Facial Mimics Recognition From Face Profile Image Sequences

Technical Report
TR-DKS-02-01

Delft, January 11th 2002

Maja Pantic
Delft University of Technology
Information Technology and Systems
Department of Mediamatics
Data and Knowledge Systems group

Ioannis Patras
University of Amsterdam
Computer Science Department
Intelligent Sensory Information
Systems group

Leon J.M. Rothkrantz
Delft University of Technology
Information Technology and Systems
Department of Mediamatics
Data and Knowledge Systems group

Facial Mimics Recognition from Face Profile Image Sequences

Maja Pantic
Delft University of Technology
ITS / Mediamatics Dept.
Delft, the Netherlands
M.Pantic@cs.tudelft.nl

Ioannis Patras
University of Amsterdam
Computer Science Dept.
Amsterdam, the Netherlands
yiannis@science.uva.nl

Leon Rothkrantz
Delft University of Technology
ITS / Mediamatics Dept.
Delft, the Netherlands
L.J.M.Rothkrantz@cs.tudelft.nl

Abstract

A robust way to discern facial mimics (facial muscle activity) in images of faces, insensitive to scale, pose, and occlusion, is still the key research challenge in automatic facial expression analysis domain. A practical method recognized as the most promising one for addressing this problem is through facial mimics analysis of multiple views of the face. However, portraits of faces or nearly frontal-views of faces constitute input data processed by current systems for automatic facial mimics analysis. To advance the existing technological framework upon which further research on automatic facial mimics analysis from multiple facial views can be based, we developed an automatic system as to analyze subtle changes in facial expressions based on profile-contour fiducial points in a profile-view video. A probabilistic classification method based on statistical modeling of the color and motion properties of the profile in the scene is proposed for tracking the profile face. From the segmented profile face, we extract the profile contour and from it, we extract 10 profile-contour fiducial points. Based on these face profile features, 20 individual facial-actions occurring alone or in a combination are recognized by a rule-based method. A recognition rate of 85% is obtained.

1. Introduction

The human face is a rich source of information about human behavior. Nonverbal facial cues regulate our social interactions [1], form our primary means to communicate emotion [2], and form visible speech signals that we use to confirm what has been said by means of lip reading [3]. Automatic analyzers of subtle facial changes, therefore, seem to have a natural place in various vision systems including the automated tools for psychological research, lip reading, bimodal speech analysis, affective computing, videoconferencing, face and visual speech synthesis, and human-behavior-aware next-generation interfaces. It is this wide range of principle driving applications that has caused an upsurge of interest in the research problems of automatic facial mimics visual analysis.

This problem domain usually includes three sub-problem areas [4]: finding faces, detecting facial features, and interpreting this information in terms of facial subtle changes or microactions such as the facial action units of the FACS system [5].

Discerning the existence and location of a face in the scene is an autonomous perceptual human ability that is rather difficult to duplicate computationally. Yet this task is a precursor to encoding the information that the facial display provides. The typical issue here is that of scale, pose, and occlusion; rigid head and body motions of the observed person usually cause changes in the viewing angle and visibility of the tracked face. As noted in [6], perhaps the most promising method for addressing this problem is through the use of multiple cameras yielding multiple views of the face. To date, however, most works on automatic facial mimics analysis have largely avoided dealing with facial views other than a frontal one: portraits [7, 8, 9] or nearly frontal-views of faces [10, 11] constitute input data processed by the current systems.

The problem of automatic facial feature detection in input images has at least three dimensions: (1) is temporal information used, (2) are the features holistic (spanning the whole face) or analytic (spanning subparts of the face), and (3) are the features view (2D) or volume (3D) based. Given this glossary, most existing methods for automatic facial mimics visual analysis are directed towards temporal analysis of analytic 2D facial features. The proposed strategies include optical flow analysis across the entire face or local facial feature regions [7, 8, 10], holistic spatial analysis [8], analytic spatial analysis [9], and analytic spatio-temporal analysis [11].

