of the facial features (eyebrows, eyes, mouth). We have used the Gradient Vector Flow snake algorithm to extract the deformable and non-rigid measurement. However, this algorithm presents some limitations: it is only used for binary images, selects manually the initial snake and extracts only one geometric object per images. We propose to modify the GVF algorithm in a way to overcome those limitations. Our method is described in section 2.1.

The FEEDTUM database was chosen because it is of the most used databases in facial expression recognition. It consists of elicited spontaneous emotions of 18 subjects within the MPEG-4 emotion-set plus added neutrality.

Kass and al. [13] have introduced the active contours or snakes. Since, numerous variants of these deformable models have been studied for multiple applications. Their utility was particularly well illustrated in medical imagery, but also in electronic surveillance domain, and spatio-temporal tracking in video [14].

The GVF method has been developed to propose a solution to some limitations of the approach as the initialization of the snake and its convergence toward the concave regions [15].

2.1 GVF Field

The GVF method proceeds in two stages to calculate the GVF field:

- Calculate the gradient of image
- Calculate the gradient vector flow

The classical snake model proposed by Kass et al., defines the active contour as a parametric curve, $r(s) = (x(s); y(s))$, that moves in the spatial domain until the energy functional in Eq. 1 reaches its minimum value.

$$E_{snake} = \int (E_{int}(r(s)) + E_{ext}(r(s))) ds \quad (1)$$

E_{int} and E_{ext} represent the internal and external energy, respectively. The internal energy enforces smoothness along the contour. A common internal energy function is defined as follows:

$$E_{int}(r(s)) = (\alpha |r'(s)|^2 + \beta |r''(s)|^2) \quad (2)$$

Where α and β are weighting parameters, r' and r'' are the first and second derivative of $r(s)$ with respect to s . The first term, also known as tension energy, prevents the snake to remain attracted to isolated points. The second term, known as bending energy, prevents the contour of developing sharp angles. Constraints based on more complex shape models, such as Fourier descriptors.

The external energy is derived from the image, so that the snake will be attracted to features of interest. Given a gray level image $I(x;y)$, a common external energy is defined as:

$$E_{ext}(I(x,y)) = -|\nabla(G_\sigma(x,y) * I(x,y))|^2 \quad (3)$$

Where ∇ is the gradient operator, $G_\sigma(x,y)$ a 2D Gaussian kernel with standard deviation σ and where $*$ is the convolution operator. Minimizing the energy function of eq. 1 results in solving the following associated Euler-Lagrange equations:

$$\alpha \frac{d^2 x(s)}{ds^2} - \beta \frac{d^4 x(s)}{ds^4} - \frac{\partial E_{ext}}{\partial x} = 0 \quad (4a)$$

$$\alpha \frac{d^2 y(s)}{ds^2} - \beta \frac{d^4 y(s)}{ds^4} - \frac{\partial E_{ext}}{\partial y} = 0 \quad (4b)$$

This can be seen as a force balance equation:

$$F_{int} + F_{ext} = 0 \quad (5)$$

These equations can be solved using gradient descent by considering $r(s)$ as a function of time, i.e. $r(s; t)$. The partial derivative of r with respect to t is then

$$\frac{dx(s,t)}{dt} = \alpha \frac{d^2 x(s,t)}{ds^2} - \beta \frac{d^4 x(s,t)}{ds^4} - \frac{\partial E_{ext}}{\partial x} \quad (6a)$$

$$\frac{dy(s,t)}{dt} = \alpha \frac{d^2 y(s,t)}{ds^2} - \beta \frac{d^4 y(s,t)}{ds^4} - \frac{\partial E_{ext}}{\partial y} \quad (6b)$$

When the snake stabilizes, i.e. when an optimum is found, the terms $\frac{dx(s,t)}{dt}$ and $\frac{dy(s,t)}{dt}$ vanish.

