Automatic recognition of facial expressions using Bayesian Belief Networks[*]

D. Datcu
Department of Information Technology and Systems
T.U.Delft, The Netherlands
D.Datcu@ewi.tudelft.nl

L.J.M. Rothkrantz
Department of Information Technology and Systems
T.U.Delft, The Netherlands
L.J.M.Rothkrantz@ewi.tudelft.nl

Abstract - *The current paper addresses the aspects related to the development of an automatic probabilistic recognition system for facial expressions in video streams. The face analysis component integrates an eye tracking mechanism based on Kalman filter. The visual feature detection includes PCA oriented recognition for ranking the activity in certain facial areas. The description of the facial expressions is given according to sets of atomic Action Units (AU) from the Facial Action Coding System (FACS). The base for the expression recognition engine is supported through a BBN model that also handles the time behavior of the visual features.*

Keywords: Facial expression recognition, tracking, pattern recognition.

1 Introduction

The study of human facial expressions is one of the most challenging domains in pattern research community. Each facial expression is generated by non-rigid object deformations and these deformations are person-dependent. The goal of our project was to design and implement a system for automatic recognition of human facial expression in video streams. The results of the project are of a great importance for a broad area of applications that relate to both research and applied topics. As possible approaches on those topics, the following may be presented: automatic surveillance systems, the classification and retrieval of image and video databases, customer-friendly interfaces, smart environment human computer interaction and research in the field of computer assisted human emotion analyses. Some interesting implementations in the field of computed assisted emotion analysis concern experimental and interdisciplinary psychiatry. Automatic recognition of facial expressions is a process primarily based on analysis of permanent and transient features of the face, which can be only assessed with errors of some degree. The expression recognition model is oriented on the specification of Facial Action Coding System (FACS) of Ekman and Friesen [6]. The hard constraints on the scene processing and recording conditions set a limited robustness to the analysis. In order to manage the uncertainties and lack of information, we set a probabilistic oriented framework up. The support for the specific processing involved was given through a multimodal data fusion platform. In the Department of Knowledge Based Systems at T.U.Delft there has been a project based on a long-term research running on the development of a software workbench. It is called Artificial Intelligence aided Digital Processing Toolkit (A.I.D.P.T.) [4] and presents native capabilities for real-time signal and information processing and for fusion of data acquired from hardware equipments. The workbench also includes support for the Kalman filter based mechanism used for tracking the location of the eyes in the scene. The knowledge of the system relied on the data taken from the Cohn-Kanade AU-Coded Facial Expression Database [8]. Some processing was done so as to extract the useful information. More than that, since the original database contained only one image having the AU code set for each display, additional coding had to be done. The Bayesian network is used to encode the dependencies among the variables. The temporal dependencies were extracted to make the system be able to properly select the right emotional expression. In this way, the system is able to overcome the performance of the previous approaches that dealt only with prototypic facial expression [10]. The causal relationships track the changes occurred in each facial feature and store the information regarding the variability of the data.

2 Related work

The typical problems of expression recognition have been tackled many times through distinct methods in the past. In [12] the authors proposed a combination of a Bayesian probabilistic model and Gabor filter. [3] introduced a Tree-Augmented-Naive Bayes (TAN) classifier for learning the feature dependencies. A common approach was based on neural networks. [7] used a neural network approach for developing an online Facial Expression Dictionary as a first step in the creation of an online Nonverbal Dictionary. [2] used a subset of Gabor filters selected with Adaboost and trained the Support Vector Machines on the outputs.

[*] **0-7803-8566-7/04/$20.00 © 2004 IEEE.**

3 Eye tracking

The architecture of the facial expression recognition system integrates two major components. In the case of the analysis applied on video streams, a first module is set to determine the position of the person eyes. Given the position of the eyes, the next step is to recover the position of the other visual features as the presence of some wrinkles, furrows and the position of the mouth and eyebrows. The information related to the position of the eyes is used to constrain the mathematical model for the point detection. The second module receives the coordinates of the visual features and uses them to apply recognition of facial expressions according to the given emotional classes. The detection of the eyes in the image sequence is accomplished by using a tracking mechanism based on Kalman filter [1]. The eye-tracking module includes some routines for detecting the position of the edge between the pupil and the iris. The process is based on the characteristic of the dark-bright pupil effect in infrared condition (see Figure 1).

