A Bayesian Approach to Recognise Facial Expressions using Vector Flows

Xiaofan Sun, Leon Rothkrantz, Dragos Datcu and Pascal Wiggers

Abstract: *Facial expressions play an important role in human nonverbal communication. They can be generated by activation and dilatation of facial muscles. In this paper we describe a system to recognize facial expressions automatically. Special areas in the face have been selected to extract features from the vector flow of visual muscle activity. To classify facial expressions Bayesian Networks have been used. The classifier has been trained and tested on video recordings from the Cohn Kanade database. It contains recordings from the six basic emotions as defined by Ekman. The model and results of testing are reported in the paper.*

Key words: *Bayesian networks, Automatic Recognition of Facial Expressions, Vector Flow, Cohn Kanade database.*

INTRODUCTION

Facial expressions play an important role in human communication. Humans are able to convey their emotional state by facial expressions. Facial expressions are also used as paralinguistic cues to regulate our conversation. Facial expressions can be used to complete our verbal communication or to reduce the ambiguity in verbal communication. The research of facial expressions has a long history. Darwin pointed to the biological roots of facial expressions and the role they play in the survival of the species including human beings. Ekman defined 6 basic emotional facial expression, happiness, sadness, disgust, anger, surprise and fear and claims that these expressions are universal. Most emotions are blended emotions and according to Ekman, a combination of the basic emotions. Facial expressions are the result of activation/dilatation of one or more of the 43 facial muscles. Unfortunately from an observation point of view the activation of facial muscles is not observable directly. The activation of facial muscles can be observed by facial movements. Ekman defined 43 basic movements in the face called Action Units. He claims that every facial expression can be described in terms of Action Units. He defined a facial action coding scheme called FACS. That system has been used by many researchers, especially those with a psychological background and focus of interest and many facial expressions are coded by the FACS system. One of the well known databases of video recordings of facial expressions is the Cohn Kanade database. The apex of the facial expressions in those recordings is fully described by the activation of the Aus.

The FACS system is based on human observations and the labelling process is manual. For many years, researchers designed systems to classify facial expressions in an automated way. Our system fits in that tradition. The innovative aspect of our system is that we first try to recognize the activation level of involved AUs. Next we can use the knowledge developed in the psychological domain to classify the emotions.

Many methods have been used in the past to classify emotional facial expressions, such as Artificial Neural Networks, Hidden Markov models, (dynamic) Bayesian networks, Support Vector machines etc. [2]. Most systems are focussed on the changing contours of the mouth, eyes and eyebrows. Special fiducially points have been defined on the contours, such as corners of the eyes and the mouth. Most methods try to localise and track those points. In case of happiness the corners of the mouth are going upwards and in case of sadness downwards. We use a different probabilistic approach. Facial expressions in video recordings can be visualised as vector flows. We define special Region Of Interest (ROI) around the underlying facial muscles. In those areas we extract parameters from the gradient field (average length and angle of direction). These parameters are fed into our Bayesian network to compute the probability of the six basic expressions. To train our system, we used the data from the Cohn Kanade database [4].

The outline of this paper is as follows. In the next section we present an overview of related work. Next we define our models. Then we will present the results of our experiments and we will end with the conclusion.

RELATED WORK

In the previous research, the most used facial expression coding tool is FACS (Facial Action Coding System) by Ekman and Friesen [1], which is intended for describing all visually detectable changes on the face produced by facial muscle activity. The FACS system is used by human observers, after extensive training, to recognize and classify subtle facial actions. Facial actions are described with objective and emotion-independent Action Units (AUs), depicted with abbreviations such as AU1, AU4 or AU1+4.

To recognize facial expression by AUs some groups use Gaussian Tree-Augmented Naive Bayes (TAN) to learn the dependencies among different facial motion features in order to classify facial expressions [3]. However, due to TAN's structure limitations, it cannot handle complex relationships between facial features, as well as temporal changes. Zhang and Ji [11] exploit a BN to classify six basic facial expressions with a dynamic fusion strategy. Datcu and Rothkrantz [5] make use of a Bayesian Network architecture which include variables for Action Unit detection and facial expression recognition.

Tian et al. [3] presented an NN-based approach, in which upper face AUs and lower face AUs are considered in NN construction. Gu and Ji [12] use a similar idea for facial event classification such as fatigue. Cohen et al. [13] further propose a classification driven stochastic structure search (SSS) algorithm for learning a BN classifier to recognize facial expression from both labeled and unlabeled data.