From several existing methods for recognition of facial mimics based on visually observable facial muscular activity, the FACS system [5] is the most commonly used in psychological research [12]. Following this trend, all of the existing methods for automatic facial mimics visual analysis, including the method proposed in this paper, interpret the facial display information in terms of the facial action units (AUs) of the FACS system. Yet none automatic system is capable of encoding the full range of facial mimics, i.e., none is capable of recognizing all 44 AUs that account for the changes in facial display. From the previous works, the automatic facial mimics analyzers presented in [11] and [9] perform the best in this aspect: they code 16 and, respectively, 27 AUs occurring alone or in a combination in frontal-view facial images.

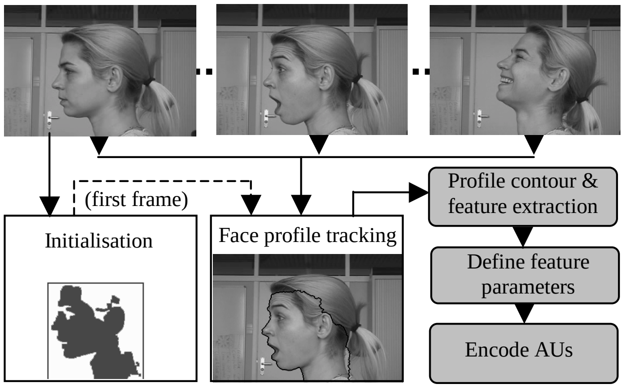

Fig. 1: Outline of the profile-based AU recognition

The research reported here addresses the problem of automatic AU coding from face profile image sequences. It was undertaken with two motivations: (1) experimental observations suggest that the facial muscular activity causing a temporal protrusion of the facial features (e.g., AU19: shown tongue, AU18: puckered lips, AU29: jaw pushed forward, etc.), which are difficult to detect in a frontal-view of the face [9], are clearly observable in a profile-view of the face; and (2) basic understanding of how to achieve automatic facial mimics analysis from human face profiles is necessary for the establishment of the technological framework upon which further research on automatic facial mimics analysis from multiple facial views can be based where procedures of greater flexibility and improved performance can evolve.

Fig. 1 outlines the proposed method. For the first frame of the input face profile image sequence a description of the local statistical properties of the tracked face profile is built based on a label field that is initiated by utilizing a HSV color-based segmentation of the face. A probabilistic classification method based on statistical modeling of the color and motion properties of the label field is used to track the label field in the rest of the sequence. The contour of the tracked face-profile region of the label field is extracted as the face profile contour. Under the assumption that the face images are in (nearly) left profile view, the left-hand-side points of the extracted profile contour that have a large curvature are extracted as feature points (profile-contour fiducials). Subtle changes in the tracked face profile are measured next. Motivated by AUs of the FACS system, these changes are represented as a set of mid-level feature parameters describing the state and motion of the feature points and the shapes formed by different sets of feature points. Based on these feature parameters, a rule-based algorithm interprets the facial display in terms of 20 AUs occurring alone or in a combination. Face profile tracking, profile contour and feature extraction, parametric feature representation and AU coding are explained in sections 2, 3, 4 and 5 respectively. Section 6 concludes the paper.

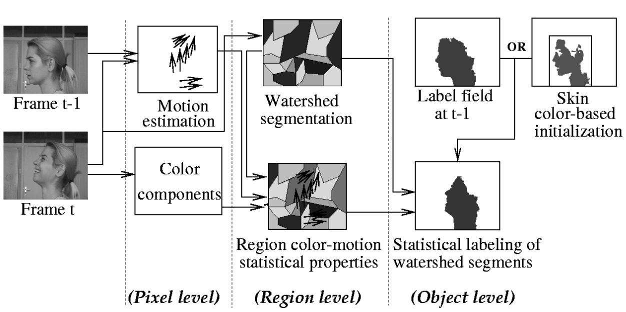

Fig. 2: Outline of the face profile tracking method

2. Face profile tracking

The first step in automatic facial mimics analysis is to ascertain the presence of a face in the scene and to locate it. In order to do so, we adapted a semi-automatic method for object-based segmentation of complex-scene image sequences [13] for the purpose of human face tracking. An outline of the adapted method is shown in Fig. 2.

