Facial Expression Recognition for Human-Robot Interaction – A Prototype

Matthias Wimmer[1], Bruce A. MacDonald[2],
Dinuka Jayamuni[2], and Arpit Yadav[2]

[1] Department of Informatics, Technische Universitat München, Germany
[2] Electrical and Computer Engineering, University of Auckland, New Zealand

Abstract. To be effective in the human world robots must respond to human emotional states. This paper focuses on the recognition of the six universal human facial expressions. In the last decade there has been successful research on facial expression recognition (FER) in controlled conditions suitable for human–computer interaction [1,2,3,4,5,6,7,8]. However the human–robot scenario presents additional challenges including a lack of control over lighting conditions and over the relative poses and separation of the robot and human, the inherent mobility of robots, and stricter real time computational requirements dictated by the need for robots to respond in a timely fashion.

Our approach imposes lower computational requirements by specifically adapting model-based techniques to the FER scenario. It contains adaptive skin color extraction, localization of the entire face and facial components, and specifically learned objective functions for fitting a deformable face model. Experimental evaluation reports a recognition rate of 70% on the Cohn–Kanade facial expression database, and 67% in a robot scenario, which compare well to other FER systems.

1 Introduction

The desire to interpret human gestures and facial expressions is making interaction with robots more human-like. This paper describes our model–based approach for automatically recognizing facial expressions, and its application to human-robot interaction.

Knowing the human user's intentions and feelings enables a robot to respond more appropriately during tasks where humans and robots must work together [9,10,11], which they must do increasingly as the use of service robots continues to grow. In robot-assisted learning, a robot acts as the teacher by explaining the content of the lesson and questioning the user afterwards. Being aware of human emotion, the quality and success of these lessons will rise because the robot will be able to progress from lesson to lesson just when the human is ready [12]. Studies of human-robot interaction will be improved by automated emotion interpretation. Currently the humans' feelings about interactions must be interpreted using questionnaires, self-reports, and manual analysis of recorded video. Robots will need to detect deceit in humans in security applications, and

G. Sommer and R. Klette (Eds.): RobVis 2008, LNCS 4931, pp. 139–152, 2008.

facial expressions may help [13]. Furthermore, natural human-robot interaction requires detecting whether or not a person is telling the truth. Micro expressions within the face express these subtle differences. Specifically trained computer vision applications would be able to make this distinction.

Today's techniques for detecting human emotion often approach this challenge by integrating dedicated hardware to make more direct measurements of the human [14,15,16]. Directly connected sensors measure blood pressure, perspiration, brain waves, heart rate, skin temperature, electrodermal activity, etc. in order to estimate the human's emotional state. In practical human-robot interactions, these sensors would need to be portable, wearable and wireless. However, humans interpret emotion mainly from video and audio information and it would be desirable if robots could obtain this information from their general purpose sensing systems, in the visual and audio domains. Furthermore, these sensors do not interfere with the human being by requiring direct connections to the human body. Our approach interprets facial expressions from video information.

Section 2 explains the state of the art in facial expression recognition (FER) covering both psychological theory and concrete approaches. It also elaborates on the specific challenges in robot scenarios. Section 3 describes our model-based approach which derives facial expressions from both structural and temporal features of the face. Section 4 presents results on a standard test set and in a practical scenario with a mobile robot.

2 Facial Expression Recognition: State of the Art

Ekman and Friesen [18] find six universal facial expressions that are expressed and interpreted in the same way by humans of any origin all over the world. They do not depend on the cultural background or the country of origin. Figure 1 shows one example of each facial expression. The Facial Action Coding System (FACS) precisely describes the muscle activity within a human face [19]. So-called Action Units (AUs) denote the motion of particular facial parts and state the involved facial muscles. Combinations of AUs assemble facial expressions. Extended systems such as the *Emotional FACS* [20] specify the relation between facial expressions and emotions.