2.2 Gradient Vector Flow

The external force field defined in the previous section requires a good initialization [15], close to the object boundary, in order to segment the object. This limitation is caused by the nature of the external force field, whose vectors point towards the object only in the proximity of the object's boundary. As we move away from the boundary the external fields rapidly become zero, therefore reducing the possibilities that a contour located in such regions will converge correctly. To overcome this problem, Xu and Prince [16] proposed another external force field $v(x;y) = (u(x;y); v(x;y))$. This vector field minimizes the following energy functional:

$$E_{GVF}(u, v) = \iint \mu \left(\frac{du^2}{dx} + \frac{du^2}{dy} + \frac{dv^2}{dx} + \frac{dv^2}{dy} \right) + |\nabla f|^2 |v - \nabla f|^2 dx dy \quad (7)$$

Where μ is a nonnegative parameter expressing the degree of smoothness of the field v and where f is an edge map, e.g. $|\nabla I|$. The first term in Eq.7 keeps the field v smooth, whereas the second term forces the field v to resemble the original edge force in the neighborhood of edges. This new external force is called gradient vector flow (GVF) field. The GVF-field can be found by solving the following associated Euler-Lagrange equations:

$$\mu \nabla^2 u - \left(u - \frac{\partial f}{\partial x} \right) \left(\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y} \right) = 0 \quad (8a)$$

$$\mu \nabla^2 v - \left(u - \frac{\partial f}{\partial x} \right) \left(\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y} \right) = 0 \quad (8b)$$

Where ∇^2 is the Laplacian operator.

2.3 Semi-Automatic Initialization of GVF Snake

We have automated the initialization procedure of the snake by a geometric shape (ellipsis) that corresponds in its shape to the region of interest (the eyes), in the goal to give solution for the number of initial snake points problem, (see Fig. 1).

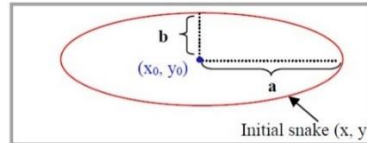


Fig. 1. Geometric shape of the initial snake.

$$x = x_0 + a \cos(t) \quad (9)$$

$$y = y_0 + b \sin(t) \quad (10)$$

Where (x_0, y_0) defines the region of interest center coordinates, (x, y) are the initial snake coordinates, a and b are respectively the width and the height of the ellipsis, t represents the step displacement of the snake.

We modify the initial snake so that is initialized automatically (see Fig. 2. a), with the GVF snake parameters: $\alpha=0.2$, $\beta=0$, $\gamma=1$, $\tau=2$, $iterMax=5$.

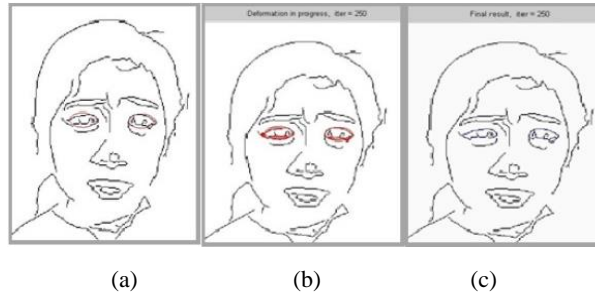


Fig. 2. a) Automatic snake initialization. b) Snake progression. c) Final snake.

The snake could converge toward the zone of concavity during 250 iterations, from a snake automatically initialized without any user intervention.

2.4 Superposition of the GVF Field and the GVF Snake

The second idea that proved very interesting to achieve, is to superpose the GVF snake and the GVF field as shown in (Fig. 3. a) in the goal to understand why the snake chose to converge toward this result.



Fig. 3. a) Superposition Image. b) Automatic snake initialization. c) Final snake.

We notice then, that the snake of (Fig. 3. c) converged well toward the ROI in the GVF sense because it perfectly integrates the zone of the potential biggest GVF.

We achieved two strategies of attributes extraction by GVF snake from video frames using FEEDTUM database. These strategies are spatio-temporal extraction and temporal extraction.