Face region tracking is addressed as a segmentation problem in two objects: the Face and the Background. For the first frame of the sequence, markers of the two objects are extracted as follows. For the face region the marker is extracted as the largest connected image component with Hue, Saturation and Value within the range [5,35], [0, 0.7] and [0.1, 0.9] respectively. In the absence of a similar model for the Background, its marker is extracted as the bounding box of the Face marker. A color-based version of a watershed segmentation algorithm provides the final segmentation for the first frame [13]. The color-based watershed segmentation provides good localization of the face given that the most prominent color edge between the Background and the Face markers is indeed the face contour. Under the assumption that in the first frame of the sequence the face is seen in a nearly vertical position, this is usually the case. However, for the rest of the image sequence, rotations of the face (see Fig. 1) can result in bounding boxes that contain large parts of the background with probable strong edges. Therefore, for the rest of the image sequence, we perform the segmentation based on tracking of the local statistical color and motion properties of the Face and the Background.

The method operates at three levels. At *Level 1* (pixel level) a feature vector is estimated for each pixel in the current frame. At *Level 2* (region level) a watershed color

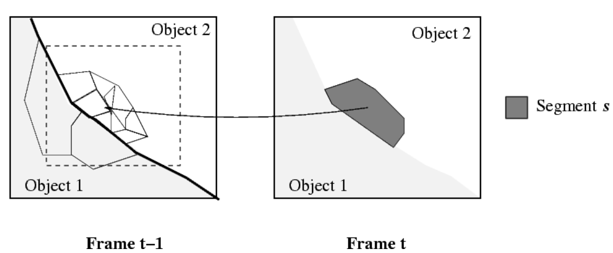

Fig. 3: Projection of segment *s* in the label field of the previous frame

Fig. 4: Results for the "pleasant surprise" sequence. 1st and 3rd columns: Superposition of the contour of the face profile region on the original frames. 2nd and 4th columns: Face profile region (i.e., the label field for segmentation in two objects). Results are shown for frames 1, 36, 42, 91, 105, and 115.

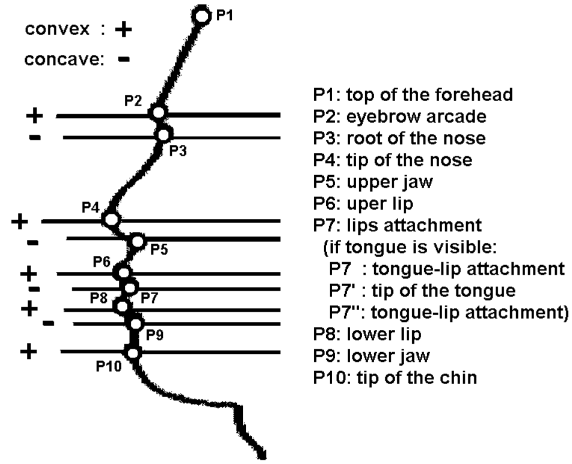

Fig. 5: Feature points (profile contour fiducials)

segmentation method decomposes the current frame in a number of color regions. The statistical properties of the color regions are estimated subsequently under the assumption that the same process, which is modeled as a multivariate Gaussian, generates the feature vectors at pixels inside the same region. At *Level 3* (object level) a labeling based on probabilistic classification of the color regions takes place. To wit, each color region is projected in the previous frame where an estimation of the label field is available (Fig. 3). For each object present in a window surrounding the area of projection, we estimate the parameters of the model (a multivariate Gaussian) describing the objects color and motion properties. Then, each color region is assigned the object label such that the joint probability of the label field and the observed color and motion features is maximized. Once each color region is labeled, the local models are re-estimated for the current frame and an iterative region-classification / model estimation procedure is performed until no color region changes its label. Fig. 4 illustrates the typical results of this method. For further details about this method and its performance, the reader is referred to [13].