Many attempts on the representation of visual information for facial expression have focused on optical flow analysis from facial action [6, 8]. Cohn et al. [9] developed an automated method of facial display analysis based on the detection of fifteen action units and action unit combinations. The features were computed using a hierarchical algorithm for estimating optical flow from sequences of face images. Yacoob and Davis [10] employ optical flow to either model muscle activities or estimate the displacements of feature points. In this paper optical flow is used to estimate the motion of features in defined ROI. Lien et al. explored HMMs for facial AU recognition. Since each AU or AU combination associates with one HMM, the approach is infeasible for covering a great number of potential AU combinations involved in facial expressions. In this paper, we first created a BN model to represent the causal relations between the ROI and facial AUs then extended the BN to a DBN model for modelling the dynamic behaviours of facial expressions in frame sequences. Colmenarez et al. [15] presented a Bayesian probabilistic approach to recognizing the face and facial expression. In this paper, the works focus on recognizing facial AUs instead of facial expression.

Model

We mentioned in the introduction that facial expressions are generated by activation of facial muscles. The visual results of muscle activation are changing contours of the mouth, eye and eyebrow. But we also observe changing texture and position of wrinkles in the face. Many emotional facial expressions are characterised by the shape of the mouth, eye and eyebrow, including some wrinkles areas. In fact many facial muscles are in the area around the mouth, eyes and eyebrow. The activity of facial muscles is limited to specific areas. To study facial movements we define special region of interest (ROIs). The idea is that one muscle activity and corresponding AU is limited to one ROI. But more than one muscle can be active in one ROI. In Fig 1 the ROIs are displayed.

The next step is to model the movement in facial expressions. We consider the movement of pixels in consecutive frames. This movement can be visualized by a vector field that represents the speed and direction of movements. We used the Lucas Kanade algorithm [6] to compute the vector flow from the video recordings. As mentioned in the introduction, we use the Cohn Kanade database. In that database about 70 persons show facial expressions from the 6 basic emotions starting from the neutral position up to the apex, showing the maximal intensity.

We are only interested in facial movements. So the assumption is that the test persons don't move their heads. Unfortunately this assumption is violated many times, so we have to correct for head movements. We used the Active Appearance Model – AAM [14] to localise the face and contours of eyes, eyebrow and mouth. We removed the background and hairs and selected only the mask of the face. This guarantees that our vector flow is concentrated of facial movements and not from the position of the head, hair etc. (see Figure 2).

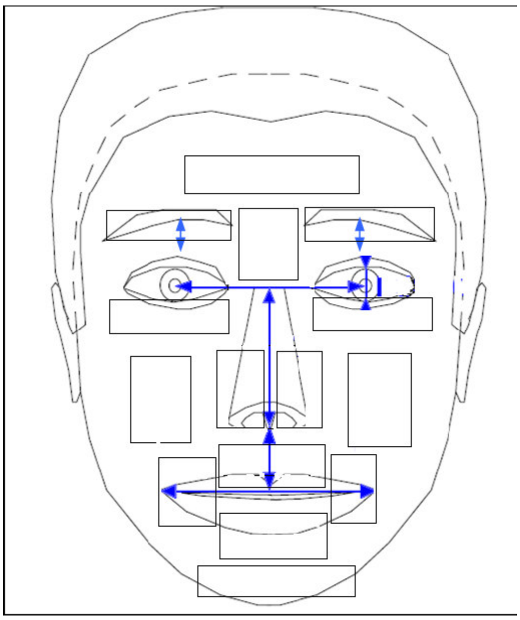

Figure 1. Model of the Area of Interest.

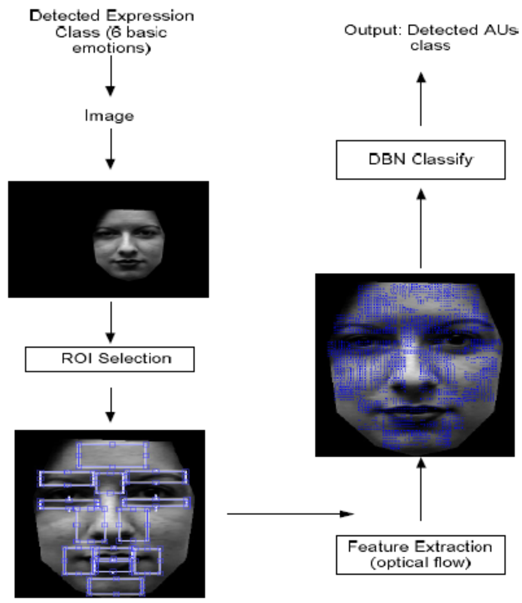

Figure 2. Flowchart of the data analysis.

