A Hidden Markov Model Based Approach for Facial Expression Recognition in Image Sequences

Miriam Schmidt, Martin Schels, and Friedhelm Schwenker

Institute of Neural Information Processing,
University of Ulm, 89069 Ulm, Germany
{miriam.k.schmidt,martin.schels,friedhelm.schwenker}@uni-ulm.de

Abstract. One of the important properties of hidden Markov models is the ability to model sequential dependencies. In this study the applicability of hidden Markov models for emotion recognition in image sequences is investigated, i.e. the temporal aspects of facial expressions. The underlying image sequences were taken from the Cohn-Kanade database. Three different features (principal component analysis, orientation histograms and optical flow estimation) from four facial regions of interest (face, mouth, right and left eye) were extracted. The resulting twelve paired combinations of feature and region were used to evaluate hidden Markov models. The best single model with features of principal component analysis in the region face achieved a detection rate of 76.4 %. To improve these results further, two different fusion approaches were evaluated. Thus, the best fusion detection rate in this study was 86.1 %.

1 Introduction

For over a hundred years people are interested in understanding emotions. One of the most important and pioneering researcher, who has dealt with this issue, was Charles Darwin [6]. By examining the interaction between facial muscles and the associated emotions, he introduced the first rules of emotion recognition. Paul Ekman developed a so-called Facial Action Coding System (FACS) [9] to encode the emotions using the facial muscles. He distinguished six basic emotions: "joy", "anger", "surprise", "disgust", "sadness" and "fear".

Hidden Markov Models (HMMs) are often used in speech recognition [15] and increasingly also for emotion recognition [20,4] because they are able to model temporal dependencies. HMMs are probabilistic models which consist of a countable number of states, transitions and corresponding emissions. HMMs are easy to model, but variable by the parameters that describe them.

Lisetti and Rumelhart [14] proposed a NN-based approach to recognize the facial expressions in which they selected different face regions manually. Lien et al. [13] compared different methods that use optical flow. Lin et al. [7] utilized principle component analysis (PCA) and hierarchical radial basis functions for facial emotion recognition.

In this study, HMMs were utilized for recognition of facial expressions in image sequences. The aim was to investigate the behavior of HMMs in facial expression recognition and to demonstrate whether HMMs are capable of recognizing emotions in image sequences sufficiently well. Hence, four regions in the image sequences were selected manually: face, mouth, right and left eye. Afterward, in each region of the image, three different features, namely principal component analysis, orientation histograms and optical flow estimation, were extracted. The resulting twelve paired combinations of feature and region were used to evaluate the HMMs. Numerous experiments to optimally adjust the HMMs were conducted. The optimal number of states and the optimal number of the normal distributions of the Gaussian Mixture Models (GMMs), which were attached to the states, were determined empirically. Additionally, two different model architectures were evaluated. To improve the over-all performance further, two approaches to fuse the results of the twelve individual models were developed.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the HMMs, GMMs and the two classifier fusion approaches. In Sect. 3 the images and the extraction of the features are described. The experiments and the results will be introduced in Sect. 4. Section 5 gives a brief summary of the paper.

2   Stochastic and Functional Principles

2.1   Hidden Markov Models

A Hidden Markov Model (HMM) $\lambda = (Z, V, \pi, A, E)$ is a statistical model which is composed of two random processes [8][16]. The first process is a Markov chain consisting of a defined number of states $Z = (z_1, ..z_n)$ and corresponding state transition probabilities which form the transition matrix $A = (a_{ij})$. The probability a_{ij} designates the probability to change from state z_i to state z_j. The third component of this process is the initial probability vector π which defines the probabilities of the states to be the initial state. The second random process defines the output: it consists of possible observations $V = (v_1...v_m)$ and the observation matrix $E = \{e_j(k)\}$ storing the probabilities $e_j(k)$ of observation v_k being produced from the state z_j. The sequence of observations provides information about the sequence of the hidden states.