3. Profile contour and feature extraction

The contour of the face profile region (referred to as "face profile contour" in the text below), generated by the face profile tracking method (Fig. 4), is utilized for further analysis of shown facial mimics. We proceed with feature points' extraction (Fig. 5) under two assumptions: (1) the face images are non-occluded nearly left profile view with possible in-plane head rotations, and (2) the first frame is in a neutral expression. After initializing the feature points in the first frame based on the face region of the initial label field, they are automatically extracted from the tracked face profile contour for the rest of the sequence.

To account for possible in-plane head rotations and variations in scale of the tracked face profile, the face profile contour is normalized in each frame based on two referential points (Fig. 5): the tip of the nose (P4) and the top of the forehead (P1). The major impulse to the usage of these referential points comes from their stability with respect to non-rigid facial features' movements: facial muscles' contractions do not cause physical displacements of these points [14]. The tip of the nose and the top of the forehead are extracted as the leftmost and, respectively, the uppermost leftmost point of the generated contour. To handle possible inaccuracies in detection of the referential points caused by inaccuracies in the segmentation of the face profile region, we exploit all: information from the previous frames, the knowledge about temporal dynamics of rigid head movements (usually they occur gradually in time) and the knowledge about the facial stability of the referential points. To wit, a small window (its height and width set to 3% of the length of P1P4 measured in frame ***t-1***) centered at the location of a referential point extracted from frame ***t-1*** is searched for the pertinent referential point in frame ***t***. If a referential point cannot be defined such that it belongs to the face profile contour determined for frame ***t***, the relevant referential point determined for frame ***t-1*** is used instead. Finally, the face profile contour is normalized by carrying out affine transformations of it such that the line P1P4 between the referential points discerned for the current frame is of the same length and orientation as the line P1P4 determined for the first frame.

To extract the feature points from the normalized face profile contour, we move from image to function analysis and treat the left-hand side of the normalized face profile contour (up to the determined referential point P1) as the profile contour function. First we extract the extremities of this function (i.e., the zero-crossings of the function's 1st order derivative). Then, given the *a priori* knowledge on where the convexities and concavities of a left face profile are, we analyze the extracted extremities to find out where the function is arched. The maximums and minimums of the function's 2nd order derivative are extracted as the feature points (Fig. 5). To ascertain correct extraction of the feature points when the tongue is visible (P7' and P7''

Search window basis: W_{PB}: 3% length($P1_{t1}P4_{t1}$) $\times$ 3% length($P1_{t1}P4_{t1}$)
P1 (defined already as the referential point): uppermost point of *f*
P4 (defined already as the referential point): leftmost point of *f*
P2: maximum of *f*" between P1 and P4; W_{P2}: 2 $*$ width(W_{PB}) $\times$ 4 $*$ height(W_{PB}); right-middle(W_{P2}) = $P2_{t1}$
P3: minimum of *f*" between P1 and P4; W_{P3} = W_{PB}; center(W_{P3}) = $P3_{t1}$
P10: lowermost maximum of *f*"; W_{P10}: 15 $*$ width(W_{PB}) $\times$ 15 $*$ height(W_{PB}); top-middle(W_{P10}) = $P10_{t1}$
P5: the first minimum of *f*" between P4 and P10; W_{P5}: width(W_{PB}) $\times$ 2 $*$ height(W_{PB}); right-middle(W_{P5}) = $P5_{t1}$
P9: the last minimum of *f*" between P4 and P10; W_{P9}: 15 $*$ width(W_{PB}) $\times$ 15 $*$ height(W_{PB}); center(W_{P9}) = $P9_{t1}$
P7: the first minimum of *f*" between P5 and P9; W_{P7}: 5 $*$ width(W_{PB}) $\times$ 5 $*$ height(W_{PB}); center(W_{P7}) = $P7_{t1}$
if P7 is the only minimum of *f*" between P5 and P9, P7' and P7" do not exist, otherwise:
P7": the second minimum of *f*" between P5 and P9; $W_{P7''}$ = W_{P7}; center($W_{P7''}$) = $P7_{t1}$
P7': maximum of *f*" between P7 and P7"; $W_{P7'}$ = W_{P7}; center($W_{P7'}$) = $P7_{t1}$
P6: maximum of *f*" between P5 and P7; W_{P6}: 4 $*$ width(W_{PB}) $\times$ 8 $*$ height(W_{PB}); center(W_{P6}) = $P6_{t1}$
P8: maximum of *f*" between P7 (P7" if exists) and P9; W_{P7}: 4 $*$ width(W_{PB}) $\times$ 8 $*$ height(W_{PB}); center(W_{P7}) = $P7_{t1}$

