# AUTOMATIC LIP TRACKING AND ACTION UNITS CLASSIFICATION USING TWO-STEP ACTIVE CONTOURS AND PROBABILISTIC NEURAL NETWORKS

Hadi Seyedarabi
*Faculty of Electrical and Computer Engineering*
*University of Tabriz, Tabriz, Iran*
email: hadis@discover.uottawa.ca

WonSook LEE
*School of Information Technology and Engineering (SITE)*
*Faculty of Engineering,*
*University of Ottawa, Canada*
email: wslee@uottawa.ca

Ali Aghagolzadeh
*Faculty of Electrical and Computer Engineering*
*University of Tabriz, Tabriz, Iran*
email: aghagol@tabrizu.ac.ir

## Abstract

*The most of the human emotions are communicated by changes in one or two of discrete facial features. Theses changes are coded as Action Units (AUs). Among the facial features mouth has most flexible deformability and it is highly complicated to track.*

*In this paper, we develop a lip shape extraction and lip motion tracking system both in static and dynamic facial images, based on a novel two step active contours model. A knowledge based system is used for estimating initial position of mouth. An oval shaped initial active is considered inside the estimated mouth region.*

*At the first step active contour locks onto stronger upper lip edges by using both high threshold Canny edge detector and balloon energy for contour deflation. Then using lower threshold image gradient as well as balloon energy for inflation, snake inflates and locks onto weaker lower lip edges.*

*Extracted lip feature points are used to extract some geometric feature points to form a feature vector which is used to classify lip images into AUs, using Probabilistic Neural Networks (PNN).*

*Experimental results show robust edge detection and reasonable classification where an average AUs recognition rate is 85.98% in image sequences and 77.44% in static images.*

*Keywords: Active contours, Action Units, lip tracking, Probabilistic Neural Networks,*

## 1. Introduction

Most computer-vision based approaches to facial expression analysis attempt to recognize only prototypic emotions. These prototypic emotional seem to be universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise, and anger. In everyday life, however, such prototypic expressions occur relatively infrequently. Instead, emotion is communicated by changes in one or two discrete features. Theses changes are coded as AUs.

Action units are visible appearance of facial muscles. Among the facial features mouth has most flexible deformability and it is highly complicated to track and shape extraction of lips and therefore 17 AUs among 44 AUs for facial features in the Facial Action Coding System (FACS) are directly associated with the movement of mouth [1]. Table 1 shows AUs used in this work that occur in the lower face and are more important in describing facial expressions.

Many studies have been done on lip tracking. Some recent studies used templates [2][3] and active contours [4]. The use of templates is valid in many cases. However, it usually requires a great amount of training and may not result in exact fit of the features. The use of active contours is appropriate especially when the feature shape is hard to represent with a simple template. Nevertheless, this method is sensitive to certain salient regions close to the desired feature. Random noise may strongly affect the deformation of active contours.

In this paper, we develop a lip shape extraction and lip motion tracking system both in static and dynamic facial images, based on a novel two step active contours model. A knowledge based system is used for estimating initial position of mouth. After that, a high threshold Canny edge detector is used for extracting mouth upper and middle lips edges. An oval shaped initial active contour is considered inside the estimated mouth region. Four energy terms are used to control motion of control points. At the first step active contour locks onto upper lip. Then using lower threshold image gradient snake inflates and locks onto lower lip edges.

Extracted lip feature points are used to extract some geometric feature points to form a feature vector which is used to classify lip images into AUs, using PNN.

Table 1. Some of FACS AUs used in this work

| AU | NAME |
|----|------|
| 12 | Lip corner pull |
| 15 | Lip corner depress |
| 17 | Chin raise |
| 20 | Lip stretch |
| 23 | Lip tighten |
| 24 | Lip press |
| 25 | Lips part |
| 26 | Jaw drop |
| 27 | Mouth stretch |

Experimental results show robust edge detection and reasonable classification rates in both static and dynamic face images where an average AUs recognition rate is 85.98% in image sequences and 77.44% in static images.