The topology of the transition matrix defines the structure of the model. A connection between two states z_i and z_j is given, if the corresponding entry a_{ij} is greater than 0. The following two models were utilized in this study: *fully connected model* with $a_{ij} > 0 \ \forall \ i, j$ and a *forward model* with $a_{ij} = 0$ for $i > j$ and $j > i + 1$, i.e. only connections to the state itself and to the following state are allowed.

One property of the HMM is that the next state only depends on the current state. The values at time $t - i$, $i > 1$ have no influence. This is called the Markov property, e.g. for a sequence of states $Q = q_1...q_L, q_i \in Z$:

$$P(q_{t+1} = z_j \mid q_t = z_i, q_{t-1} = z_k, ...) = P(q_{t+1} = z_j \mid q_t = z_i) \ . \qquad (1)$$

There are three basic problems associated with HMMs [16]. Each of it can be solved with a specific dynamic programming algorithm:

Decoding Problem. Given the parameters of the model $\lambda = (Z, V, \pi, A, E)$ and a observed output sequence $O = O_1, ..., O_L$. Evaluate the most likely state sequence which could have generated the output sequence.
Solution: *Viterbi algorithm.*

Evaluation Problem. Given the parameters of the model $\lambda = (Z, V, \pi, A, E)$. Compute the probability of an observed output sequence $O = O_1, ..., O_L$.
Solution: *forward algorithm.*

Learning Problem. Given a set of output sequences $O^1, ..., O^L$ and the structure of the model $\lambda = (Z, V)$. Determine the transition matrix A, the observation matrix E and the initial probability vector π so that the probability for this HMM, producing $O^1, ..., O^L$, is the maximum value.
Solution: *Baum-Welch algorithm*[1].

The HMMs were utilized for classification. Therefore, one HMM λ_i, $i = 1, ..., n$ for each emotion class was trained using data of one class. The probabilities $P(O \mid \lambda_i)$ for an unclassified observation O and the i-th HMM were estimated with the forward algorithm. For numerical reasons, the logarithm of the probabilities was used in the computation because they could become very small. The maximum of the n achieved values $P(O \mid \lambda_i)$, one for each HMM, lead to the most likely class. This architecture is called a log-likelihood estimator.

2.2   Gaussian Mixture Models

A Gaussian mixture model (GMM) $g(f_1, ..., f_m)$ is a probabilistic model for estimation of probability density functions [3]. It combines m Gaussian distributions $f_1, ..., f_m$. Each distribution f_i is defined as $f_i = (\mu_i, \Sigma_i, \alpha_i)$. The value μ_i is the expected value and thus the center of the normal distribution. The value Σ_i defines the covariance matrix of the i-th distribution. The third value α_i stands for the weight of the i-th distribution in the probability density function with $\sum_{i=1}^{m} \alpha_i = 1$.

The individual distributions f_i are assumed to be stochastically independent and defined as followed:

$$f_i(x \mid \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^{N/2} \mid \Sigma_i \mid}} exp\Big(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\Big) \ . \qquad (2)$$

This defines the total probability $P(X \mid f_1, ..., f_m)$ with:

$$P(X \mid f_1, ..., f_m) = \sum_{i=1}^{m} \alpha_i f_i(X, \mu_i, \Sigma_i) \ . \qquad (3)$$

In this study, one GMM in each state of the HMM was utilized to define the observation probabilities E, which were mentioned in Sect. 2.1. It should be

---

[1] The Baum-Welch algorithm is an instance of the expectation-maximization algorithm.

noted that GMMs cannot take temporal dependencies into account as HMMs do. The parameters of the GMM are trained with an expectation maximization algorithm.

2.3   Classifier Fusion

Combining classifiers is a promising approach to improve classifier results. A team of classifiers, which is intended to be fused, needs to be accurate and diverse [11]. While accuracy of a classifier is clear, diversity means that if a sample is classified falsely, not all classifiers should agree on a wrong class-label. In this paper different feature views on the data are constructed to produce diversity and two very simple but intuitive fusion techniques were evaluated.

Vote-Fusion. The results of the log-likelihood estimators lead at best all to the same class, but normally they are different. To nullify this problem, the first fusion method vote the results of selected log-likelihood estimators.