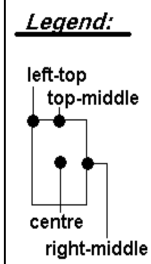

Fig. 6: Definitions of the feature points *P* and the related search windows W_P given in the order in which the points are extracted from the profile contour function *f* defined for frame *t*. Legend: *t1* denotes the first frame and *f*" denotes the 2nd order derivative of *f*.

Feature points motion	
up/down (P) = $y_{Pt1} - y_{Pt}$ If ***up/down(P)*** < 0, point moves up.	***in/out(P)*** = $x_{Pt1} - x_{Pt}$ If ***in/out(P)*** > 0, point moves outward.
Feature points state	
If P9 equals P7, ***absent(***P9***)***. If there is no maximum of f" between P5 and P7, ***absent(***P6***)***. Similarly for P7', P7" and P8 (see Fig. 5).	***increase/decrease(AB)*** = $AB_{t1} - AB_t$, where $AB = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$ If ***increase/decrease(AB)*** < 0, distance ***AB*** increases.
Shapes formed by feature points	
If for a set of points P = $\{p_i \mid i \in \{1,n\}, p_1=\boldsymbol{A}, p_n=\boldsymbol{B}\}$ $(\exists\, p_j, p_k \in P \mid 1<j<<n, 1<<k<n) \mid (\forall\, p_l, j \le l \le k, x_{pl} = x_{pl+1})$, then ***angular(AB)*** = *true*. The physic meaning of this parameter is shown in Fig. 8.	
If for a set of points P = $\{p_i \mid i \in \{1,n\}, p_1=\boldsymbol{A}, p_n=\boldsymbol{B}\}$ min_x(P_t) < min_x(P_{t1}), where min_x(P) is the minimal x-coordinate value that the points from P can have, then ***increased_curvature(AB)*** = *true*. The physic meaning of this parameter is shown in Fig. 8.	

Fig. 7: Parametric representation of face-profile-contour features for AU recognition

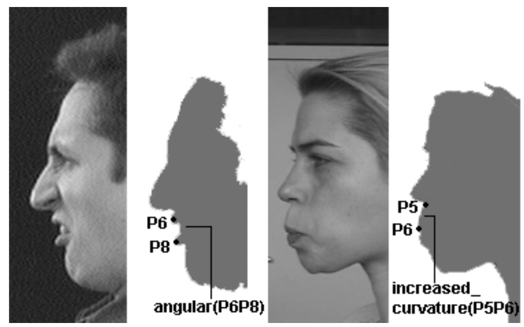

Fig. 8: Physic meanings of the "shape" mid-level feature parameters

exist), we extract the feature points in the particular order (Fig. 6). To handle possible inaccuracies in feature points' detection caused by inaccuracies in the segmentation of the face profile region (e.g., like in frame 91 shown in Fig. 4), we exploit both the knowledge about facial anatomy and information from the previous frames. To wit, for each feature point ***P***, a standard "search" window W_P has been defined with respect to the possible directions and magnitudes of the motion on the skin surface affecting the temporal location of ***P***. For instance, non-rigid movements of the eyebrows (raised and/or frowned eyebrows) cause upward and/or outward movement of P2 and the upward pull of the skin along the nose (i.e. wrinkled nose) causes outward and downward movement of P2. This and the fact that no facial muscle activity can push the eyebrow arcade either to the nasal bone or to the middle of the forehead, define the size and the positioning of the search window W_{P2} given in Fig. 6. The feature point P_t is determined further for frame ***t*** such that it represents a specific zero crossing (Fig. 5, 6) of the 1st order derivative of the profile contour function defined for frame ***t*** and, at the same time, belongs to the W_P set around the location of P_{t1} discerned for frame ***t1***. If P_t cannot be defined, P_{t-1} is used instead.