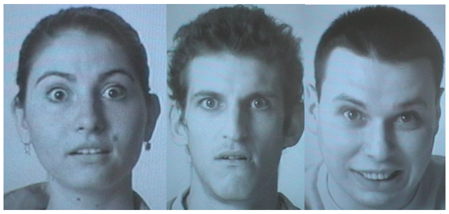


Figure 1. The dark-bright pupil effect in infrared

However, the eye position locator may not perform well in some contexts as poor illuminated scene or the rotation of the head. The same might happen when the person wears glasses or has the eyes closed. The inconvenience is managed by computing the most probable eye position with Kalman filter. The estimation for the current frame takes into account the information related to the motion of the eyes in the previous frames. The Kalman filter relies on the decomposition of the pursuit eye motion into a deterministic component and a random component. The random component models the estimation error in the time sequence and further corrects the position of the eye. It has a random amplitude, occurrence and duration. The deterministic component concerns the motion parameters related to the position, velocity and acceleration of the eyes in the sequence. The acceleration of the motion is modeled as a Gauss-Markov process. The autocorrelation function is as presented in formula (1):

$$R(\tau) = \sigma^2 e^{-\beta|\tau|} \qquad (1)$$

The equations of the eye movement are defined according to the formula (2). In the model we use, the state vector contains an additional state variable according to the Gauss-Markov process. $u(t)$ is a unity Gaussian white noise.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t) \qquad (2)$$

$$z = \begin{bmatrix} \sqrt{2\sigma^2\beta} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

The discrete form of the model for tracking the eyes in the sequence is given in formula (3). $\phi = e^{F\Delta t}$, w are the process Gaussian white noise and v is the measurement Gaussian white noise.

$$\begin{aligned} x_k &= \phi_k + w_k \\ z_k &= H_k + v_k \end{aligned} \qquad (3)$$

The Kalman filter method used for tracking the eyes presents a high efficiency by reducing the error of the coordinate estimation task. In addition to that, the process does not require a high processor load and a real time implementation was possible.

4 Face representational model

The facial expression recognition system handles the input video stream and performs analysis on the existent frontal face. In addition to the set of degree values related to the detected expressions, the system can also output a graphical face model. The result may be seen as a feedback of the system to the given facial expression of the person whose face is analyzed and it may be different of that. One direct application of the chosen architecture may be in design of systems that perceive and interact with humans by using natural communication channels. In our approach the result is directly associated to the expression of the input face (see Figure 2). Given the parameters from the expression recognition module, the system computes the shape of different visual features and generates a 2D graphical face model.

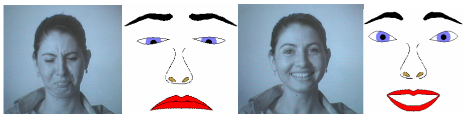


Figure 2. Response of the expression recognition

The geometrical shape of each visual feature follows certain rules that aim to set the outlook to convey the appropriate emotional meaning. Each feature is reconstructed using circles and simple polynomial functions as lines, parabola parts and cubic functions. A five-pixel window is used to smooth peaks so as to provide shapes with a more realistic appearance. The eye upper and lower lid was approximated with the same cubic function. The eyebrow's thickness above and below the middle line was calculated from three segments as a parabola, a straight line and a quarter of a circle as the inner corner. A thickness function was added and subtracted to and from the middle line of the eyebrow. The shape of the mouth varies strongly as emotion changes from sadness to happiness or disgust. The manipulation of the face for setting a certain expression implies to mix different emotions. Each emotion has a percentage value by which they contribute to the face general expression. The new control set values for the visual features are computed by the difference of each emotion control set and the neutral face control set, and make a linear combination of the resulting six vectors.

5 Visual feature acquisition

The objective of the first processing component of the system is to recover the position of some key points on the face surface. The process starts with the stage of eye coordinate detection. Certified FACS coders coded the image data. Starting from the image database, we processed each image and obtained the set of 30 points according to Kobayashi & Hara model [9]. The analysis was semi-automatic. A new transformation was involved then to get the key points as described in figure 3. The coordinates of the last set of points were used for computing the values of the parameters presented in table 2. The preprocessing tasks implied some additional requirements to be satisfied. First, for each image a new coordinate system was set. The origin of the new coordinate system was set to the nose top of the individual. The value of a new parameter called *base* was computed to measured the distance between the eyes of the person in the image. The next processing was the rotation of all the points in the image with respect to the center of the new coordinate system. The result was the frontal face with correction to the facial inclination. The final step of preprocessing was related to scale all the distances so as to be invariant to the size of the image. Eventually a set of 15 values for each of the image was obtained as the result of preprocessing stage. The parameters were computed by taking both the variance observed in the frame at the time of analysis and the temporal variance. Each of the last three parameters was quantified so as to express a linear behavior with respect to the range of facial expressions analyzed. The technique used was Principal Component Analysis oriented pattern recognition for each of the three facial areas. The technique was first applied by Turk and Pentland for face imaging [11]. The PCA processing is run separately for each area and three sets of eigenvectors are available as part of the knowledge of the system. Moreover, the labeled patterns associated with each area are stored (see Figure 4). The computation of the eigenvectors was done offline as a preliminary step of the process. For each input image, the first processing stage extracts the image data according to the three areas. Each data image is projected through the eigenvectors and the pattern with the minimum error is searched.