Probability-Fusion. The other fusion approach did not combine the results but the probabilities $P(O \mid \lambda_i)$ of selected log-likelihood estimators. To combine several estimators, the class-wise sums are computed. A summation of the logarithmic values is equivalent to a multiplication of the probabilities. Only then is the maximum determined to obtain the most likely class. This implies statistic independence of the models, which is unfortunately not fully given since the features were generated from the same sequence.

3   Data Collection

The Cohn-Kanade dataset is a collection of image sequences with emotional content [5], which is available for research purposes. It contains image sequences which were recorded in a resolution of 640×480 (sometimes 490) pixels with a temporal resolution of 33 frames per second. Every sequence is played by an amateur actor who is filmed from a frontal view. The sequences always start with a neutral facial expression and end with the full blown emotion which is one of the six categories "fear", "joy", "sadness", "disgust", "surprise" or "anger".

To acquire a suitable label, the sequences were presented to 15 human labelers (13 male and two female). The sequences were presented as a video. After the play-back of a video the last image remained on the screen and the test person was asked to select a label. Thus, a label for every sequence was created as the majority vote of the 15 different opinions. The result of the labeling procedure is given in Table 1 showing the confusion matrix of the test persons according to the majority of all persons. The resulting data collection showed to be highly imbalanced: the class "joy" (105 samples) occurred four times more often than the class "fear" (25 samples) and in addition, this expression could not be identified by the test persons.

In all automatic facial expression recognition systems some relevant features are extracted from the facial image first and these feature vectors are utilized to train some type of classifier to recognize the facial expression. One problem is here how to categorize the emotions: one way is to model emotions through

Table 1. Confusion matrix of the human test persons against the majority of all 15 votes (left). The right column shows the share of the facial expressions in the data set (hardened class labels).

maj.\test pers.	joy	ang.	sur.	disg.	sad.	fear	no. samples
joy	**104**	0	0	0	0	1	105
ang.	0	**39**	0	6	3	1	49
sur.	1	0	**72**	0	1	17	91
disg.	1	12	1	**54**	1	12	81
sad.	0	6	2	2	**70**	1	81
fear	1	1	3	6	1	**13**	25

a finite set of emotional classes such as anger, joy, sadness, etc, another way is to model emotions using continuous scales such as valence (the pleasantness of the emotion) and arousal (the level of activity) of an expression [12]. In this paper, a discrete representation in six emotions was used. Finding the most relevant features is definitely the most important step in designing a recognition system. In our approach, prominent facial regions such as the eyes, including the eyebrows, the mouth and for comparison the full facial region have been considered. For these four regions orientation histograms, principal components and optical flow features have been computed. Principal components (eigenfaces approach) are very well known in face recognition [19] and orientation histograms were successfully applied for the recognition of hand gestures [10] and faces [18], both on single images. In order to extract the facial motion in these regions, optical flow[2] features from pairs of consecutive images have been computed, as suggested in [17].

4   Experiments and Results

In this section, experiments which concern both the number of the distributions and the architecture of the HMMs are presented. The results of the achieved single models will be shown and the results of the two different fusion methods will be demonstrated.

4.1   Adjusting the Number of Gaussian-Components

We started with simple models with only one state and mixture. By doing so, a model is constructed which neglects any sequential dependencies. Thus, the transition matrix was just a vector. The results of the twelve feature-region pairs with this plain model are shown in Table 2. The next step was to improve the results by optimizing the models.

Numerical test runs were conducted with 10-fold cross-validation. The number of states were evaluated from 1 to 9 and the number of mixture components

---

[2] We were using a biologically inspired optical flow estimator which was developed by the Vision and Perception Science Lab of the Institute of Neural Processing at the University of Ulm [1,2].

from 1 to 4. The transition matrices were fully connected. The results of this process are shown in Table 2. The adjustments on the number of distributions led to higher detection rates than they were in the plain model. The experiments showed that by increasing the number of states, the corresponding number of mixture components per state became smaller. We determined that a total number of 8-20 for all distributions is sufficient for this application. One negative effect of a high number of distributions was, that some emotions were recognized very poorly: since the more states and mixtures were chosen, the more parameters had to be estimated. There were a few emotions with only little data (see Table 1), so the fine-tuning could not be adjusted well enough and it led to overfitting. Additionally, there was no model which was optimal for all emotions. The emotion "joy" had the best recognition rate with PCA-features in the region face, whereas the emotion "anger" was best recognized in the region of the left eye and features with optical flow estimation. The best results for the different emotions and the corresponding models are shown in Table 3.