4. Parametric feature representation

Each AU of the FACS system is anatomically related to contraction of a specific facial muscle [5]. Contractions of facial muscles produce motion in the skin surface and deform the shape and location of the facial components (eyebrows, mouth, chin, etc.). Some of these changes in facial expression are observable from the changes in the tracked face profile contour and the related feature points. To classify detected changes of the face profile contour in terms of facial muscle activity (i.e., in terms of AUs of the FACS system), these changes should be represented first as a set of suitable feature parameters.

We defined 6 mid-level feature parameters in total: two describing the motion of the feature points, two describing their state, and two describing shapes formed by a set of feature points. The definitions of the parameters, which are calculated for each frame, are given in Fig. 7.

5. Facial mimics recognition

The last step in automatic facial mimics analysis is to translate the extracted facial expression information (i.e., the calculated feature parameters) into a description of shown facial changes such as the AU-coded description of shown facial expression. To achieve this, we utilize a rule-based method that encodes 20 AUs occurring alone or in a combination in the current frame of the input sequence.

5.1. Utilized rule-based method

Motivated by the rules of the FACS system, each of the rules utilized for AU recognition is defined in terms of the predicates of the mid-level representation (Fig. 7) and each encodes a single AU in a unique way according to the relevant FACS rule. The FACS rules for 20 AUs to be recognized and the pertinent rules utilized by our method are given in Fig.9.

To represent the relations between the utilized rules, we apply a relational list (R-list). The utilized R-list is a four-tuple list, where the first two columns identify the conclusion clause of a certain rule that forms the premise clause of another rule identified in the next two columns of the R-list. Based on this R-list, the rule-based method proposed for AU recognition applies fast direct chaining as the inference procedure. To wit, the procedure starts with the first internally stored rule and then searches the R-list to find if the conclusion of the fired rule forms a premise of another rule that will be fired in the next loop.