## 2. Lip Tracking System

### 2.1. Lip Position

A knowledge based system on a lip position extracted from lip positions on 97 subjects in Cohn-Kanade face database [5] is used for estimating initial position of mouth. Then a rectangular is automatically located around the estimated mouth location. After that, a high threshold Canny edge detector is used for extracting mouth strong edges and improving initial mouth position. Using high threshold edge detector moves initial snake toward upper lip by removing some artifacts like weak edges and shadow effects and we did not need to locate the initial snake very close to lip edges. We use a median filter before edge detection to remove random noises and a Gaussian filter used in Canny edge detector to decrease system sensitivity to illumination difference. These strong edges belong to upper and middle lips edges. Fig.1 shows high threshold Canny edge detector for lip region in different illumination images.

An oval shaped initial active contour (snake) with 28 control points is considered inside the estimated mouth region. These control points are equally spaced except in left and right corners where density of these points is doubled to maximize better tracking results on lip corners

### 2.2. Active Contours

The active contour model algorithm, first introduced by Kass et al. [6], deforms a contour to lock onto features of interest within in an image Usually the features are lines, edges, and/or object boundaries. An active contour is an ordered collection of n points in the image plane:

$$V = \{v_1, v_2, ..., v_n\}$$
$$v_i = (x_i, y_i), i = 1, 2, ... n \tag{1}$$

The points in the contour iteratively approach the boundary of an object through the solution of an energy minimization problem. For each point in the neighborhood of $v_i$, an energy term is computed:
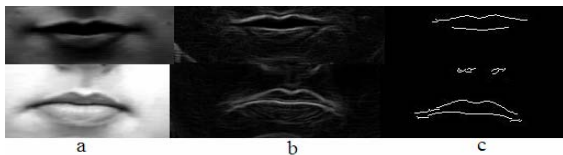
$$E_i = E_{int}(v_i) + E_{ext}(v_i) \tag{2}$$



Fig. 1. a) Original lip image b) magnitude of Gradient c) Canny edge detector with T=0.2

Where $E_{int}(v_i)$ is an energy function dependent on the shape of the contour and $E_{ext}(v_i)$ is an energy function dependent on the image properties, such as the gradient, near point $v_i$.

The internal energy function used herein is defined as follows:

$$E_{int}(v_i) = cE_{con}(v_i) + bE_{bal}(v_i) \tag{3}$$

Where $E_{con}(v_i)$ is the continuity energy that enforces the shape of the contour and $E_{bal}(v_i)$ is a balloon force that causes the contour to grow or shrink, **c** and **b** provide the relative weighting of the energy terms.

The external energy function attracts the deformable contour to interesting features, such as object boundaries, in an image. Image gradient and intensity are obvious characteristics to look at. The external energy function used herein is defined as follows:

$$E_{ext}(v_i) = mE_{img}(v_i) + gE_{grad}(v_i) \tag{4}$$

where $E_{img}(v_i)$ is an expression that attracts the contour to high or low intensity regions and $E_{grad}(v_i)$ is an energy term that moves the contour towards edges. Again, the constants, **m** and **g**, are provided to adjust the relative weights of the terms.

### 2.3. Two- Step Lip Active Contour

In this paper, we develop a lip shape extraction and lip motion tracking system both in static and dynamic facial images, based on a novel two- step active contours model. Four energy terms are used to control motion of control points. The points in the contour iteratively approach the outer mouth edges through the solution of a two-step energy minimization problem. One of the advantages of the proposed method is that we do not need to locate the initial snake very close to lip edges. At the first step active contour locks onto stronger upper lip edges by using both high threshold Canny edge detector and balloon energy for contour deflation. Then using lower threshold image gradient as well as balloon energy for inflation, snake inflates and locks onto weaker lower lip edges. In this stage 14 upper control points were fixed and only lower 14 points inflates to find lower lip edges. Fig. 2 and 3 show flowchart of proposed two- step algorithm and results.
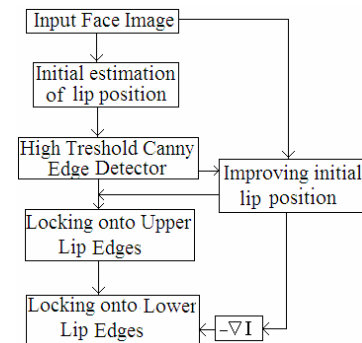