Table 2. Results for expression concerning the number of states, the corresponding number of mixtures per state and the detection rate for all twelve models. In case of more than one state, the transition matrix was fully connected.

no.	feature	region	no. st/mix	det. rate	no. st/mix	det. rate
1	PCA	face	1 / 1	0.727	9 / 2	0.742
2	PCA	mouth	1 / 1	0.641	8 / 2	0.671
3	PCA	right eye	1 / 1	0.421	7 / 2	0.444
4	PCA	left eye	1 / 1	0.472	3 / 2	0.479
5	Orientation histograms	face	1 / 1	0.678	4 / 2	0.710
6	Orientation histograms	mouth	1 / 1	0.678	4 / 3	0.732
7	Orientation histograms	right eye	1 / 1	0.440	4 / 2	0.473
8	Orientation histograms	left eye	1 / 1	0.417	9 / 2	0.475
9	Optical flow	face	1 / 1	0.627	8 / 2	0.638
10	Optical flow	mouth	1 / 1	0.639	9 / 2	0.646
11	Optical flow	right eye	1 / 1	0.437	7 / 3	0.471
12	Optical flow	left eye	1 / 1	0.449	8 / 4	0.491

Table 3. Best recognition rate and corresponding model for each emotion

emotion	det. rate	feature	region	no. st/mix
joy	0.981	PCA	face	9 / 2
anger	0.816	Optical flow	left eye	8 / 4
surprise	0.978	Orientation histograms	mouth	4 / 3
disgust	0.593	Optical flow	face	8 / 2
sadness	0.889	Orientation histograms	mouth	4 / 3
fear	0.200	PCA	mouth	8 / 2

4.2   Investigating Temporal Dependence

In this subsection, the temporal properties of the learned models and the underlying data is studied. To study the influence of the topology of the HMM to the classifier, fully connected models were compared with forward models (see Sect. 2.1). It was discovered that the transition matrix of the fully connected model approximated the transition matrix of the forward model during training. Since the detection rate of the forward model was higher, we concluded that the forward model can model the time dependence better in this case. By using the forward model, also computation time can be saved because not all possible entries in the transition matrix have to be adapted. Another advantage of the forward model is the avoidance of local minima. The fully connected model has a higher risk to descend toward a local minimum [8].

The expectation values of the GMMs (see Sect. 2.2) form the centers of the normal Gaussian distributions. For initialization, the k-means procedure was used to ensure that the centers cover the entire feature space. To investigate whether the centers are different from each other after the training, the pairwise distances of the expected values were calculated. If the observed probabilities in the states are equal, using HMMs does not provide any benefit. Figure 1 shows the differences of an example with four states and two mixtures per state as grayscale values. It can be observed, that both, the distances between centers in different states as well as centers within a state, differ. Thus, there is a corresponding difference between the data assigned to the respective states. This can be demonstrated more descriptive by producing average images of each state. Figure 2 shows average images of the emotion "joy" made from 103 sequences. The images were assigned to the states using the Viterbi algorithm.
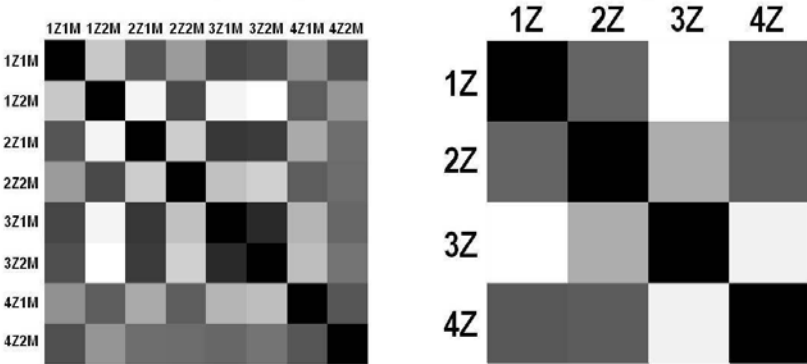