The Cohn–Kanade Facial Expression Database (CKDB) contains a number of 488 short image sequences of 97 different persons showing the six universal facial expressions [17]. It provides researchers with a large dataset for experiment and benchmarking. Each sequence shows a neutral face at the beginning and then

happiness anger disgust sadness fear surprise

Fig. 1. The six universal facial expressions as they occur in [17]

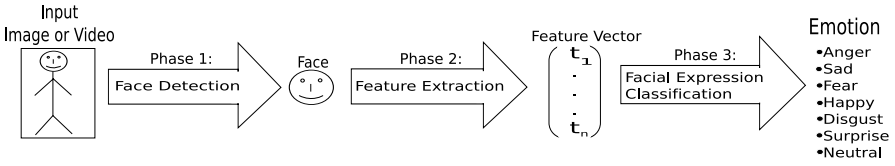

Fig. 2. Procedure for recognizing facial expressions according to Pantic et al. [3]

develops into the peak expression. Furthermore, a set of AUs has been manually specified by licensed FACS experts for each sequence. Note that this database does not contain natural facial expressions, but volunteers were asked to act them. Furthermore, the image sequences are taken in a laboratory environment with predefined illumination conditions, solid background and frontal face views. Algorithms that perform well with these image sequences are not necessarily appropriate for real-world scenes.

2.1 The Three-Phase Procedure

The computational task of FER is usually subdivided into three subordinate challenges shown in Figure 2: face detection, feature extraction, and facial expression classification [3]. Others add pre– and post–processing steps [4].

In Phase 1, the human face and the facial components must be accurately located within the image. On the one hand, sometimes this is achieved automatically, as in [1,21,22]. Most automatic approaches assume the presence of a frontal face view. On the other hand, some researchers prefer to specify this information by hand, and focus on the interpretation task itself, as in [23,24,25,26].

In Phase 2, the features relevant to facial expressions are extracted from the image. Michel et al. [21] extract the location of 22 feature points within the face and determine their motion between a neutral frame and a frame representative for a facial expression. These feature points are mostly located around the eyes and mouth. Littlewort *et al.* [27] use a bank of 40 Gabor wavelet filters at different scales and orientations to extract features directly from the image. They perform convolution and obtain a vector of magnitudes of complex valued responses.

In Phase 3, the facial expression is derived from the extracted features. Most often, a classifier is learned from a comprehensive training set of annotated examples. Some approaches first compute the visible AUs and then infer the facial expression by rules stated by Ekman and Friesen [28]. Michel *et al.* [21] train a Support Vector Machine (SVM) that determines the visible facial expression within the video sequences of the CKDB by comparing the first frame with the neutral expression to the last frame with the peak expression. Schweiger and Bayerl [25] compute the optical flow within 6 predefined face regions to extract the facial features. Their classification uses supervised neural networks.

2.2 Challenges of Facial Expression Recognition

FER is a particularly challenging problem. Like speech, facial expressions are easier to generate than to recognize. The three phases represent different challenges. While Phase 1 and Phase 2 are confronted with image interpretation challenges, Phase 3 faces common Machine Learning problems. The challenge of one phase increases if the result of the previous one is not accurate enough.

The first two phases grasp image components of different semantic levels. While human faces have many similarities in general shape and layout of features, in fact all faces are different, and vary in shape, color, texture, the exact location of key features, and facial hair. Faces are often partially occluded by spectacles, facial hair, or hats. Lighting conditions are a significant problem as well. Usually, there are multiple sources of light, and hence multiple shadows on a face.

The third phase faces typical Machine Learning challenges. The features must be representative for the target value, i.e. the facial expression. The inference method must be capable to derive the facial expression from the facial features provided. The learning algorithm has to find appropriate inference rules. Being the last phase, it depends most on accurate inputs from the preceding phases.

Cohen *et al.* [24] use a three-dimensional wireframe model consisting of 16 different surface patches representing different parts of the face. Since these patches consist of Bézier volumes, the model's deformation is expressed by the Bézier volume parameters. These parameters represent the basis for determining the visible facial expression. Cohen *et al.* integrate two variants of Bayesian Network classifiers, a Naive Bayes classifier and a Tree-Augmented-Naive Bayes classifier. Whereas the first one treats the motion vectors to be independent, the latter classifier assumes dependencies between them, which facilitate the interpretation task. Further improvements are achieved by integrating temporal information. It is computed from measuring different muscle activity within the face, which is represented by Hidden Markov Models (HMMs).

Bartlett *et al.* [29] compare recognition engines on CKDB and find that a subset of Gabor filters using AdaBoost followed by training on Support Vector Machines gives the best results, with 93% correct for novel subjects in a 7–way choice of expressions in real time. Note that this approach is tuned to the specifics of the images within CKDB. In contrast, approaches for a robot scenario must be robust to the variety of real-world images.