2.5 Spatio-Temporal Extraction Of The Attributes

The spatio-temporal extraction, consist in making a spatial extraction of eyes contour by GVF snake, for each frame regardless of the others (see Fig. 4). The result of temporal extraction is defined by the set of results of all frames sequence, that we created while gathering pictures of three expressions (neutral, disgust, joy) by the software Ulead Medium Studio Pro 7.0, in order to test the performances of the GVF snake.

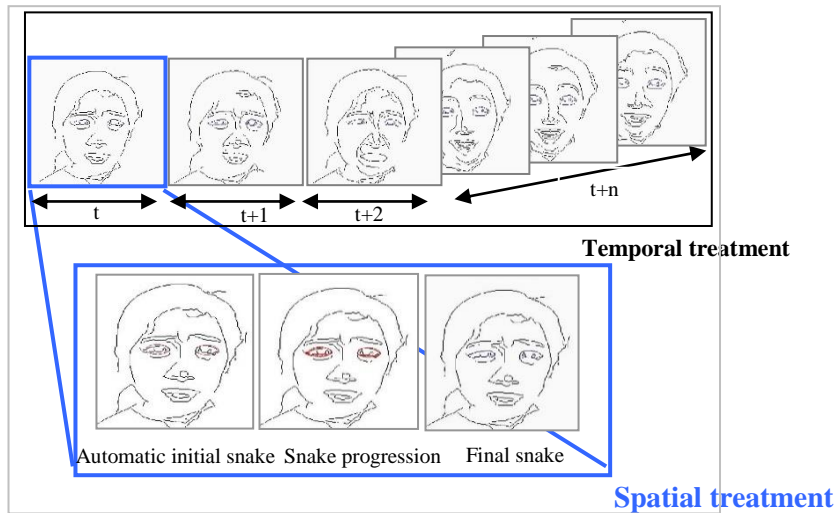


Fig. 4. Spatio-temporal extraction of ROI by GVF snake.

2.6 Temporal Extraction of the Attributes

This extraction, consist in initializing snake of the first frame, then the resulting contour of frame t is used as an initialization of the contour of frame $t + 1$ (see Fig. 5). The GVF force should range as far as the ROI can move between two subsequent frames as:

$$(Initial\ snake)|_{t=0} = Ellipsis\ shape.$$

$$(Initial\ snake)|_{t=t+1} = (final\ snake)|_{t=t}.$$

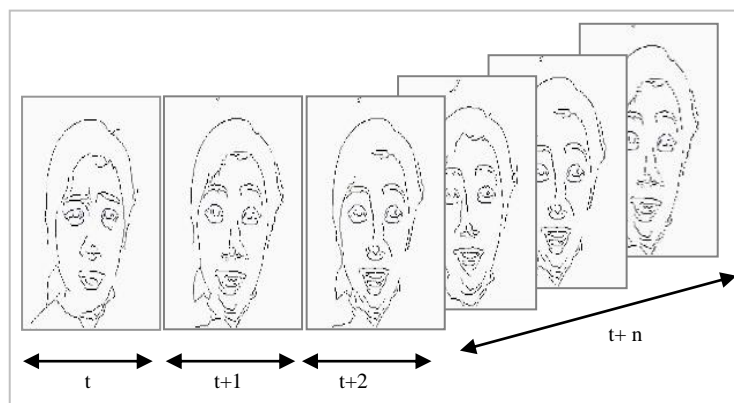


Fig. 5. Temporal extraction of ROI by GVF snake.

3 Application

We applied our GVF snake algorithm, using the spatio-temporal extraction strategy described above. As seen in (Fig. 6), we initialized three snakes to segment the regions of the eyes, the eyebrows, and the mouth.

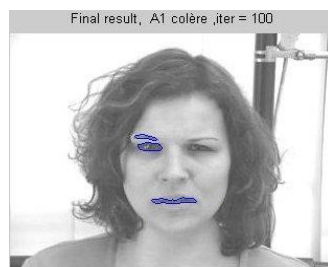


Fig. 6. Extraction of the ROI of FEEDTUM database by GVF snake.