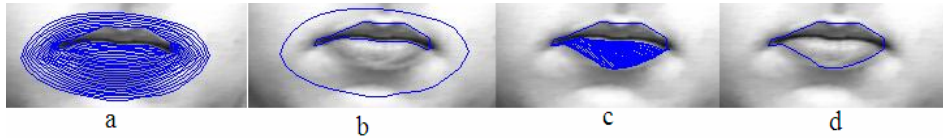Fig. 2. Two-step active contour algorithm

Fig. 3. Two step lip tracking a- deflating initial snake and finding upper strong edges b- initial and final snake at the end of first step c- fixing 14 upper control points and inflating 14 lower points to find lower weak edges d- final lip contour



Fig. 4. Lip tracking in image sequences using proposed algorithm

In image sequences two-steps active contour is applied in the first frame (which is supposed that mouth is not open) and then the final snake is used as an initial snake in the next frame. Fig. 4 shows some results of tracking in image sequences.

## 3. Feature Extraction

Extracted lip feature points are used to extract some geometric feature points to form a feature vector, representing lip shape in static and dynamic face images. Fig. 5 shows selected 9 feature points in lip contour.

Using geometrical position of these feature points, we extract the following features:

- Openness of mouth: average vertical distance of points 1-8 and 4-8.
- Width of mouth: horizontal distance of points 3 and 6.
- Chin rise: vertical distance of point 8 from origin.
- Lip corners distance: average vertical distance of points 3 and 6 from origin
- Quadratic curvature parameters for points 1, 2 and 3.
- Quadratic curvature parameters for points 3, 7 and 8.

Mid-Point between inner eye corners is determined as origin and distance between inner eye corners (IED) is used for normalizing features in static images. Fig. 6 shows the origin and normalizing factor.

In the image sequences calculated features (except curvature parameters) in the first and the last frames are normalized using the following equation:

$Norm\_fature = (Last\_frame - First\_frame)/First\_frame$ (5)

Quadratic curvature parameters are calculated in static images as well as only the last frame in image sequences. To calculate and normalize these parameters, origin is transferred to point 3 that reduces curvature parameters from 3 to 2, also horizontal distance of points 3 and 8, is normalized to one.

Calculated features form a $8 \times 1$ feature vector which is used in classifying action units.
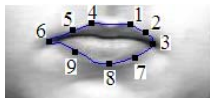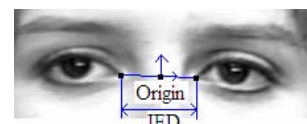


Fig. 5. Mouth feature points

## 4. Classification of AUs

### 4.1. PNN Classifier

PNN is a variant of the Radial Basis Function Neural Networks (RBFNN) and attempts have been carried out to make the learning process in this type of classification faster than normally required for the multi-layer feed forward neural networks.

The construction of PNN involves an input layer, a hidden layer and an output layer with feed forward architecture. The input layer of this network is a set of R units, which accept the elements of an R-dimensional input feature vector. The input units are fully connected to the hidden layer with Q hidden units (RBF units). Q is the number of input/target training pairs. Each target vector has K elements. One of these elements is 1 and the rest are 0. Thus, each input vector is associated with one of K classes.

When an input vector is presented in the input layer, the hidden layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The output layer sums these contributions for each class of inputs to produce its net output as a vector of probabilities. Finally, a compete transfer function on the output of the output layer picks up the maximum of these probabilities, and produces a 1 for that class and 0 for the other classes [3].