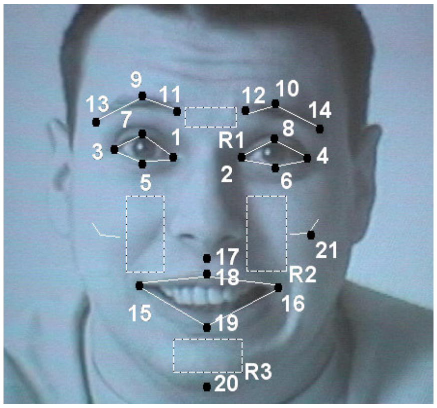


Figure 3. The model facial key points and areas

The label of the extracted pattern is then fed to the quantification function for obtaining the characteristic output value of each image area. Each value is further set as evidence in the probabilistic BBN.

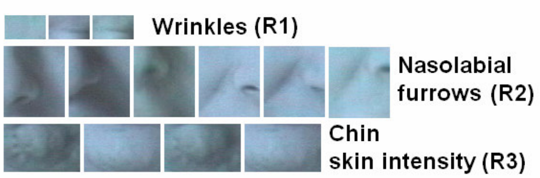


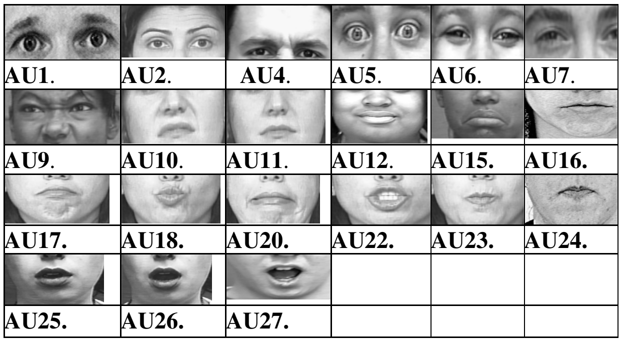
Figure 4. Examples of patterns used in PCA recognition

6 Data preparation

The Bayesian Belief Network encodes the knowledge of the existent phenomena that triggers changes in the aspect of the face. The model does include several layers for the detection of distinct aspects of the transformation. The lowest level is that of primary parameter layer. It contains a set of parameters that keeps track of the changes concerning the facial key points. Those parameters may be classified as static and dynamic. The static parameters handle the local geometry of the current frame. The dynamic parameters encode the behavior of the key points in the transition from one frame to another. By combining the two sorts of information, the system gets a high efficiency of expression recognition. An alternative is that the base used for computing the variation of the dynamic parameters is determined as a previous tendency over a limited past time. Each parameter on the lowest layer of the BBN has a given number of states. The purpose of the states is to map any continuous value of the parameter to a discrete class. The number of states has a direct influence on the efficiency of recognition. The number of states for the low-level parameters does not influence the time required for obtaining the final results. It is still possible to have a real time implementation even when the number of states is high.

The only additional time is that of processing done for computing the conditioned probability tables for each BBN parameter, but the task is run off-line. According to the method used, each facial expression is described as a combination of existent Action Units (AU).

Table 1. The used set of Action Units



One AU represents a specific facial display. Among 44 AUs contained in FACS, 12 describe contractions of specific facial muscles in the upper part of the face and 18 in the lower part. The table 1 presents the set of AUs that is managed by the current recognition system.