These snakes progress at the same time and by the same parameters. Let us note that it was very difficult to find the adequate regulating parameters of the three snakes.

3.1 Modeling the ROI by MPEG-4 Description

We exploited the present information in the skeletons of the sequences of segmented pictures by the GVF snake (Fig. 7).

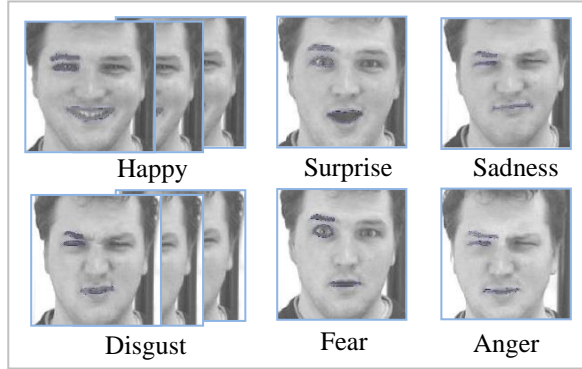


Fig. 7. Skeletons of the six facial expressions extracted by the GVF snake (FEEDTUM database).

We took the work in [17] as a basis, according to the description of the MPEG-4 norm of the six universal expressions (Expressions \rightarrow translation of these descriptions) (see Fig. 8. a). We extract automatically points of interest from segmented ROI to calculate the 5 Euclidian distances (see Fig. 8. b).

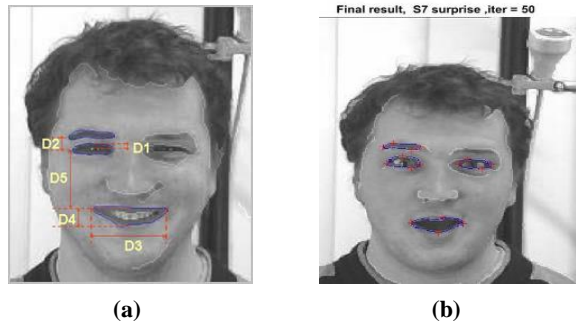


Fig. 8. a) The facial expression model by MPEG-4 description. b) Automatic extraction of points of interest from GVF snake results.

The neuter is represented by five distances takes as reference. For the rest of the expressions their description will be a combination of comparison of their distances in relation to those of the neutral expression. That, every facial expression (in our case, video sequence = 4 frames) is represented by four normalized features vectors.

We have applied this normalization procedure:

The change of axis (O, X, Y) linked to image, to (O', U_h, U_v) linked to face is established as follows:

$$x_{o'} = \frac{x_{og} + x_{od}}{2} \quad (11)$$

$$y_{o'} = \frac{y_{og} + y_{od}}{2} \quad (12)$$

Where $x_{o'}$ and $y_{o'}$ are the coordinates of new origin. $x_{od}, y_{od}, x_{og}, y_{og}$ are respectively coordinates of right and left eyes (see Fig. 9).

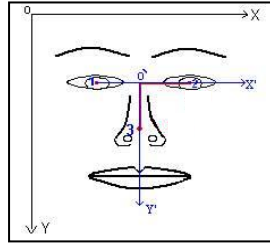


Fig. 9. Normalization with changing landmark.

$$Uh = \sqrt{[x_{od} - x_{og}]^2 + [y_{od} - y_{og}]^2} \quad (13)$$

$$Uv = \sqrt{[x_{nez} - x_{o'}]^2 + [y_{nez} - y_{o'}]^2} \quad (14)$$

After having made a change of landmark, we can calculate the coordinate of the points of interest by taking:

$$X = \frac{x - x_{o'}}{U_h/2} \quad (15)$$

$$Y = \frac{y - y_{o'}}{U_v} \quad (16)$$

With, x, y are coordinates of points of interest in old landmark (o, x, y). The facial expressions remain invariant for either the change of scale or slant of the individual head.