2.3 Additional Challenges in a Robot Scenario

In contrast to human–machine interactions with say desktop computers, during interactions with robots the position and orientation of the human is less constrained, which makes facial expression recognition more difficult. The human may be further from the robot's camera, e.g he or she could be on the other side of a large room. The human may not be directly facing the robot. Furthermore, he or she may be moving, and looking in different directions, in order to take part in the robot's task. As a result it is difficult to ensure a good, well illuminated view of the human's face for FER.

To make matters worse, robots are mobile. There is no hope of placing the human within a controlled space, such as a controlled lighting situation, because the human must follow the robot in order to continue the interaction. Most image subtraction approaches cannot cope, because image subtraction is intended to separate facial activity from the entire surrounding that is expected to be static. In contrast, our model-based approach copes with this challenge, because of the additional level of abstraction (the model).

A technical challenge is to meet the real time computational constraints for human-robot interaction. The robot must respond to human emotions within a fraction of a second, to have the potential to improve the interaction.

Many researchers have studied FER for robots. Some model–based methods cannot provide the real time performance needed for human–robot interaction, because model fitting is computationally expensive and because the required high resolution images impose an additional computational burden [30]. Kim *et al* [30] use a set of rectangular features and train using AdaBoost. Yoshitomi *et al* fuse speech data, face images and thermal face images to help a robot recognize emotional states [31]. An HMM is used for speech recognition. Otsuka and Ohya [32] use HMMs to model facial expressions, for recognition by robots. Zhou *et al* use an embedded HMM and AdaBoost for real-time FER by a robot [33].

3 Model-Based Interpretation of Facial Expressions

Our approach makes use of model-based techniques, which exploit *a priori* knowledge about objects, such as their shape or texture. Reducing the large amount of image data to a small set of model parameters facilitates and accelerates the subsequent facial expression interpretation, which mitigates the computational

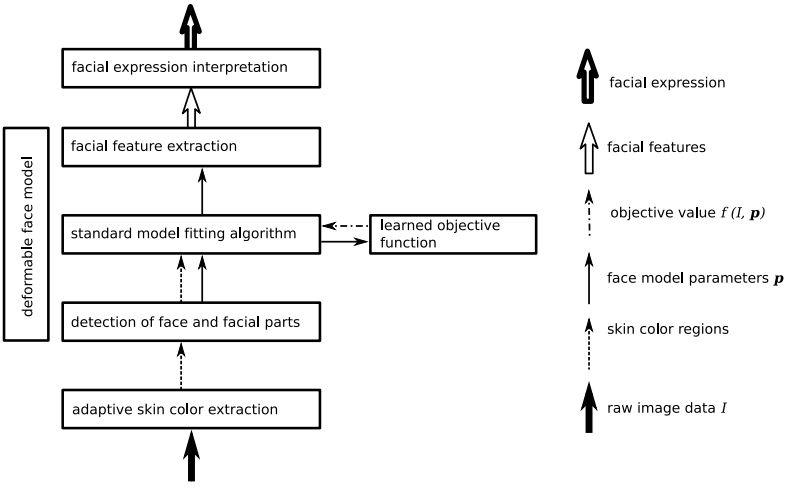

Fig. 3. Model-based image interpretation splits the challenge of image interpretation into computationally independent modules

problems often associated with model–based approaches to FER. According to the usual configuration [34], our model-based approach consist of seven components, which are illustrated in Figure 3.

This approach fits well into the three-phase procedure of Pantic *et al.* [3], where skin color extraction represents a pre-processing step, mentioned by Chibelushi *et al.* [4]. Phase 1 is contains by the core of model-based techniques: the model, localization, the fitting algorithm, and the objective function. Phase 2 consists of the facial feature extraction, and Phase 3 is the final step of facial expression classification.

3.1 The Deformable Face Model

The model contains a parameter vector $\boldsymbol{p}$ that represents its possible configurations, such as position, orientation, scaling, and deformation. Models are mapped onto the surface of an image via a set of feature points, a contour, a textured region, etc. Our approach makes use of a statistics-based deformable model, as introduced by Cootes *et al.* [35]. Its parameters $\boldsymbol{p} = (t_x, t_y, s, \theta, \boldsymbol{b})^T$ comprise the affine transformation (translation, scaling factor, and rotation) and a vector of deformation parameters $\boldsymbol{b} = (b_1, \ldots, b_B)^T$. The latter component describes the configuration of the face, such as the opening of the mouth, the direction of the gaze, and the raising of the eye brows, compare to Figure 4. In this work, $B = 17$ to cover all necessary modes of variation.

