
The Facial Expression Module

Carmen Frank, Johann Adelhardt, Anton Batliner, Elmar Nöth, Rui Ping Shi, Viktor Zeißler, Heinrich Niemann

Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
{frank, adelhardt, batliner, noeth, shi, zeissler,
niemann}@informatik.uni-erlangen.de,

Summary. In current dialogue systems the use of speech as an input modality is common. But this modality is only one of those human beings use. In human–human interaction people use gestures to point or facial expressions to show their moods as well. To give modern systems a chance to read information from all modalities used by humans, these systems must have multimodal user interfaces. The SMARTKOM system has such a multimodal interface that analyzes facial expression, speech and gesture simultaneously. Here we present the module that fulfills the task of facial expression analysis in order to identify the internal state of a user.

In the following we first describe the state of the art in emotion and user state recognition by analyzing facial expressions. Next, we describe the *facial expression recognition module*. After that we present the experiments and results for recognition of user states. We summarize our results in the last section.

1 State of the Art

The knowledge about a user’s facial expression can be used, e.g., to recognize that a helpless user needs some help or to adapt presented information in case of confusion. The systems dealing with classification of facial images, which were developed over the past years, can be differentiated according to the following characteristics:

- sequence or single image
- recognizable classes
- used features
- used classifier

For the field of **single images**, the decision for a class is done with respect to one single image. Methods using single images can be found in Wiskott (1997); Chen and Huang (2002), Kumar and Poggio (2000), and Thomaz et al. (2001). **Images sequences** are used in Kaiser et al. (1998), Essa and Pentland (1997), and Müller et al. (2001). These methods and similar ones use the variation over time, for instance the optical flow.

Some systems cannot be associated with single-image or image-sequence methods. Samples of such hybrid systems can be found in Schwerdt et al. (2000), Lien

et al. (1998a), and Tian et al. (2001). The first system classifies each frame of a sequence, and the results form a trajectory of changes in the face, which is classified using an eigenspace. Both other systems use a *neutral* image to find relative changes to the current one. This comparison allows users to determine the intensity of the facial expression.

Depending on the field of application, a facial expression recognition system must have a special **category set**, to which the faces should be classified. Very common is the concept of six basic emotions (happiness, sadness, anger, fear, disgust and surprise) used by Ekman (Lyons et al., 1999; Thomaz et al., 2001; Otsuka and Ohya, 1997). Subsets of these emotions are also used. Often there is one additional expression, “raised eyebrows” (Müller et al., 2001; Schwerdt et al., 2000; Essa and Pentland, 1995).

Besides these complex categories, there are several methods which use action units (AUs) (Ekman and Friesen, 1978) as a category set: Lien et al. (1998a), Tian et al. (2000), Tian et al. (2001), Otsuka and Ohya (1999), and Braathen et al. (2001). AUs are the smallest units in facial movements similar to phonemes for speech. The major interest to the AUs arises from their relevance in field of psychology. Because AUs are a standard technique for coding facial expressions, automatic coding would provide a much larger set of data for further experiments in psychology.

There are only a few papers concerning classification of user states that are not part of the basic emotions. Most systems that detect states necessary for a human computer interaction (attention, point of interest, stress, fatigue, etc.) use touch sensors. One example for the recognition of fatigue from an image sequence is shown in Li et al. (1999).

AUs are not only used as a category system, but also as **features**. Some systems using AUs as features are explained in Lien et al. (1998b), Otsuka and Ohya (1999), and Cohn et al. (1998).

For feature extraction to expression classification, one can distinguish between template-based and analytic methods. Template-based methods match a face model to a face. Then positions of facial features are extracted from this model, or muscle movements are derived. Template-based methods use graphs (Wiskott, 1997), models of facial features, e.g., the mouth (Oliver et al., 1997; Tian et al., 2000), or work with hand-segmented facial features. Essa explains (Essa and Pentland, 1997) a two-step system by matching a triangle mesh, using optical flow (Simoncelli, 1993) to the face and matching the result to a muscle model of the face.

In feature-based methods one can often find systems using wavelets. Two common types of wavelets are Haar wavelets (Kumar and Poggio, 2000) and Gabor wavelets (Lyons et al., 1999; Hu et al., 2002).