3.2 Analysis Of Distances Features

To recognize the facial expressions from video sequences, time delay neural network classifier is used with the 5 feature distances in inputs layer, and the 7 expressions in output layer trained by the backpropagation gradient algorithm (with 90% of good recognition). Once the facial expressions are recognized, the corresponding control signals are sent to a real mobile robot (Pioneer-3DX) using serial port via the Robot Operating System (establishing communication) which makes the robot move forward, backward, left, right.

The control signals are a sequence of the distance features that define a scenario displacement of robot corresponding to recognized facial expressions like:

Move forward:

IF D5 (surprise) \gg D5 (fear) \Leftrightarrow D5 (surprise) \gg $T_1=1,991034 \Rightarrow$ Move forward.

Move backward:

IF D5 (joy) \ll D5 (disgust) \Leftrightarrow D5 (joy) \ll $T_2=1,77888 \Rightarrow$ Move backward.

Turn right:

IF D1 (fear) \gg D1 (neutral) \Leftrightarrow D1 (fear) \gg $T_3=1,711864 \Rightarrow$ Turn right.

Turn left:

IF D2 (disgust) \gg D2 (neutral) \Leftrightarrow D2 (disgust) \gg $T_4=0,561078 \Rightarrow$ Turn left.

Where T_i defines thresholds values deduced from analyzing the features distances evolution in all facial expressions recognized of an individual to the database (see Fig. 10).

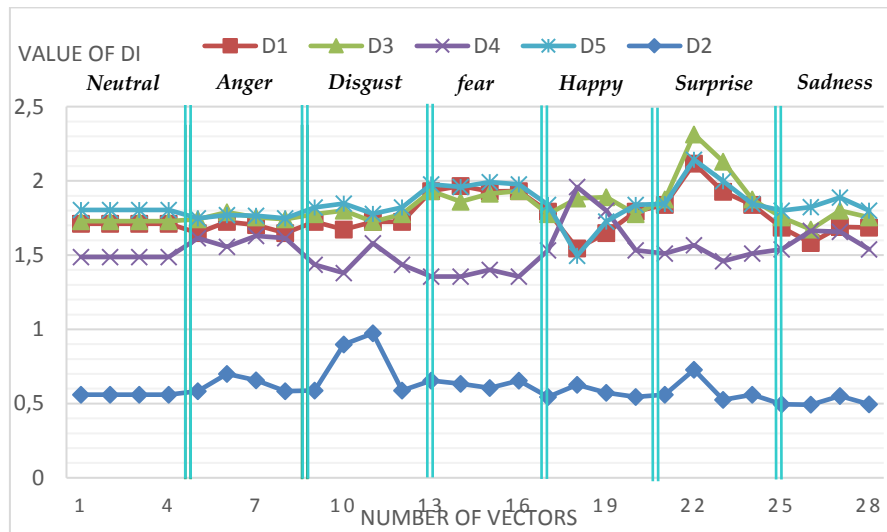


Fig. 10. Evolution of D_i in the 7 expressions 'Anton'.

An example of scenarios that takes place in the Intelligent Systems Research Laboratory (LARESI) is seen in (Fig. 11): forward \rightarrow forward \rightarrow turn right \rightarrow forward \rightarrow turn left \rightarrow forward, corresponding to: D5, D5, D1, D5, D2, D5, D5. We could check that, the robot reacts to all these commands successfully with 100% accuracy.

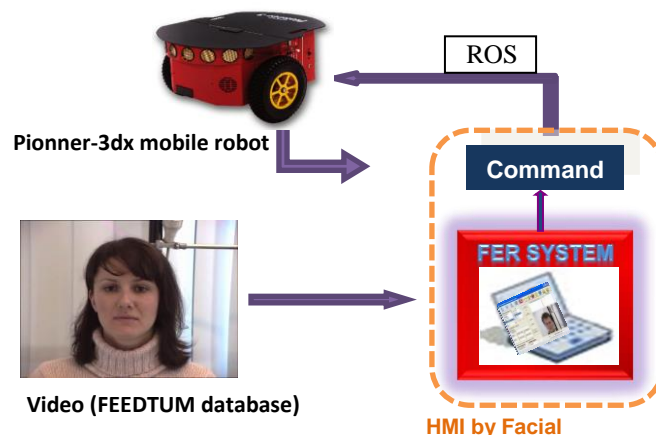


Fig. 11. Mobile robot command by the facial expression recognition system.