### 4.2. Image Database

In this work, we used Cohn-Kanade database that consists of expression sequences of subjects, starting from a neutral expression and ending with the peak of the facial expression. Subjects sat directly in front of the camera and performed a series of the facial expressions that included the six primary and also some single AUs. We used a subset of 180 image sequences containing six basic emotions for 30 subjects



Fig. 6. Origin and normalization factor

emotions. Those AUs which are important to the communication of the emotion and were occurred at least 25 times in our database are selected. This frequency criterion ensures sufficient data for training and testing. For each person there are on average of 12 frames for each expression (after eliminating alternate frames). Image sequences for the frontal views are digitized into $640 \times 490$ pixel array with 8 bits grayscale.

## 4.3. Experimental Results

We used the sequence of 144 (80%) subjects as training sequences, and the sequence of the remaining 36 (20%) subjects as test sequences. This test is repeated five times, each time leaving different subjects out. The number of the input layer units in the PNN classifier is equal to 8, the number of extracted features, the number of the hidden layer units equals to $144 \times 9$, the number of training pairs and that of the output layers is 9, which corresponds to selected 9 lower face AUs.

Last frame in image sequences which represents peak of facial expressions is used to evaluate performance of the system for static images. Table 2 shows the recognition rate for static images and image sequences.

## 5. Conclusions

In this paper, we developed a lip shape extraction and lip motion tracking system both in static and dynamic facial images, using a novel two step active contours model. A knowledge based system was used for estimating initial position of mouth. After that, a high threshold Canny edge detector was used for extracting upper and middle lips edges. An oval shaped initial active contour was considered inside the estimated mouth region. At the first step active contour locked onto stronger upper lip edges. Then using lower threshold image gradient, snake inflated and locked onto weaker lower lip edges. Extracted lip feature points were used to extract

some geometric feature points to form a feature vector which was used to classify lip images into AUs, using PNN.

Proposed two-step active contour showed robust tracking results in different illumination conditions and we did not need to locate the initial snake very close to lip edges.

Experimental results showed robust edge detection and reasonable classification rates in both static and dynamic face images where an average AUs recognition rate is 85.98% in image sequences and 77.44% in static images.

## References

[1] P. Ekman and W.V. Friesen, *FACIAL ACTION CODING SYSTEM (FACS).* Consulting Psychologists Press, Inc., 1978.
[2] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Facial Expressions Animation and Lip Tracking Using Facial Characteristic Points and Deformable Model*,".* *Int. Journal of Information Technology (IJIT)*, Vol.1 No. 4, 284-287, 2004
[3] Aghagolzadeh, A., Seyedarabi, H.; Khanmohammadi, S." Single and Composite Action Units Classification in Facial Expressions by Feature-Points Tracking and RBF Neural Networks," *Ukrainian Int. conf. signal/Image processing, UkrObraz 2004*, Kiev, Ukraine 2004, 181-184.
[4] Zhilin Wu, Petar S. Aleksic, and Aggelos K. Katsaggelos "Lip Tracking for MPEG-4 Facial Animation," *Proc. of the Fourth IEEE Int. Conf. on Multimodal Interfaces (ICMI'02),* 293-298
[5] T. Kanade, J. Cohn, and Y. Tian. *Comprehensive database for facial expression analysis*, 2000
[6] Michael Kass, Andrew Witkin, and Demtrri Terzopoulos. "Snakes: Active contour models*," Int. J. of Computer Vision*, 321-331, 1988.

Table 2: Recognition results for image sequences and static images

| AU | Image sequences | | Static Images | |
|---|---|---|---|---|
| 12 | 31/35 | 88.57% | 31/35 | 88.57% |
| 15 | 27/29 | 93.10 % | 26/29 | 89.66% |
| 17 | 73/82 | 89.02% | 63/82 | 76.83% |
| 20 | 26/30 | 86.67 % | 18/30 | 60.00% |
| 23 | 24/29 | 82.76% | 21/29 | 72.41% |
| 24 | 23/32 | 71.88% | 20/32 | 62.50% |
| 25 | 52/59 | 88.14% | 46/59 | 77.96% |
| 26 | 6/10 | 60.00% | 7/10 | 70.00% |
| 27 | 20/22 | 90.9%1 | 22/22 | 100% |
| **Average** | 282/328 | 85.98 % | 254/328 | 77.44% |