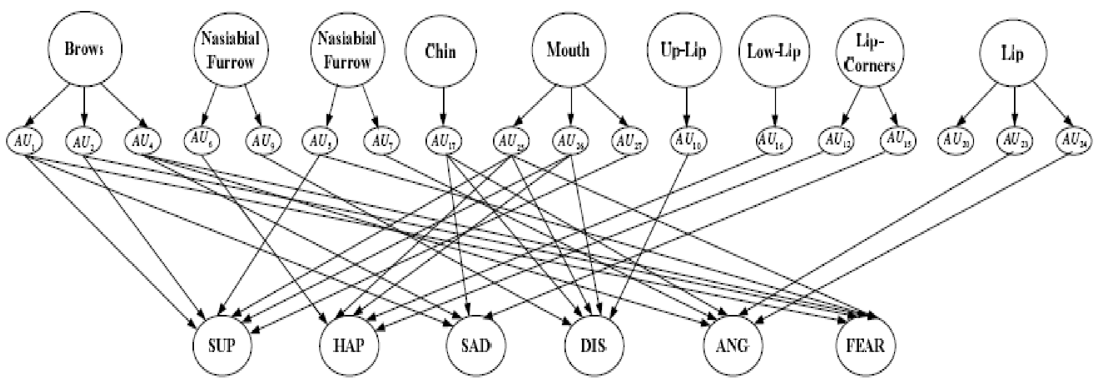

Figure 3. 3-layered Bayesian model.

Experimental Results

The selection of parameters as input of our Bayesian network is a complex process. The Lukas Kanade algorithm has many parameters to choose. Also the choice of ROI is arbitrary. To validate our choices we computed for every of the six basic emotions and ROI the average length and orientation of the vector flow. In Fig 4 we displayed the results. We observe a lot of variation and the average vectors have the expected size and orientation. For example in case of happiness the corners of the mouth are pulled upwards so the vector should point in North direction. In case of sadness, we expect the opposite direction. In case of Fear and Surprise the eyes are fully open and in case of sadness only the upper eyelid is falling down. We observe the expected results in the in the displayed vector field.

A Bayesian network is a probabilistic belief network that is composed of variable nodes (chance, decision, deterministic) and directed arcs (a directed acyclic graph) or temporal arcs. Each node stands for a random variable with discrete values, and each arc has been used to present a dependence (causal) relationship between two nodes. To implement our Bayesian network model we used GeNIe developed by M. Druzdzel at Pittsburg University [7]. To compute the probabilities in the Conditional Probability tables we trained the network on the Cohn Kanade database. The samples showing the apex of specific facial displays have attached annotations based on sequences of Action Units. We conducted an investigation on the relation between the activation of Action Units and facial expressions. In this way, we identified and used the most relevant face samples for each facial expression category. This approach aimed to solve the problem of insufficient data for training the classifiers for automatic detection of combinations of AUs.

The GeNIe tools offers a learning algorithm based on the gradient descend method or Expectation maximisation (EM) method.

As shown in Table 1, the image sequences were assigned to training and testing sets. In the first set S1, the sequences were randomly selected, so the same subjects possibly appeared in both training and testing sets. In the second set S2 no subject was allowed to appear in both training and testing sets, that is to say, the training data and testing data were totally different. The way in assigning training and testing sets is the same as Cohn's in [6].

Table 1. Splitting of dataset in training set and test.

Data Set		number of Sequences	Single AUs																		
			1	2	4	5	6	7	9	10	12	15	16	17	20	23	24	25	25	26	27
S1	Train	689	23	34	24	37	22	21	19	17	24	26	29	28	32	17	19	19	28	27	28
	Test	184	9	10	7	13	10	9	11	10	12	10	10	11	10	9	7	8	9	9	10
S2	Train	505	26	38	25	29	29	26	22	19	22	23	19	18	34	33	28	25	29	29	31
	Test	150	7	8	6	5	6	7	8	6	9	7	8	6	8	9	9	11	11	9	10

The features used by the dynamic BN are extracted by using a tool that first determines the region of interest and secondly prepares the features vectors given the set of face regions.

Finally, we apply the BN model to recognize single AUs and some important AU combinations. The average classification rate for the single AUs are between 80% and 90% and for the AU combinations are above 90%.

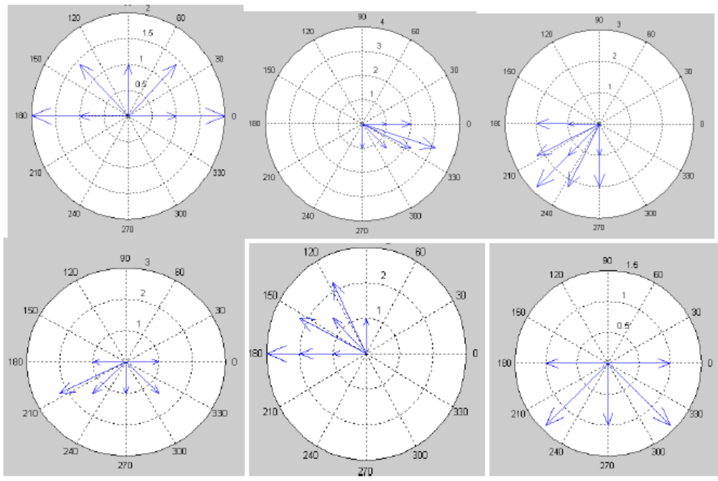

Figure 4. Average of the vector flow parameters for the six basis emotions in ROI.