3.2 Skin Color Extraction

Skin color, as opposed to pixel values, represents highly reliable information about the location of the entire face and the facial components and their boundaries. Unfortunately, skin color occupies a cluster in color space that varies with

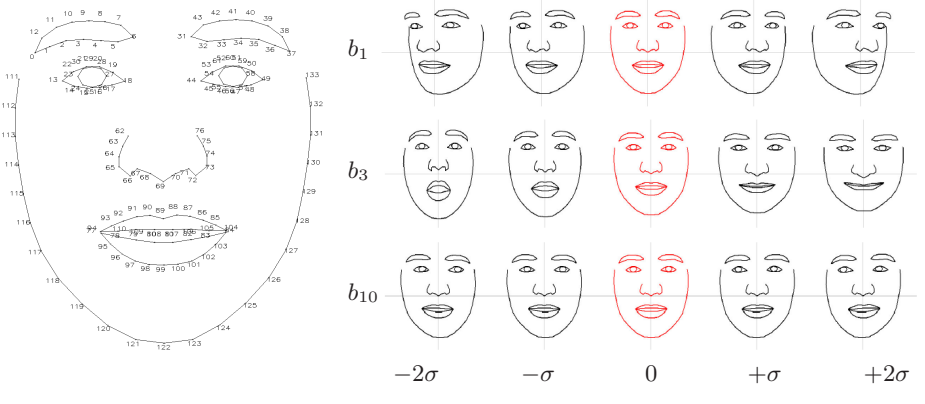

Fig. 4. Our deformable model of a human face consists of 134 contour points that represent the major facial components. The deformation by a change of just one parameter is shown in each row, ranging from -2σ and 2σ, as in Cootes *et al.* [36]. b_1 rotates the head. b_3 opens the mouth. b_{10} changes the direction of the gaze.

the scenery, the person, and the camera type and settings. Therefore, we conduct a two-step approach as described in our previous publication [37]. We first detect the specifics of the image, describe the skin color as it is visible in the given image. In a second step, the feature values of the pixels are adjusted by these image specifics. This results a set of simple and quick pixel features that are highly descriptive for skin color regions of the given image.

This approach facilitates distinguishing skin color from very similar colors such as those of the lips and eyebrows. The borders of the skin regions are clearly extracted, and since most of these borders correspond to the contour lines of our face model, see Figure 4, this supports model fitting in the subsequent steps.

3.3 Localization of the Face and the Facial Components

The localization algorithm computes an initial estimate of the model parameters. The subsequent fitting algorithm is intended to refine these values further. Our system integrates the approach of Viola and Jones [38], which detects a rectangular bounding box around the frontal face view. From this information, we derive the affine transformation parameters of our face model.

Additionally, rough estimation of the deformation parameters $\boldsymbol{b}$ improves accuracy. A second iteration of the Viola and Jones object locator is used on the previously determined rectangular image region around the face. We specifically learn further algorithms to localize the facial components, such as eyes and mouth.[1] In the case of the eyes, the positive training data contains images of eyes, whereas the negative training data consists of image patches that are taken from the vicinity of the eyes, such as the cheek, the nose, or the brows. Note that the learned eye locator is not able to accurately find the eyes within a complex image, because images usually contain a lot of information that was not part of our specific training data. However, the eye locator is highly appropriate to determine the location of the eyes given a pure face image or a facial region within a complex image.

3.4 The Objective Function

The objective function $f(I, \boldsymbol{p})$ yields a comparable value that specifies how accurately a parameterized model $\boldsymbol{p}$ describes an image I. It is also known as the likelihood, similarity, energy, cost, goodness, or quality function. Without loss of generality, we consider lower values to denote a better model fit. The fitting algorithm, as described in Section 3.5, searches for the model that describes the image best by determining the global minimum of the objective function.

Traditionally, the calculation rules of objective functions are manually specified by first selecting a small number of image features, such as edges or corners, and then combining them by mathematical operators, see Figure 5. Afterwards, the appropriateness of the function is subjectively investigated by inspecting its

[1] Locators for facial components, which are part of our system, can be downloaded at: `http://www9.in.tum.de/people/wimmerm/se/project.eyefinder`

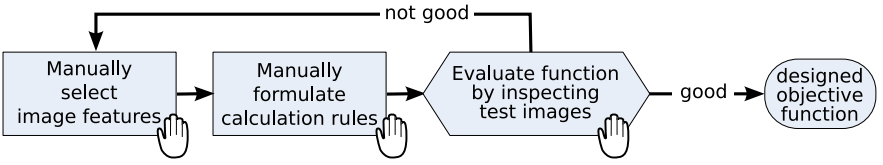

Fig. 5. The traditional procedure for designing objective functions is elaborate and error-prone

results on example images and example model parameterizations. If the result is not satisfactory the function is modified or designed from scratch. This heuristic approach relies on the designer's intuition about a good measure of fitness. Our earlier works [39,40] show that this methodology is erroneous and tedious.