Table 2. The set of visual feature parameters

	Static	Dynamic	Visual Feature
P_1	$dy(2,12)$	$\Delta P_1 / P_1^s(0)$	*Eyebrow*
P_2	$dy(2,14)$	$\Delta P_2 / P_2^s(0)$	*Eyebrow*
P_3	$m(\angle 12,10,14)$	$\Delta P_3 / P_3^s(0)$	*Eyebrow*
P_4	$dy(6,8)$	$\Delta P_4 / P_4^s(0)$	*Eye*
P_5	$dy(2,6)$	$\Delta P_5 / P_5^s(0)$	*Eye*
P_6	$m(\angle 2,6,4)$	$\Delta P_6 / P_6^s(0)$	*Eye*
P_7	$dy(17,18)$	$\Delta P_7 / P_7^s(0)$	*Mouth*
P_8	$dy(16,17)$	$\Delta P_8 / P_8^s(0)$	*Mouth*
P_9	$dy(17,19)$	$\Delta P_9 / P_9^s(0)$	*Mouth*
P_{10}	$dx(15,16)$	$\Delta P_{10} / P_{10}^s(0)$	*Mouth*
P_{11}	$dy(18,19)$	$\Delta P_{11} / P_{11}^s(0)$	*Mouth*
P_{12}	$dy(2,21)$	$\Delta P_{12} / P_{12}^s(0)$	*Cheek*
P_{13}	$f(R_1)$	$\Delta f(R_1) / f_0^s(R_1)$	*Forehead*
P_{14}	$f(R_2)$	$\Delta f(R_2) / f_0^s(R_2)$	*Nasolabial*
P_{15}	$f(R_3)$	$\Delta f(R_3) / f_0^s(R_3)$	*Chin*

An important characteristic of the AUs is that they may act differently in given combinations. According to the behavioral side of each AU, there are additive and non-additive combinations. In that way, the result of one non-additive combination may be related to a facial expression that is not expressed by the constituent AUs taken separately. In the case of the current project, the AU sets related to each expression are split into two classes that specify the importance of the emotional load of each AU in the class. By means of that, there are primary and secondary AUs. The AUs being part of the same class are additive. The system performs recognition of one expression as computing the probability associated with the detection of one or more AUs from both classes. The probability of one expression increases, as the probabilities of detected primary AUs get higher. In the same way, the presence of some AUs from a secondary class results in solving the uncertainty problem in the case of the dependent expression but at a lower level.

Table 3. The dependency between AUs and intermediate parameters

AU	Action	Encoding
AU1	Inner corner of the eyebrow raised	P1
AU2	Outer corner of the eyebrow raised	P2
AU4	Eyebrows lowered and horizontal	P1,P3
AU5	Eyes widened	P4
AU6	Cheeks Raised and eyes narrowed	P12,P4
AU7	Lower eyelid raised and horizontal	P5,P6
AU9	Upper lip raised, activity in the area between the eyebrows and the superior part of the nasolabial	P7,P13,P14
AU10	Upper lip raised, activity in the nasolabial area	P7,P14
AU11	Activity in the nasolabial area	P14
AU12	Lip corners raised and laterally	P8,P10
AU15	Lip corner lowered and inward	P8,P10
AU16	Lower lip lowered, mouth stretched laterally	P9,P10
AU17	Activity in the chin area	P15
AU18	Lips pursed	P10,P11
AU20	Lips stretched horizontally	P10
AU22	Lips funneled	P10,P11
AU23	Lips tightened	P10,P11
AU24	Lips pressed together	P10,P11
AU25	Lips parted	P10,P11
AU26	Jaw lowered	P9
AU27	Mouth stretched open	P10,P11

The conditioned probability tables for each node of the Bayesian Belief Network were filled in by computing statistics over the database. The Cohn-Kanade AU-Coded Facial Expression Database contains approximately 2000 image sequences from 200 subjects ranged in age from 18 to 30 years. Sixty-five percent were female, 15 percent were African-American and three percent were Asian or Latino. All the images analyzed were frontal face pictures. The original database contained sequences of the subjects performing 23 facial displays including single action units and combinations. Six of the displays were based on prototypic emotions (joy, surprise, anger, fear, disgust and sadness).