The systems for facial expression **classification** found in literature differ not only in the used features but also in the used classifier. They can be subdivided into model-, rule- and neural network-based methods. As model-based methods, principal component analysis, PCA (Schwerdt et al., 2000; Thomaz et al., 2001; Kirby and Sirovich, 1990) or support vector machines SVM (Braathen et al., 2001) are used. If the classification is done using image sequences and not only single images, the information of the single frames can be combined. For this task Hidden

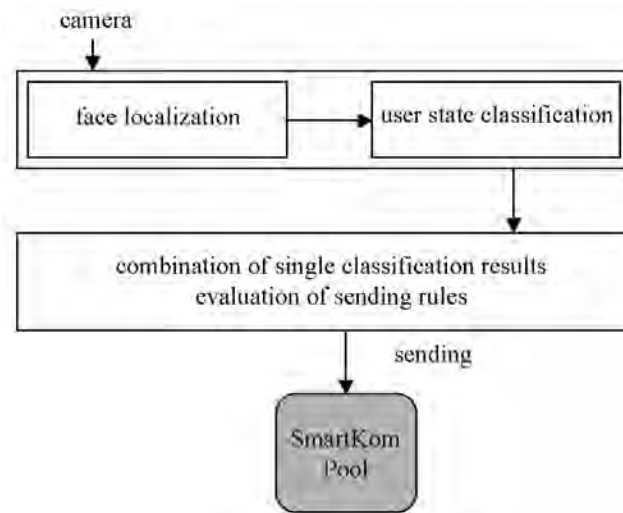


Fig. 1. The general architecture of the facial expression module

Markov Models are well known (Otsuka and Ohya, 1997; Müller et al., 2001; Oliver et al., 1997; Lien et al., 1998a). Rule-based methods include rules to map features to classes. In Chen and Huang (2002) the results from different classification stages are fused using handmade rules. An expert system using AUs for classifying facial expressions is explained in Pantic (2001). Methods based on neural nets can be found in Tian et al. (2000), Hu et al. (2002), and Zhang et al. (1998).

Furthermore, methods and systems for facial expression classification can be differentiated using facts like

- the use of handmade markers in the face
- the initialization of a sequence by hand
- the selection of features by hand
- person dependence or independence

The presented facial expression module uses a localization technique based on color because this is a feature common to all faces. As classifiers, eigenspaces PCA and SVM are used. Both allow the use of pixel values as features instead of segmented facial features such as eyebrows and lip corners. The motivation for not hand selecting features is the success of probabilistic, segmentation-free methods for spontaneous speech recognition. Facial expressions during a multimodal dialogue are equivalent to spontaneous utterances.

2 Module Description

The facial expression module consists of three main parts: **localization**, **classification**, and **combination and sending** of data (Fig. 1) in order to fulfill the task of



Fig. 2. Original image (*left*); result when skin color segmentation is done (*right*). The image parts with blue and green color shades disappear, but white shades and natural colors stay

facial expression recognition as part of a multimodal dialogue system. They are explained in more detail in the following sections.

The facial expression recognition does not try to detect the basic emotions like *disgust*, etc., but tries to detect user states. User states are all these internal states of a user that influence his or her interaction with the automatic system. For an information system similar to SMARTKOM, user states like *hesitant* or *angry* are suitable.

2.1 Localization

The localization of a face in the SMARTKOM environment has several constraints. There is at most one person using SMARTKOM, but this person is unknown to the system. The background is dynamic and is also unknown. It is not possible to use motion information to locate a person. For this reason we decided to use facial color as a first step in person localization.

Because skin color is not very clearly defined in the RGB color space, we use the YUV color space. In Fig. 2 one can see the results of a skin color segmentation. A nontypical image is used for better illustration: nonfacial skin color can be found as well. It is possible to eliminate “cold” colors. The resulting pixels are used as hypotheses for the further calculations.

To find the real position of the face, the facial expression module uses a holistic classifier and a filter. The holistic classifier can be a SVM, see Osuna et al. (1996) or PCA, see Turk and Pentland (1991a).