AU1 rule 1	Pulls the inner portion of the eyebrows upward, causes the skin in the centre of the forehead to wrinkle horizontally. IF ***up/down***(P2) < 0 THEN AU1
AU4 rule 2	Pulls the eyebrows closer together, produces a bulge between the eyebrows, lowers the eyebrows slightly. IF ***in/out***(P2) > 0 AND ***increase/decrease***(P2P3) ≤ 0 THEN AU4
AU8 rule 3	Pulls the lips towards each other, parts the lips. IF NOT (AU9 OR AU12 OR AU15 OR AU17 OR AU18 OR AU20) AND ***angular***(P6P8) = *true* THEN AU8
AU9 rule 4	Wrinkles the nose, lowers the brows, produces a bulge between the brows and the root of the nose, raises the upper lip. IF ***in/out***(P2) > 0 AND ***increase/decrease***(P2P3) > 0 THEN AU9
AU10 rule 5	Raises the upper lip, deepens the nasolabial furrow, does not wrinkle the nose. IF ***increase/decrease***(P2P3) ≤ 0 AND ***increase/decrease***(P5P6) > 0 AND ***in/out***(P6) > 0 THEN AU10
AU12 rule 6	Pulls the lip corners upward obliquely. IF ***in/out***(P6) < 0 AND ***in/out***(P8) < 0 AND ***increase/decrease***(P5P6) > 0 THEN AU12
AU15 rule 7	Pulls the corners of the lips downward, stretches the lips slightly, flattens the skin of the chin boss. IF ***up/down***(P6) > 0 AND ***up/down***(P8) > 0 AND ***increased_curvature***(P5P6) = *false* THEN AU15
AU16 rule 8	Pulls the lower lip downward laterally, causes the lower lip to protrude. IF ***increase/decrease***(P8P10) > 0 AND ***up/down***(P8) > 0 AND ***in/out***(P8) > 0 THEN AU16
AU17 rule 9	Pushes the chin boss and the lower lip upward and stretches the skin on the chin boss. IF NOT (AU28 OR AU28t OR AU28b) AND ***in/out***(P10) < 0 THEN AU17
AU18 rule 10	Pushes the mouth forward medially, de-elongates the mouth, causes the lips to protrude forwards (as by saying “fool”). IF ***in/out***(P6) > 0 AND ***in/out***(P8) > 0 AND ***increase/decrease***(P5P6) ≤ 0 AND ***increase/decrease***(P8P10) ≤ 0 AND ***increase/decrease***(P6P8) < 0 THEN AU18
AU19 rule 11	Causes at least the tip of the tongue to be visible. IF ***absent***(P7’) = *false* AND ***absent***(P7”) = *false* THEN AU19
AU20 rule 12	Pulls the lips backward laterally, flattens the skin of the lips and the chin boss. IF ***in/out***(P6) < 0 AND ***in/out***(P8) < 0 AND ***increase/decrease***(P5P6) ≤ 0 AND ***increase/decrease***(P6P8) ≤ 0 AND ***increased_curvature***(P5P6) = *false* THEN AU20
AU23 rule 13	Tightens the lips slightly making the lips appear more narrow. IF NOT (AU28 OR AU28t OR AU28b) AND ***increase/decrease***(P6P8) > 0 AND ***increase/decrease***(P6P8) < ½$P6_{t1}P8_{t1}$
AU24 rule 14	Presses the lips together, tightens and narrows the lips to a small extent. IF NOT (AU28 OR AU28t OR AU28b) AND ***increase/decrease***(P6P8) > 0 AND ***increase/decrease***(P6P8) ≥ ½$P6_{t1}P8_{t1}$
AU25 rule 15	Parts the lips, does not parts the jaws. IF ***increase/decrease***(P6P8) > 0 AND ***increase/decrease***(P4P10) ≥ 0 THEN AU25
AU26 rule 16	Parts the lips, parts the jaws, does not stretches the mouth. IF ***increase/decrease***(P4P10) < 0 AND ***increase/decrease***(P4P10) > -½height(W_{P10}) THEN AU26
AU27 rule 17	Stretches the mouth as lower jaw is pulled down. IF ***increase/decrease***(P4P10) < 0 AND ***increase/decrease***(P4P10) ≤ -½height(W_{P10}) THEN AU27
AU28,t,b rule 18	AU28: lips sucked into the mouth. AU28t: Upper lip sucked into the mouth. AU28b: Bottom lip sucked into the mouth. IF ***absent***(P6) AND ***absent***(P8) THEN AU28. IF ***absent***(P6) THEN AU28t. IF ***absent***(P8) THEN AU28b.
AU29 rule 19	Pushes the jaw forward making the chin to stick out and the lower teeth to extend in front of upper teeth. IF ***in/out***(P10) > 0 THEN AU29
AU36t,b rule 20	AU36t: Pushes the tongue under the upper lip, causes a bulge above the upper lip. AU36b: Pushes the tongue under the lower lip, causes a bulge below the lower lip. IF ***increased_curvature***(P5P6) = *true* THEN AU36t. IF ***absent***(P9) THEN AU36b.

Fig. 9: The descriptions of 20 AUs to be recognized and the related rules utilized for AU recognition

If such a relation does not exist, the procedure tries to fire the rule that in the internal storage comes after the rule that has fired last. A rule is fired only if all clauses of the rule's premise are true. To prevent firing of a rule more than once, we utilize a list of fired rules (LFR). If a rule has fired, the rule number is added to the LFR.

5.2. Test data set and recognition results

Although AU-coded facial expression image databases are available in general, these databases contain portraits or nearly frontal-views of human faces. Since these data are not suitable for testing our face-profile-based AU encoder, we generated our own test data.

The test data set has been created with the help of 5 certified FACS coders drawn from college personnel. The acquired face-profile image sequences represent a number of demographic variables including ethnic background (European, Asian and South American), gender (60% female) and age (20 to 35 years). The subjects were asked to perform series of facial expressions that included single AUs and combinations of those. Forty image sequences of variable length (110 to 240 frames) of nearly left-profile view of subjects' faces were recorded by utilizing a CCD digital PAL camera. The size of the face region in each frame was 135×175 pixels or more. Each sequence began with a neutral expression with no head rotation.