To avoid these shortcomings, our former publications [39,40] propose to learn the objective function from training data generated by an ideal objective function, which only exists for previously annotated images. This procedure enforces the learned function to be approximately ideal as well. Figure 6 illustrates our five-step methodology. It splits the generation of the objective function into several partly automated independent pieces. Briefly, this provides several benefits: first, automated steps replace the labor-intensive design of the objective function. Second, the approach is less error-prone, because annotating example images with the correct fit is much easier than explicitly specifying calculation rules that need to return the correct value for all magnitudes of fitness while considering any facial variation at the same time. Third, this approach does not need any expert knowledge in computer vision and no skills of the application domain and therefore, it is generally applicable. This approach yields robust and accurate objective functions, which greatly facilitate the task of the fitting algorithms.

3.5 The Fitting Algorithm

The fitting algorithm searches for the model that best describes the face visible in the image, by finding the model parameters that minimize the objective function. Fitting algorithms have been the subject of intensive research and evaluation,

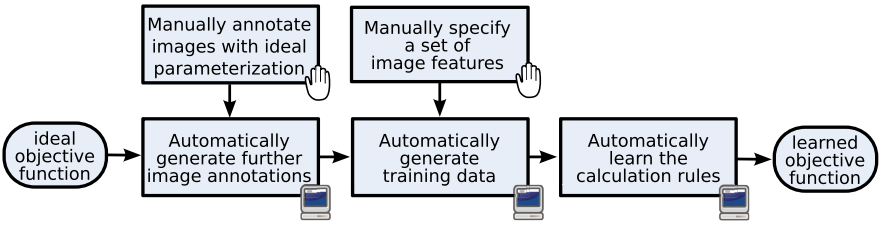

Fig. 6. The proposed method for learning objective functions from annotated training images automates many critical decision steps

see Hanek [41] for an overview and categorization. Since we adapt the objective function rather than the fitting algorithm to the specifics of the application, our approach is able to use any of these standard methods. Due to real-time requirements, the experiments in Section 4 have been conducted with a quick *hill climbing* algorithm. Carefully specifying the objective function makes this local optimization method nearly as accurate as a global optimization strategy.

3.6 Facial Feature Extraction

Two aspects characterize facial expressions: They turn the face into a distinctive state [27] and the muscles involved show a distinctive motion [21,25]. Our approach considers both aspects in order to infer the facial expression, by extracting structural and temporal features. This large set of features provides a fundamental basis for the subsequent classification step, which therefore achieves high accuracy.

Structural Features. The deformation parameters $\boldsymbol{b}$ of the model describe the constitution of the visible face. The examples in Figure 4 illustrate their relation to the facial configuration. Therefore, $\boldsymbol{b}$ provides high-level information to the interpretation process. In contrast, the model's transformation parameters t_x, t_y, s, and θ do not influence the facial configuration and are not integrated into the interpretation process.

Temporal Features. Facial expressions emerge from muscle activity, which deforms the facial components involved. Therefore, the motion of particular feature points within the face is able to give evidence about the facial expression currently performed. In order to meet real-time requirements, we consider a small number of facial feature points only. The relative location of these points is derived from the structure of the face model. Note that these locations are not specified manually, because this assumes the designer is experienced in analyzing

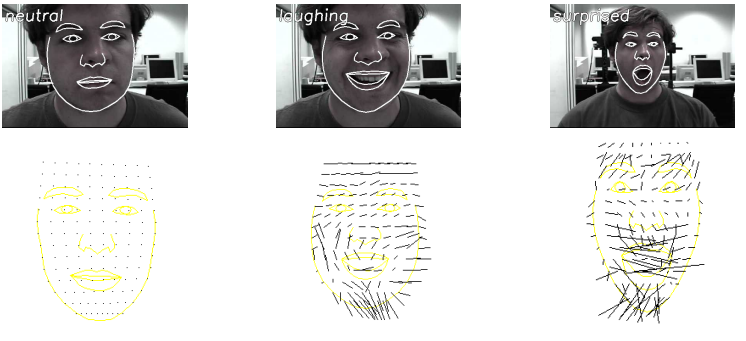

Fig. 7. Model-based techniques greatly support the task of facial expression interpretation. The parameters of a deformable model give evidence about the currently visible state of the face. The lower section shows $G = 140$ facial feature points that are warped in the area of the face model, showing example motion patterns from facial expressions.