A problem when separating faces from “nonfaces” is that a nonface is not clearly defined. It is not possible to collect a set of all kinds of nonfaces. Therefore, the holistic classifiers sometimes select an area in an image as a face which is, for a human observer, clearly not a face. This problem can be reduced by a filter. We use a bank of circle frequency (CF) filters (Kawato and Tetsutani, 2002). The CF filter tries to find an area with alternation of light and dark values, which can be found near the between-eyes point (dark eyes, light forehead and nose). The effects of this filter are illustrated in Fig. 3.



Fig. 3. The *left* face is the original face used to calculate the CF filter. The *middle* image shows the pixels with high weighting in light values. For a better illustration, both images are combined (*right*)

2.2 Classification

After the position of the face/the person is localized, the user state of this person can be classified. The classification is done by holistic classifiers as well. The method proposed by us for the recognition of facial expressions is a modification of a standard eigenspace classification for user identification. Eigenspace methods are well known in the field of face recognition (Turk and Pentland, 1991b; Yambor et al., 2000; Moghaddam and Pentland, 1994). In a standard face recognition system, one eigenspace for each person is created using different face images. The set of face images for each person is used to create a probability distribution or a representative for this person in the face space. Later, when classifying an image of an unknown person, this image is projected onto the face spaces. The probability distribution or representative that best matches the new image is chosen as the class in question.

To create an eigenspace with training images, a partial Karhunen–Loève transformation, also called principal component analysis (PCA), is used. This transformation is a dimensionality reduction scheme that maximizes the scatter of all projected samples, using N sample images of a person $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ with values in an n -dimensional feature space. Let μ be the mean image of all feature vectors. The total scatter matrix is then defined as

$$S_T = \sum_{k=1}^N (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T. \quad (1)$$

In PCA, the optimal projection W_{opt} to a lower-dimensional subspace is chosen to maximize the determinant of the total scatter matrix of the projected samples

$$W_{opt} = \arg \max_W |W^T S_T W| = [w_1, w_2, \dots, w_m], \quad (2)$$

where $\{w_i | i = 1, 2, \dots, m\}$ is the set of n -dimensional eigenvectors of S_T corresponding to the set of decreasing eigenvalues. These eigenvectors have the same dimension as the input vectors and are referred to as “eigenfaces”. In Fig. 4 the first four eigenfaces of the *angry* eigenspace are shown. In the following we assume that high-order eigenvectors correspond to high eigenvalues. Therefore, high-order eigenvectors contain more relevant information.

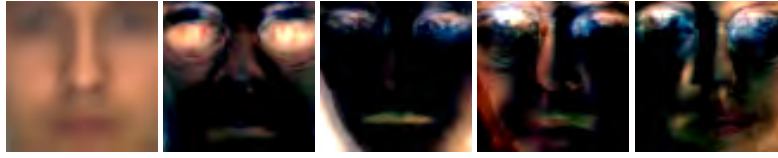


Fig. 4. The *left* image is the average image; the following images are the first four eigenvectors of the *angry* eigenspace (Frank and Nöth, 2001)

For face classification a new image is projected to each eigenspace and the eigenspace which best describes the input image is selected. This is accomplished by calculating the residual description error.

Imagine we have training sets F_κ of l samples \mathbf{y}_i with similar characteristics for each class Ω_κ , $\kappa \in 1, \dots, k$. There is different illumination, different face shape, etc., in each set F_κ . Reconstructing one image \mathbf{y}_i with each of our eigenspaces results in k different samples \mathbf{y}^κ . The reconstructed images do not differ in characteristics like illumination, because this is modeled by each eigenspace. But they differ in the facial expression of specific regions, such as the mouth or the eyes area.

With a set of eigenspaces for each class Ω_κ we receive distances v_κ of a test image \mathbf{y}_i to each class

$$v_\kappa = \|\mathbf{y}_i - \mathbf{y}^\kappa\|^2 \quad (3)$$

$$.k = \arg \min_{j \in 0..k} v_\kappa. \quad (4)$$

An image is attributed to a class k with minimum distance as criterion.

In addition to the modeling of separate models for each user state, we use only a subset of the available pixels, i.e., only “significant” pixels; the extraction of this subset is described in Frank and Nöth (2003). This method creates a map of significance of pixels and allows us to delete less significant ones. In Fig. 5 one can see such a map (leftmost image). Only pixels significant for facial expression (mouth and eye areas) are selected as features for the holistic classifier.