7 Inference with BBN

The expression recognition is done computing the anterior probabilities for the parameters in the BBN (see Figure 5). The procedure starts by setting the probabilities of the parameters on the lowest level according to the values computed at the preprocessing stage. In the case of each parameter, evidence is given for both static and dynamic parameters. Moreover, the evidence is set also for the parameter related to the probability of the anterior facial expression. It contains 6 states, one for each major class of expressions. The aim of the presence of the anterior expression node and that associated with the dynamic component of one given low-level parameter, is to augment the inference process with temporal constrains. The structure of the network integrates parametric layers having different functional tasks. The goal of the layer containing the first AU set and that of the low-level parameters is to detect the presence of some AUs in the current frame. The relation between the set of the low-level parameters and the action units is as it is detailed in table 4. The dependency of the parameters on AUs was determined on the criteria of influence observed on the initial database. The presence of one AU at this stage does not imply the existence of one facial expression or another. Instead, the goal of the next layer containing the AU nodes and associated dependencies is to determine the probability that one AU presents influence on a given kind of emotion. The final parametric layer consists of nodes for every emotional class. More than that, there is also one node for the current expression and another one for that previously detected. The top node in the network is that of current expression. It has two states according to the presence and absence of any expression and stands for the final result of analysis. The absence of any expression is seen as a neutral display of the person’s face on the current frame. While performing recognition, the BBN probabilities are updated in a bottom-up manner. As soon as the inference is finished and expressions are detected, the system reads the existence probabilities of all the dependent expression nodes. The most probable expression is that given by the larger value over the expression probability set.

Table 4. The emotion projections of each AU combination

	Primary AUs						Secondary Aus			
Surprise	1+2	5	25	26	27					
Fear	1+2	4	5	7	20	25	6	10	11	
Happy	6+12	12					7	16	25	26
Sadness	15+17	4+17	1				7	24	20+25	
Disgust	4	7+9	17	6+7			23	24	25	
Anger	4	4+7	23	24			17			

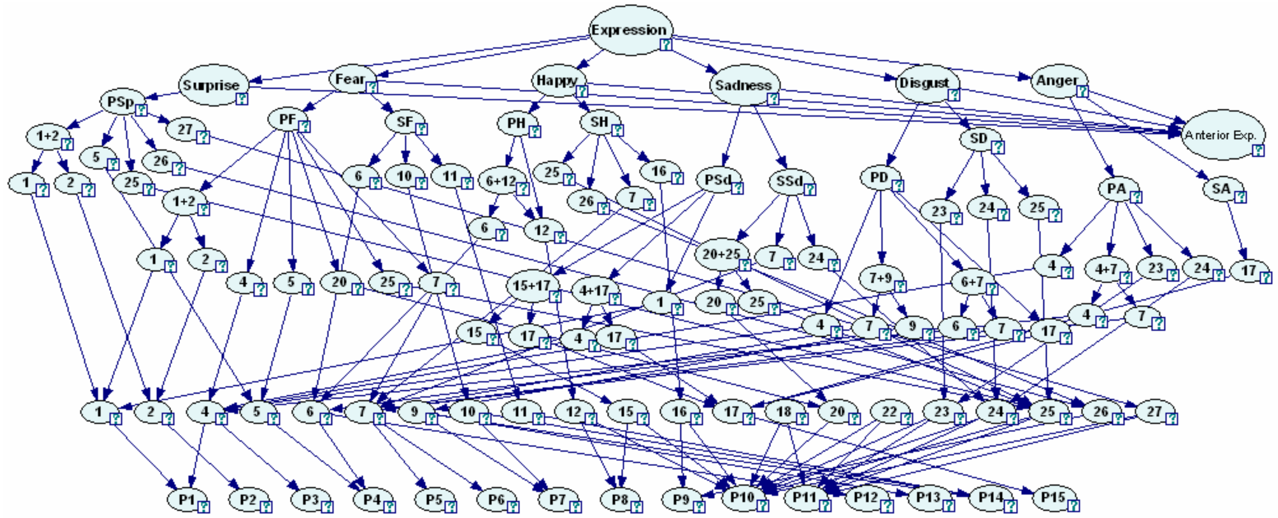


Figure 5. BBN used for facial expression recognition

8 Results

The implementation of the model was made using C/C++ programming language. The system consists in a set of applications that run different tasks that range from pixel/image oriented processing to statistics building and inference by updating the probabilities in the BBN model. The support for BBN was based on S.M.I.L.E. (Structural Modeling, Inference, and Learning Engine), a platform independent library of C++ classes for reasoning in probabilistic models [5]. S.M.I.L.E. is freely available to the community and has been developed at the Decision Systems Laboratory, University of Pittsburgh. The library was included in the AIDPT framework. The implemented probabilistic model is able to perform recognition on six emotional classes and the neutral state. By adding new parameters on the facial expression layer, the expression number on recognition can be easily increased. Accordingly, new AU dependencies have to be specified for each of the emotional class added. In figure 7 there is an example of an input video sequence. The recognition result is given in the graphic containing the information related to the probability of the dominant facial expression (see Figure 6).