Table 1. Confusion matrix and recognition rates of our approach

ground truth	classified as						recognition rate
	surprise	happiness	anger	disgust	sadness	fear	
surprise	28	1	1	0	0	0	93.33%
happiness	1	26	1	2	3	**4**	70.27%
anger	1	1	14	2	2	1	66.67%
disgust	0	2	1	10	3	1	58.82%
sadness	1	2	2	2	22	1	73.33%
fear	1	**5**	1	0	2	13	59.09%
mean recognition rate							**70.25%**

facial expressions. In contrast, a moderate number of G equally distributed feature points is generated automatically, shown in Figure 7. These points should move uniquely and predictably for any particular facial expression. The sum of the motion of each feature point i during a short time period is $(g_{x,i}, g_{y,i})$ for the motion in (x, y), $1 \leq i \leq G$. In our proof-of-concept, the period is 2 seconds, which covers slowly expressed emotions. Furthermore, the motion of the feature points is normalized by subtracting the model's affine transformation. This separates the motion that originates from facial expressions from the rigid head motion.

In order to acquire robust descriptors, Principal Component Analysis determines the H most relevant motion patterns (Principal Components) that are contained within the training sequences. A linear combination of these motion patterns describes each observation approximately. Since $H \ll 2G$, this reduces the number of descriptors by enforcing robustness towards outliers as well. As a compromise between accuracy and runtime performance, we set $G = 140$ and $H = 14$. Figure 7 visualizes the obtained motion of the feature points for some example facial expressions.

3.7 Facial Expression Classification

With the knowledge of the B structural and the H temporal features, we concatenate a feature vector $\boldsymbol{t}$, see Equation 1. It represents the basis for facial expression classification. The structural features describe the model's configuration within the current image and the temporal features describe the muscle activity during a small amount of time.

$$\boldsymbol{t} = (b_1, ..., b_B, h_1, ..., h_H)^T \tag{1}$$

A classifier is intended to infer the correct facial expression visible in the current image from the feature vector $\boldsymbol{t}$. 67% of the image sequences of the CKDB form the training set and the remainder the test set. The classifier used is a Binary Decision Tree [42], which is robust and efficient to learn and execute.

4 Experimental Evaluation

We evaluate the accuracy of our approach by applying it to the unseen fraction of the CKDB. Table 1 shows the recognition rate and the confusion matrix of

Table 2. Recognition rate of our system compared to state-of-the-art approaches

facial expression	Results of our approach	Results of Michel *et al.* [21]	Results of Schweiger *et al.* [25]
Anger	66.7%	66.7%	75.6%
Disgust	64.1%	58.8%	30.0%
Fear	59.1%	66.7%	0.0%
Happiness	70.3%	91.7%	79.2%
Sadness	73.3%	62.5%	60.5%
Surprise	93.3%	83.3%	89.8%
Average	71.1%	71.8%	55.9%

each facial expression. The results are comparable to state of the art approaches, shown in Table 2. The facial expressions happiness and fear are confused most often. This results from similar muscle activity around the mouth, which is also indicated by the sets of AUs in FACS for these two emotions.

The accuracy of our approach is comparable to that of Schweiger *et al.* [25], also evaluated on the CKDB. For classification, they favor motion from different facial parts and determine Principal Components from these features. However, Schweiger *et al.* manually specify the region of the visible face whereas our approach performs an automatic localization via model-based image interpretation. Michel *et al.* [21] also focus on facial motion by manually specifying 22 feature points predominantly located around the mouth and eyes. They utilize a Support Vector Machine (SVM) for determining one of six facial expressions.

4.1 Evaluation in a Robot Scenario

The method was also evaluated in a real robot scenario, using a B21r robot, shown in Figure 8 [43,44]. The robot LCD panel shows a robot face, which is able to make facial expressions, and moves its lips in synchronization with the robot's voice The goal of this work is to provide a robot assistant that is able to express feelings and respond to human feelings.