2.3 Combination and sending

The classification step produces results for every captured frame, but a facial expression has a longer duration than a single frame. Normally, it starts with a neutral expression, then the expression of the user state evolves and goes back to the neutral expression. Therefore, the single classification results are combined in order to send fewer messages to the SMARTKOM pool.

There are some rules to decide whether a message should be sent or not:

1. A new user state is detected.
2. No new user state has been detected for a long time.
3. The intensity of the recent user state changed significantly.



Fig. 5. The *leftmost* image shows the map of significance for a *smile–anger* classification. Different thresholds are used to create the other images (white pixels represent significant ones)

The second rule is necessary because the system needs to know that nothing happened, but the facial expression module is still “alive” and keeps watching the user. The values for *long time* and *significantly* used in rules 2 and 3 can be set using the GUI of the module. This is necessary because the amount of data processed by the facial expression module depends on the computational resources needed by other modules.

The sent data are a combination of single classification resulting from the last analyzed frames. The older a single classification result is, the lower is its influence to the overall result.

3 Experiments

For examination of the proposed classification methods we need a special dataset, including video sequences and corresponding user state labels. In the WOZ dataset of the SMARTKOM project (Schiel and Türk, 2006), one can only find a very small subset showing expressive user state. Therefore, we collected our own multimodal dataset with acted user states (Adelhardt et al., 2006). For the following experiments we used sequences of ten persons with, for a human being, recognizable expressions, in a leave-one-out manner. The whole face is used to create four eigenspaces for four facial expression classes. When a naive user tried to express an utterance in an angry manner, he/she often displayed neutral facial expressions before and after speaking, thus not every image of a sequence shows the same user state. The same is true for the other user states. The labeling of each single image is not possible due to the complexity. A consequence of that is a low recognition rate for neutral, because each class-specific dataset includes neutral faces. The classification of faces with internal movements, produced due to speaking, is very difficult. We achieve a low recognition rate of 32%. The confusion matrix is shown in Table 1. A problem is the user state *angry*. Anger a facial expression that is shown in many different ways by different users, whereas joy is consistently marked by risen lip corners.

The same procedure applied to a dataset (presented in Martinez and Benavente (1998)) of mugshots, showing four different facial expressions, yields 59%. The high recognition rate of the still images are due to reasons given above.

Table 1. Confusion matrix of user state recognition with facial expression data (%)

Reference user state	Results			
	Neutral	Joyful	Angry	Hesitant
Neutral	7	23	36	33
Joyful	5	54	22	20
Angry	4	62	17	16
Hesitant	6	12	35	48

One idea to solve the latter problem is the elimination of face regions influenced by speaking. To find out which part of a face is able to identify which facial expression, we performed experiments using only parts of a face for classification.

For this task it is necessary to know the position of eyes, nose and mouth and to have full-blown facial expression. Therefore, we use the database presented in Martinez and Benavente (1998). So the first-mentioned problem (not every frame shows an expression) does not influence our results and we can concentrate on the face regions.

To classify face regions, the faces are split into three parts (eye, nose, and moutharea). Samples of these three regions can be seen in Fig. 6. When only using one facial region for expression classification, the region centered at the mouth gives the best results (79%). The eye area achieves 63%, and the nose area 65%. The confusion matrices for all three face regions are shown in Table 2. It is noticeable that the different regions have different capabilities in indentifying different facial expressions. The mouth area gives really good results identifying *smile* and *shout* expressions, but *neutral* and *anger* are mixed up more often. However, these two classes can be told apart by the eye region. Astonishingly, the eye area had a good classification rate for identifying shouting. The reason for this is obvious when viewing images with the facial expression *shout* (Fig. 7). Normally, a human closes or pinches his/her eyes while shouting. This is one feature to separate *shout* from *open the mouth* (which is not part of the dataset). All other expressions are classified with nearly equal recognition rates.