Table 2. Recognition results of a section of AUs.

Action Units	AU Recognition Rate																	
	1	2	4	5	6	7	9	10	12	15	16	17	20	23	24	25	26	27
	89.4	79.8	88.2	86.7	81.9	79.6	79.9	85.1	79.9	87.4	92.1	88.1	85.1	80.2	81.1	84.1	81.1	81

CONCLUSIONS AND FUTURE WORK

This paper describes a BN model for recognizing the "action units" of a face using video sequence images as input. The features were detected by using an optimal estimation optical flow method coupled with a physical (muscle) model describing the facial structure. These muscle action patterns are then used for analysis, recognition, and synthesis of facial expressions. Our analysis tool consists of three main parts: 1) Region of Interest Selection, 2) Feature Extraction, and 3) Image Classification.

Bayesian Networks proved to be a powerful and flexible methodology for representing and computing probabilistic models in a stochastic process. In past decades' the optical flows has been used to either model muscle activities or estimate the displacements of feature points but in this thesis we find the nine interested regions (ROI) which contains the most complex motion by entropy maximum algorithm. Furthermore, the results were statistically analyzed by compass diagram to find out the major ranges of directions and velocities of vector flows in each ROI. Finally, we apply the BN model to recognize single AUs and some important AU combinations. The average classification rate for the single AUs are between 80% and 90% and for the AU combinations are above 90%.

REFERENCES

[1] Ekman, P., W. V. Friesen, The facial action coding system: A technique for measurement of facial movement. 1978.

[2] Essa, I. A., A. P. Pentland, Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 9, pp. 757-763, 1997.

[3] Tian, Y., T. Kanade, J. F.Cohn, Recognizing Action Units for Facial Expression Analysis. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, 2001.

[4] Kanade, T., J. F. Cohn, Y. Tian, Comprehensive database for facial expression analysis. IEEE Int. Con. on Automatic Face and Gestures Recognition, France, 2000.

[5] Datcu, D., L. J. M. Rothkrantz, Automatic recognition of facial expressions using Bayesian Belief Networks. In IEEE 0-7803-8566-7/04, 2004.

[6] Cohn, J. F., J. J. Lien, A. J. Zlochower, Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression. Department of Electrical Engineering University of Pittsburgh.

[7] Onisko, A., M. J. Druzdzel, H. Wasyluk, Learning Bayesian Network Parameters from Small Data Sets: Application of Noisy-OR Gates. In Working Notes of the Workshop on Bayesian and Causal Networks: From Inference to Data Mining, 12th European Conference on Artificial Intelligence (ECAI-2000), Berlin, August 2000.

[8] Barron, J. L., N. A. Thacker, Tutorial: Computing 2D and 3D Optical Flow. www.tinavision.net, 2005.

[9] Cohn, J. F., A. J. Zlochower, J. Lien, T. Kanade, Automated Face Analysis by Feature Point Tracking Has High Concurrent Validity with Manual FACS Coding. Department of Psychology, University of Pittsburgh and Robotics Institute, Carnegie Mellon University.

[10] Yacoob, Y., L. S. Davis, Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 6, June 1996.

[11] Zhang, Y., Q. Ji, Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 699-714, 2005.

[12] Gu, H., Q. Ji, Facial event classification with task oriented dynamic Bayesian network. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004. vol.2, pp. II-870 - II-875, 2004.

[13] Cohen, I., N. Sebe, F. G. Gozman, M. C. Cirelo, T. S. Huang, Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-595 - I-601, 2003.

[14] Cootes, T. F., G. J. Edwards, C. J. Taylor. Active appearance models. IEEE TPAMI, 23(6), pp. 681–685, 2001.

[15] Colmenarez, A., B. Frey, T. S. Huang, Embedded Face and Facial Expression Recognition. Proceedings of CVPR 99, pp. 592-597, 1999.

ABOUT THE AUTHOR

Prof. drs. dr. Leon J. M. Rothkrantz, SEWACO, Netherlands Defence University, Department of Mediamatica, Delft University of Technology, Phone: +31152787504, E-mail: L.J.M.Rothkrantz@tudelft.nl

Dr. ir. Pascal Wiggers, Department of Mediamatica, Delft University of Technology, Phone: +31152787504, E-mail: P.Wiggers@tudelft.nl