The robot's 1/3in Sony cameras are just beneath the face LCD, and one is used to capture images of the human in front of the robot. Since the embedded robot computer runs Linux and our model based facial expression recognizer runs under Windows, a file transfer protocol server provided images across a wireless network to a desktop Windows PC.

The robot evaluation scenario was controlled to ensure good lighting conditions, and that the subject's face was facing the camera and forming a reasonable sized part of the image (at a distance of approximately 1m). 120 readings of three facial expressions were recorded and interpreted. Figure 8 shows the results, which are very similar to the bench test results above. The model based FER was able to process images at 12 frames per second, although for the test the speed was reduced to 1 frame per second, in order to eliminate confusion when the human subject changed.

ground truth	classified as			recognition rate
	neutral	happiness	surprise	
neutral	85	11	19	71%
happiness	18	83	25	69%
surprise	17	26	76	63%
mean recognition rate				**67%**

Fig. 8. Set up of our B21 robot and the confusion matrix and recognition rates in a robot scenario

5 Summary and Conclusion

To be effective in the human world robots must respond to human emotional states. This paper focuses on the recognition of the six universal human facial expressions. Our approach imposes lower computational requirements by specializing model-based techniques to the face scenario. Adaptive skin color extraction provides accurate low-level information. Both the face and the facial components are localized robustly. Specifically learned objective functions enable accurate fitting of a deformable face model. Experimental evaluation reports a recognition rate of 70% on the Cohn–Kanade facial expression database, and 67% in a robot scenario, which compare well to other FER systems. The technique provides a prototype suitable for FER by robots.

References

1. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation, and recognition of facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 19, 757–763 (1997)
2. Edwards, G.J., Cootes, T.F., Taylor, C.J.: Face Recognition Using Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406–1607, pp. 581–595. Springer, Heidelberg (1998)
3. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE TPAMI 22, 1424–1445 (2000)
4. Chibelushi, C.C., Bourel, F.: Facial expression recognition: A brief tutorial overview. In: CVonline: On-Line Compendium of Computer Vision (2003)
5. Wimmer, M., Zucker, U., Radig, B.: Human capabilities on video-based facial expression recognition. In: Proc. of the 2nd Workshop on Emotion and Computing – Current Research and Future Impact, Oldenburg, Germany, pp. 7–10 (2007)
6. Schuller, B., et al.: Audiovisual behavior modeling by combined feature spaces. In: ICASSP, vol. 2, pp. 733–736 (2007)
7. Fischer, S., et al.: Experiences with an emotional sales agent. In: André, E., et al. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 309–312. Springer, Heidelberg (2004)

8. Tischler, M.A., et al.: Application of emotion recognition methods in automotive research. In: Proceedings of the 2nd Workshop on Emotion and Computing – Current Research and Future Impact, Oldenburg, Germany, pp. 50–55 (2007)
9. Breazeal, C.: Function meets style: Insights from emotion theory applied to HRI. 34, 187–194 (2004)
10. Rani, P., Sarkar, N.: Emotion-sensitive robots – a new paradigm for human-robot interaction. In: 4th IEEE/RAS International Conference on Humanoid Robots, vol. 1, pp. 149–167 (2004)
11. Scheutz, M., Schermerhorn, P., Kramer, J.: The utility of affect expression in natural language interactions in joint human-robot tasks. In: 1st ACM SIGCHI/SIGART conference on Human-robot interaction, pp. 226–233. ACM Press, New York (2006)
12. Picard, R.: Toward agents that recognize emotion. In: Imagina 1998 (1998)
13. Bartlett, M.S., et al.: Measuring facial expressions by computer image analysis. Psychophysiology 36, 253–263 (1999)
14. Ikehara, C.S., Chin, D.N., Crosby, M.E.: A model for integrating an adaptive information filter utilizing biosensor data to assess cognitive load. In: Brusilovsky, P., Corbett, A.T., de Rosis, F. (eds.) UM 2003. LNCS, vol. 2702, pp. 208–212. Springer, Heidelberg (2003)
15. Vick, R.M., Ikehara, C.S.: Methodological issues of real time data acquisition from multiple sources of physiological data. In: Proc. of the 36th Annual Hawaii International Conference on System Sciences - Track 5, p. 129.1. IEEE Computer Society, Los Alamitos (2003)
16. Sheldon, E.M.: Virtual agent interactions. PhD thesis, Major Professor-Linda Malone (2001)
17. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: International Conference on Automatic Face and Gesture Recognition, France, pp. 46–53 (2000)
18. Ekman, P.: Universals and cultural differences in facial expressions of emotion. In: Nebraska Symposium on Motivation 1971, vol. 19, pp. 207–283. University of Nebraska Press, Lincoln, NE (1972)
19. Ekman, P.: Facial expressions. In: Handbook of Cognition and Emotion, John Wiley & Sons Ltd., New York (1999)
20. Friesen, W.V., Ekman, P.: Emotional Facial Action Coding System. Unpublished manuscript, University of California at San Francisco (1983)
21. Michel, P., Kaliouby, R.E.: Real time facial expression recognition in video using support vector machines. In: Fifth International Conference on Multimodal Interfaces, Vancouver, pp. 258–264 (2003)
22. Cohn, J., et al.: Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. Psychophysiology 36, 35–43 (1999)
23. Sebe, N., et al.: Emotion recognition using a cauchy naive bayes classifier. In: Proc. of the 16th International Conference on Pattern Recognition (ICPR 2002)., vol. 1, p. 10017. IEEE Computer Society, Los Alamitos (2002)
24. Cohen, I., et al.: Facial expression recognition from video sequences: Temporal and static modeling. CVIU special issue on face recognition 91(1-2), 160–187 (2003)
25. Schweiger, R., Bayerl, P., Neumann, H.: Neural architecture for temporal emotion classification. In: André, E., et al. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 49–52. Springer, Heidelberg (2004)
26. Tian, Y.L., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE TPAMI 23, 97–115 (2001)