So if one of the face regions is occluded or for other reasons is not available for classification, it is possible to recognize facial expressions. But this is not the case in all situations. While using an automatic dialogue system, a user has to speak. Speaking changes the mouth area but does not affect the user state. This makes the mouth

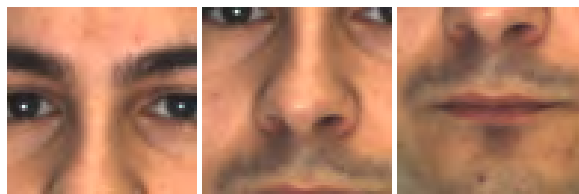
**Fig. 6.** The three parts of a face used for facial expression classification

Table 2. Confusion matrix (values in %) of facial expression recognition for the face areas. The best result in each line is in **boldface**: the eye area (**top**, overall recognition rate: 63%), the nose area (**middle**, overall recognition rate: 65%), the mouth area (**bottom**, overall recognition rate: 79%)

Eye area				
Reference	Results			
facial expression	<i>smile</i>	<i>neutral</i>	<i>anger</i>	<i>shout</i>
<i>smile</i>	60	13	19	8
<i>neutral</i>	14	52	29	5
<i>anger</i>	9	26	59	6
<i>shout</i>	5	8	4	83
Nose area				
Reference	Results			
facial expression	<i>smile</i>	<i>neutral</i>	<i>anger</i>	<i>shout</i>
<i>smile</i>	77	6	7	10
<i>neutral</i>	1	46	47	6
<i>anger</i>	2	25	71	2
<i>shout</i>	22	7	5	66
Mouth area				
Reference	Results			
facial expression	<i>smile</i>	<i>neutral</i>	<i>anger</i>	<i>shout</i>
<i>smile</i>	97	2	1	0
<i>neutral</i>	2	55	43	0
<i>anger</i>	0	33	67	0
<i>shout</i>	2	2	0	96

area classifier less trustworthy during speaking. A fusion of the classification results of other region classifiers may improve the overall result. Therefore, we examine the fusion of classification results.

3.1 Fusion

Each facial expression may be accentuated in another face region. We developed rules to combine results of the classifier for each face region dependent on user states. For the class *shout* it is clear (96%) that the mouth area is a trustworthy indicator.

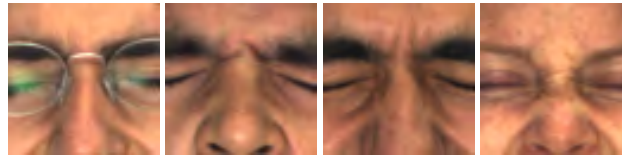


Fig. 7. Some samples of eye regions where the mouth concerning facial expression *shout* is clearly visible

The last section of Table 2 also shows that *smile* is classified correctly (with 97%). So the first two rules are:

Rule 1

IF	mouth area classifier classifies shout
THEN	overall result is shout

Rule 1

IF	mouth area classifier classifies smile
THEN	overall result is smile

These two rules are based on the low false alarm rate of the mouth area classifier for the classes *smile* and *shout* (4% for *smile* and 0% for *shout*). The classification rate is 96% and 97% for *shout* and *smile*, respectively. The high recognition rates combined with low false alarm rate makes the mouth area a trustworthy face region for classification of *smile* and *shout*.

During speaking it is not possible to use the mouth area for classification, because this face region has considerable changes due to moving lips. In this case the other regions of the face should be used for facial expression classification. The decision whether the person is speaking or not at the moment can be obtained from a speech recognition system.

The following experiments use only test images with classes not equal to *shout*. The eye area classifier achieves 57%, and the nose area classifier 64%. We introduce the following rules:

Rule 3

IF	both classifiers have same result
THEN	overall result is that result

Rule 4

IF	(one classifier says shout) AND (one classifier says smile)
THEN	overall result is smile

Rule 5

IF	(eyes classifier says neutral) AND (nose classifier says anger)
THEN	overall result is neutral

Rule 6

IF	not classified yet
THEN	overall result is that of nose classifier

The rules were developed using a subset of the total data. This subset was deleted from the test set in order to not influence recognition rates. Using these rules gives an improvement of the recognition rate of 10% (eye area), resp. 3% (nose area).



Fig. 8. Neither the eye nor the nose area classifier was able to classify these images. The displayed facial expressions are (from left to right): *neutral*, *smile*, *anger*. The first one was classified as *smile* (eye area), *shout* (nose area); The second as *anger* (both), the third as *smile* (eye area) and *neutral* (nose area)

The recognition rate of combining the eye area classifier and the nose area classifier achieves 67%. All classifiers missed the right decision for 23% of the images. It is not possible to solve the classification for these images. Some of these misclassified images are shown in Fig. 8. The confusion matrix for the fusion of results from eye and nose area classifiers is given in Table 3.