9 Conclusion

In the current paper we've described the development steps of an automatic system for facial expression recognition in video sequences. The inference mechanism was based on a probabilistic framework. We used the Cohn-Kanade AU-Coded Facial Expression Database for building the system knowledge. It contains a large sample of varying age, sex and ethnic background and so the robustness to the individual changes in facial features and behavior is high. The BBN model takes care of the variation and degree of uncertainty and gives us an improvement in the quality of recognition. As off now, the results are very promising and show that the new approach presents high efficiency. An important contribution is related to the tracking of the temporal behavior of the analyzed parameters and the temporal expression constrains.

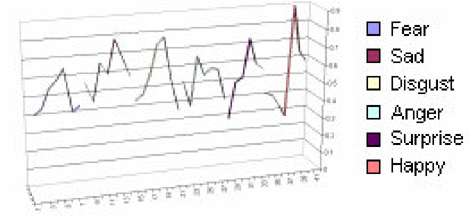


Figure 6. Dominant Emotional expression in sequence

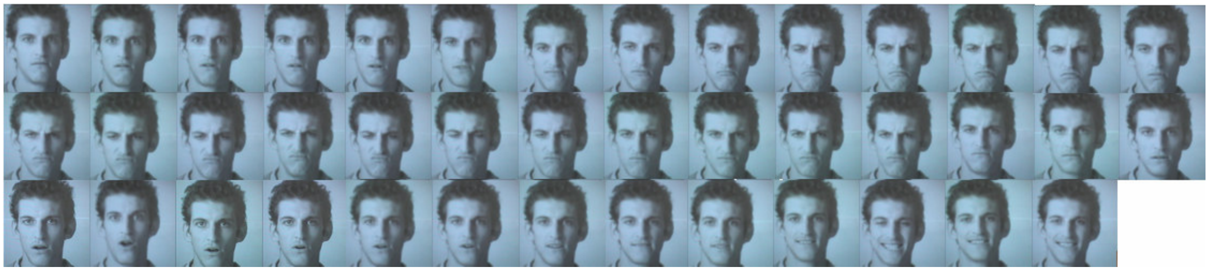


Figure 7. Example of facial expression recognition applied on video streams

References

[1] W. A-Almageed, M. S. Fadali, G. Bebis 'A non-intrusive Kalman Filter-Based Tracker for Pursuit Eye Movement' Proceedings of the 2002 American Control Conference Alaska, 2002

[2] M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan 'Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction' IEEE Workshop on Face Processing in Video, Washington 2004

[3] I. Cohen, N. Sebe, A. Garg, M. S.Lew, T. S. Huang 'Facial expression recognition from video sequences' Computer Vision and Image Understanding, Volume 91, pp 160 - 187 ISSN: 1077-3142 2003

[4] D. Datcu, L. J. M. Rothkrantz 'A multimodal workbench for automatic surveillance' Euromedia Int'l Conference 2004

[5] M. J. Druzdzel 'GeNIe: A development environment for graphical decision-analytic models'. In Proceedings of the 1999 Annual Symposium of the American Medical Informatics Association (AMIA-1999), page 1206, Washington, D.C., November 6-10, 1999

[6] P. Ekman, W. V. Friesen 'Facial Action Coding System: Investigator's Guide' Consulting Psychologists Press, 1978

[7] E. J. de Jongh, L .J. M. Rothkrantz 'FED – an online Facial Expression Dictionary' Euromedia Int'l Conference 2004

[8] T. Kanade, J. Cohn, Y. Tian 'Comprehensive database for facial expression analysis' Proc. IEEE Int'l Conf. Face and Gesture Recognition, pp. 46-53, 2000

[9] H. Kobayashi and F. Hara. 'Recognition of Mixed Facial Expressions by Neural Network' IEEE International workshop on Robot and Human Communication, 381-386, 1972

[10] M. Pantic, L. J. M. Rothkrantz 'Toward an Affect-Sensitive Multimodal Human-Computer Interaction' IEEE proceedings vol. 91, no. 9, pp. 1370-1390, 2003

[11] M. Turk, A. Pentland 'Face recognition using eigenfaces, Proc. CVPR, pp. 586-591 (1991)

[12] X. Wang, X. Tang 'Bayesian Face Recognition Using Gabor Features' Proceedings of the 2003 ACM SIGMM Berkley, California 2003