Metadata were associated with the acquired test data given in terms of AUs scored by 4 certified FACS coders (other than the coder captured in the judged sequence). As the actual test data set, we use 32 image sequences for which the overall inter-coders' agreement about displayed AUs is above 75%. The AU-coded descriptions of shown facial expressions obtained by human FACS coders were compared further to those produced by our method. The results of this comparison are given in Table 1. Most of misidentifications arise from confusions between similar AUs (e.g., AU10 and AU18) and from subtle activations that remained unnoticed by human observers (e.g., AU26 behind the closed mouth, AU24 instead of AU23).

Table 1: Recognition results for the upper face AUs (AU1, AU4, AU9), the AUs affecting the mouth and those affecting the jaw (AU17, AU26, AU27, AU29): # denotes the number of AUs' occurrences, C denotes correctly recognized AUs' occurrences, M denotes missed AUs' occurrences, IC denotes incorrectly recognized AUs' occurrences.

	#	C	M	IC	Rate
upper face	18	15	1	2	83.3%
mouth	82	66	5	11	80.5%
jaw	36	33	1	2	90.9%
Total:	136	114	7	15	**84.9%**

6. Conclusions

In this paper, we introduced an automatic system for analyzing subtle changes in facial expression based on changes in face profile contour tracked in a nearly left-profile-view image sequence. The significance of this contribution is in the following: (1) it extends the state of the art in automatic facial mimics visual analysis in several directions, including the number of AUs, the difference in AUs and the facial view handled, and (2) it provides the basics of how to achieve automatic AU coding in face profile image sequences upon which further research on facial mimics analysis from multiple facial views can be based.

However, the algorithm presented in the paper may be refined. For instance, once the referential points have been located, they can be used for stabilizing the face profile region yielding a more robust face profile tracking. Also, although the developed system demonstrates concurrent validity with manual FACS coding of test data set, more extensive field trials and more elaborate quantitative validation studies are necessary to confirm this finding.

References

[1] L. van Poecke, *Nonverbal Communication*, Garant-Uitgever, Apeldorn, the Netherlands, 1996.

[2] D. Keltner and P. Ekman, "Facial expression of emotion", *Handbook of Emotions*, Guilford Press, New York, USA, 2000, pp. 236-249.

[3] D.G. Stork and M.E. Hennecke, *Speech-reading by man and machine*, NATO/Springer-Verlag, New York, USA, 1996.

[4] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art", *IEEE TPAMI*, 2000, vol. 22, no. 12, pp. 1424-1445.

[5] P. Ekman and W. Friesen, *Facial Action Coding System*, Consulting Psychologist Press, Palo Alto, USA, 1978.

[6] A. Pentland, "Looking at people", *IEEE TPAMI*, 2000, vol. 22, no. 1, pp. 107-119.

[7] J. Cohn, A. Zlochower, J. Lien, and T. Kanade, "Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding", *Psychophysiology*, 1999, vol. 36, pp. 35-43.

[8] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis", *Psychophysiology*, 1999, vol. 36, pp. 253-263.

[9] M. Pantic and L.J.M. Rothkrantz, "Expert system for automatic analysis of facial expressions", *Image and Vision Computing*, 2000, vol. 18, no. 11, pp. 881-905.

[10] J.J. Lien, T. Kanade, J.F. Cohn, and C.C. Li, "Automated facial expression recognition based on FACS action units", *Proc. IEEE FG*, 1998, pp. 390-395.

[11] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis", *IEEE TPAMI*, 2001, vol. 23, no. 2, pp. 97-115.

[12] P. Ekman, "Methods for measuring facial action", *Handbook of methods in nonverbal behavior research*, Cambridge University Press, Cambridge, USA, 1982, pp. 45-90.

[13] I. Patras, E. Hendriks, and R. Lagendijk, "A semi-automatic method for segmentation of image sequences with local region-based classification", *Int'l Conf. Signal and Image Processing*, 2000, pp. 205-210.

[14] L. Harmon, et al., "Identification of human face profiles by computer", *Pattern Recognition*, 1978, vol. 10, pp. 301-312.