27. Littlewort, G., et al.: Fully automatic coding of basic expressions from video. Technical report, University of California, San Diego, INC., MPLab (2002)
28. Ekman, P., Friesen, W. (eds.): The Facial Action Coding System: A Technique for The Measurement of Facial Movement. Consulting Psychologists Press, San Francisco (1978)
29. Bartlett, M., et al.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: CVPR 2005. IEEE Computer Society Conference, vol. 2, pp. 568–573 (2005)
30. Kim, D.H., et al.: Development of a facial expression imitation system. In: International Conference on Intelligent Robots and Systems, Beijing, pp. 3107–3112 (2006)
31. Yoshitomi, Y., et al.: Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face. In: Proc. of the 9th IEEE International Workshop on Robot and Human Interactive Communication, Osaka, Japan, pp. 178–183 (2000)
32. Otsuka, T., Ohya, J.: Recognition of facial expressions using hmm with continuous output probabilities. In: 5th IEEE International Workshop on Robot and Human Communication, Tsukuba, Japan, pp. 323–328 (1996)
33. Zhou, X., et al.: Real-time facial expression recognition based on boosted embedded hidden markov model. In: Proceedings of the Third International Conference on Image and Graphics, Hong Kong, pp. 290–293 (2004)
34. Wimmer, M.: Model-based Image Interpretation with Application to Facial Expression Recognition. PhD thesis, Technische Universitat München, Fakultät für Informatik (2007)
35. Cootes, T.F., Taylor, C.J.: Active shape models – smart snakes. In: Proc. of the 3rd BMVC, pp. 266–275. Springer, Heidelberg (1992)
36. Cootes, T.F., et al.: The use of active shape models for locating structures in medical images. In: IPMI, pp. 33–47 (1993)
37. Wimmer, M., Radig, B., Beetz, M.: A person and context specific approach for skin color classification. In: Proc. of the 18th ICPR 2006, vol. 2, pp. 39–42. IEEE Computer Society Press, Los Alamitos, CA, USA (2006)
38. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR, Kauai, Hawaii, vol. 1, pp. 511–518 (2001)
39. Wimmer, M., et al.: Learning local objective functions for robust face model fitting. In: IEEE TPAMI (to appear, 2007)
40. Wimmer, M., et al.: Learning robust objective functions with application to face model fitting. In: DAGM Symp., vol. 1, pp. 486–496 (2007)
41. Hanek, R.: Fitting Parametric Curve Models to Images Using Local Self-adapting Separation Criteria. PhD thesis, Department of Informatics, Technische Universität München (2004)
42. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, California (1993)
43. Yadav, A.: Human facial expression recognition. Technical report, Electrical and Computer Engineering, University of Auckland, New Zealand (2006)
44. Jayamuni, S.D.J.: Human facial expression recognition system. Technical report, Electrical and Computer Engineering, University of Auckland, New Zealand (2006)