4 Conclusion

Facial expression recognition in an automatic dialogue system context has two additional problems compared to facial expression recognition known from the literature (Lien et al., 1998b; Essa and Pentland, 1995; Müller et al., 2001): the expressions are weak because users try to hide them, and users are speaking. The latter, a normal behavior for a dialogue, results in facial changes near the mouth not related to internal user states.

In cases where the mouth area could not be used for classification, a classification and fusion of local area classifiers allows the detection of the current user state. A disadvantage is that the combination results depend on the facial expression. Each face region displays each facial expression with different weights. The combination rules shown above obtain an improvement of 3% (resp., 10%) absolute to 67%.

Table 3. Confusion matrix (values in %) of fusion of results from eye and nose area classifiers; overall recognition rate is 67%

Reference facial expression	Results			
	<i>smile</i>	<i>neutral</i>	<i>anger</i>	<i>shout</i>
<i>smile</i>	76	8	11	5
<i>neutral</i>	9	66	19	6
<i>anger</i>	3	36	58	3

References

- J. Adelhardt, C. Frank, E. Nöth, R.P. Shi, V. Zeißler, and H. Niemann. Multimodal Emogram, Data Collection and Presentation, 2006. In this volume.
- B. Braathen, M.S. Bartlett, G. Littlewort, and J.R. Movellan. First Steps Towards Automatic Recognition of Spontaneous Facial Action Units. In: *Proc. ACM Conf. on Perceptual User Interfaces (PUI'03)*, pp. 319–242, Vancouver, Canada, 2001.
- X.W. Chen and T. Huang. Facial Expression Recognition: A Clustering-Based Approach. *Pattern Recognition Letters*, 24:1295–1302, 2002.
- J.F. Cohn, A.J. Zlochower, J.J. Lien, and T. Kanade. Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expressions. In: *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 390–395, 1998.
- P. Ekman and W.V. Friesen. *The Facial Action Coding System — A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.
- I.A. Essa and A.P. Pentland. Facial Expression Recognition Using a Dynamic Model and Motion Energy. In: *Proc. 5th Int. Conf. on Computer Vision*, pp. 360–367, Cambridge, MA, 1995.
- I.A. Essa and A.P. Pentland. Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):757–763, 1997.
- C. Frank and E. Nöth. Automatic Pixel Selection for Optimizing Facial Expression Recognition Using Eigenfaces. In: *Pattern Recognition, Proc. 25rd DAGM Symposium*, pp. 378–385, Magdeburg, Germany, 2001.
- C. Frank and E. Nöth. Optimizing Eigenfaces by Face Masks for Facial Expression Recognition. In: *Computer Analysis of Images and Patterns – CAIP 2003*, LNCS, pp. 1–13, Berlin Heidelberg New York, 2003. Springer.
- T. Hu, L.C.D. Silva, and K. Sengupta. A Hybrid Approach of NN and HMM for Facial Emotion Classification. *Pattern Recognition Letters*, 23:1303–1310, 2002.
- S. Kaiser, T. Wehrle, and S. Schmidt. Emotional Episodes, Facial Expressions, and Reported Feelings in Human-Computer Interactions. In: *Proc. 10th Conf. of the Int. Society for Research on Emotion*, pp. 82–86, Würzburg, Germany, 1998. ISRE Publications.
- S. Kawato and N. Tetsutani. Real-Time Detection of Between-the-Eyes With a Circle Frequency Filter. In: *Proc. 5th Asian Conference on Computer Vision (ACCV)*, pp. 442–447, Melbourne, Australia, 2002.
- M. Kirby and L. Sirovich. Application of the Karhunen–Loève Procedure for the Characterization of Human Faces. *TPAMI*, 12(1):103–108, 1990.
- V.P. Kumar and T. Poggio. Learning-Based Approach to Real Time Tracking and Analysis of Faces. In: *Automatic Face and Gesture Recognition 2000*, pp. 96–101, 2000.
- H. Li, A. Lundmark, and R. Forchheimer. Video Based Human Emotion Estimation. In: *Int. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, Sept. 1999.