4 Conclusion

In this paper, the command of mobile robot using facial expressions, is the second application that we achieved in the setting of human machine interaction, intended especially to people having a reduced mobility (motor handicap) and that have only face movements to express their intents. Moreover, we conceived a dynamic facial expression model using active contour segmentation (GVF snake) by following a temporal extraction of the regions of interests. Then we got the 5 distance features while applying the MPEG-4 norm on the result of extraction automatically. These distance features are sent as signal command via ROS to a mobile robot. The results are very encouraging to achieve this application in real world from live video, like intelligent wheelchairs, human computer interaction and security systems.

References

1. Mittal, R., Srivastava, P., George, A., Mukherjee, A.: Autonomous robot control using facial expressions. *International Journal of Computer Theory and Engineering*, Vol. 4, No. 4, (2012).
2. Ju, D. J. S., Kim, E. Y., Intelligent wheelchair (IW) interface using face and mouth recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551-564, (1999).

3. Faria, P. M., Braga, R. A., Valgode, M. E., Reis, L. P.: Interface framework to drive an intelligent wheelchair using facial expressions. In proc. Third IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), Nara, Japan, pp. 124-129, (1998).
4. Shim, B., Kang, K. K., Lee, W. W., Won, S. J.B., Han, S.H.: An intelligent control of mobile robot based on voice command. In proc. International Conference on Control, Automation and Systems 2010, South Korea, pp. 2107, (2010).
5. Irfan, E., Pentland, A.: Coding, Analysis Interpretation and Recognition of Facial Expressions. IEEE Trans, Vol.19, pages 757-763. On PAMI (1997).
6. Black, M.J., Yacoob, Y.: Recognizing facial expressions in image sequences using local parametrized models of image motion. International Journal of Computer Vision, pages 23-48, (1997).
7. Tian, Y., Kanade, T. Cohn, J.: Recognition Actions Units for Facial Expression Analysis, IEEE Trans, Vol.23, N°2, pages 97-115. On PAMI, (2001).
8. Lanitis, A., Taylor, C. J., Coots, T. F.: Automatic interpretation and coding of faces images using flexible models. IEEE: Trans, pages 743-756. On PAMI, (1997).
9. Shinza, Y., Saito, Y., Kenmochi, Y., Kotani, K.: Facial expression analysis by integrating information of feature-point positions and grey levels of facial images. IEEE ICIP, Vancouver, Canada (2000). Author, F.: Article title. Journal 2(5), 99-110 (2016).
10. Zekhnine, C., Berrached, N. E.: Realization of facial expressions recognition system by MLP neural network. In Fifth Conference on Electric Engineering, polytechnic military School, Bordj El Bahri, Algiers, Algeria (2005).
11. wallhoff@ei.tum.de. <http://www.mmk.ei.tum.de/~waf>.
12. Zekhnine, C., Berrached, N. E.: Command of an arm manipulator robot by facial expression. In second International Conference on Image and Signal Processing and their Applications. University Mohamed Khider of Biskra, Algeria (2010).
13. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. IJCV, 1(4):321-331, (January 1988).
14. Cohen, I. : Modèles Déformables 2-D et 3-D: Application à la Segmentation d'Images Médicales. (1992).
15. Vylder, J.D., Ochoa, D., Wilfried, P., Chaerle, L., Van Der Straeten, D.: Tracking Multiple Objects Using Intensity-Gvf Snakes. EUSIPCO (2009).
16. Xu, C., Pince, J. L.: Generalized gradient vector flow external forces for active contours. Signal Processing 71, pages 131-13, (1998).
17. Hammal, Z., Caplier, A., Rombaut, M. : Classification d'expressions faciales par la théorie de l'évidence. LIS INPG, Grenoble, (2004).