- J.J. Lien, T. Kanade, J.F. Cohn, and C.C. Li. A Multi-Method Approach for Discriminating Between Similar Facial Expressions, Including Expression Intensity Estimation. In: *Proc. Computer Vision and Pattern Recognition (CVPR'98)*, pp. 853–859, 1998a.
- J.J. Lien, T. Kanade, J.F. Cohn, and C.C. Li. Automated Facial Expression Recognition Based on FACS Action Units. In: *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 390–395, 1998b.
- M.J. Lyons, L. Budynek, and S. Akamatsu. Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(12):1357–1362, 1999.
- A.M. Martinez and R. Benavente. The AR FAcE DAtabase. Technical Report 24, Computer Vision Center, Purdue University, West Lafayette, IN, 1998.
- B. Moghaddam and A.P. Pentland. Face Recognition Using View-Based and Modular Eigenspaces. In: *Automatic Systems for the Identification and Inspection of Humans, SPIE'94*, vol. 2257, 1994.
- M. Müller, F. Wallhoff, F. Hülsken, and G. Rigoll. Facial Expression Recognition Using Pseudo 3-D Hidden Markov Models. In: *Pattern Recognition, Proc. 23rd DAGM Symposium*, pp. 291–297, Munich, Germany, 2001.
- N. Oliver, A.P. Pentland, and F. Berard. LAFTER: Lips and Face Real Time Tracker. In: *Proc. Computer Vision and Pattern Recognition (CVPR'97)*, pp. 123–129, Puerto Rico, 1997.
- E.E. Osuna, R. Freund, and F. Girosi. Support Vector Machines: Training and Application. Technical Report A. I. Memo No. 1602, Massachusetts Institute of Technology, Cambridge, MA, 1996.
- T. Otsuka and J. Ohya. Recognizing Multiple Persons' Facial Expressions Using HMM Based on Automatic Extraction of Significant Frames From Image Sequences. In: *Int. Conf. on Image Processing (ICIP 1997)*, pp. 546–549, Oct 1997.
- T. Otsuka and J. Ohya. Extracting Facial Motion Parameters by Tracking Feature Points. In: *Proc. 1st Int. Conf. AMCP'98*, LNCS 1554, pp. 433–444, Berlin Heidelberg New York, 1999. Springer.
- M. Pantic. *Facial Expression Analysis by Computational Intelligence Techniques*. PhD thesis, Faculteit der Informatietechnology en Systemen, TU Delft, The Netherlands, 2001.
- F. Schiel and U. Türk. Wizard-of-Oz Recordings, 2006. In this volume.
- K. Schwerdt, D. Hall, and J.L. Crowley. Visual Recognition of Emotional States. In: *Proc. Int. Conf. on Multimodal Interfaces (ICMI'00)*, pp. 41–48, Beijing, China, 2000.
- E.P. Simoncelli. *Distributed Analysis and Representation of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, 1993.
- C.E. Thomaz, D.F. Gillies, and R.Q. Feitosa. Using Mixture Covariance Matrices to Improve Face and Facial Expression Recognitions. In: *Proc. 3rd Int. Conf. of Audio- and Video-Based Biometric Person Authentication AVBPA R01*, LNCS 2091, pp. 71–77, Berlin Heidelberg New York, June 2001. Springer.

- Y. Tian, T. Kanade, and J.F. Cohn. Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(2):97–115, 2001.
- Y.L. Tian, T. Kanade, and J.F. Cohn. Recognizing Lower Face Action Units for Facial Expression Analysis. In: *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 484–490, 2000.
- M. Turk and A.P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991a.
- M. Turk and A.P. Pentland. Face Recognition Using Eigenfaces. In: *Proc. Computer Vision and Pattern Recognition (CVPR'98)*, pp. 586–591, 1991b.
- L. Wiskott. Phantom Faces for Face Analysis. *Proc. 7th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'97, Kiel*, 1296:480–487, 1997.
- W.S. Yambor, B.A. Draper, and J.R. Beveridge. Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures. In: *2nd Workshop on Empirical Evaluation Methods in Computer Vision*, Dublin, Ireland, 2000.
- Z. Zhang, M. Lyon, M. Schuster, and S. Akamatsu. Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron. In: *Proc. 3rd Int. Conf. on Automatic Face and Gesture Recognition*, pp. 454–459, Nara, Japan, 1998.