Fig. 1. Pairwise distances of the centers of the Gaussian distributions as grayscale values: the brighter the grayscale value, the greater the distance. On the diagonal are the distances to itself (no distance = black). (iZjM designates the i-th mixture M of the j-th state Z).
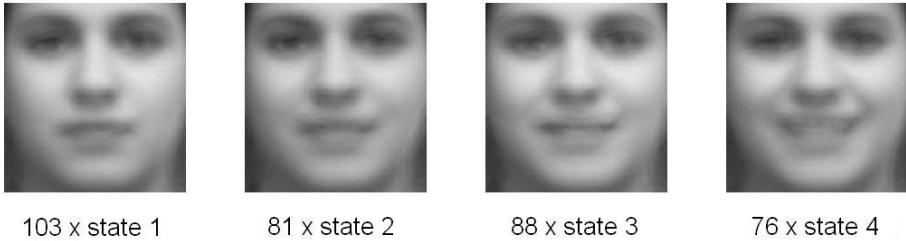
103 x state 1   81 x state 2   88 x state 3   76 x state 4

Fig. 2. Average pictures for each state. A four state model of the emotion "joy" was utilized and the paths of 103 sequences were computed.

4.3 Results of the Single Models

Based on the considerations above, the individual models were developed. Forward models with the optimized configurations for the number of states and mixture components were used (see Table 2). The best single model with a total detection rate of 76.4 % was the model with PCA-features in the region face. Table 4 shows the confusion matrix of this model. The emotions "joy", "anger" and "disgust" were often confused. This occured particularly in the eye-regions. One reason for this could be that for these three emotions either the eyes have to be pinched together or the cheek muscles goes up because of a smile. However, with "surprise", "sadness" and "fear" the eyes are open. The emotion "fear" was often confused with the emotion "surprise", in which the greatest confusion occurs in the region mouth. The open mouth of the two emotions might play a role here.

Table 4. Confusion matrix of the best single model: features with PCA of the region face. The columns stand for the labeled emotion, the rows for the detected emotions. The recognition rate of each single emotion is shown in the last column.

det.emo\emo.	joy	ang.	sur.	disg.	sad.	fear
joy	**102**	4	0	11	2	5
ang.	0	**26**	0	8	6	0
sur.	0	0	**82**	2	3	7
disg.	3	9	3	**47**	1	3
sad.	0	10	5	8	**68**	5
fear	0	0	1	5	1	**5**
det.rate	0.981	0.531	0.901	0.580	0.840	0.200

4.4 Results of the Fused Models

To improve the total detection rate even further, the results of the twelve individual models were fused. For a better overview, the individual models were coded according to the number of the rows of Table 2. For instance, model 1

Table 5. Top ten combinations with vote-fusion and probability-fusion of the developed twelve single models. The coding of the models corresponds to the first column in Table 2.

vote-fusion		probability-fusion	
no. of models	det. rate	no. of models	det. rate
1 2 5 6 9 10	0.817	1 2 5 6 7 9 10 11	0.861
1 2 5 6 9	0.815	1 2 3 5 6 9 10 12	0.859
1 2 6 8 9 10	0.815	1 2 3 5 6 9 10 12	0.859
1 2 6 9 10 11	0.815	1 2 5 6 7 9 10 12	0.859
1 2 4 5 6 9 10	0.815	1 5 6 9	0.857
1 2 6 7 9 10	0.813	2 5 6 9 10 12	0.857
1 5 6 9 10 12	0.813	1 2 4 5 6 7 9	0.857
1 2 6 9 12	0.810	1 2 4 5 6 9 11	0.857
1 2 5 6 8 9 10	0.810	1 5 6 9 10	0.854
1 2 5 6 10 11 12	0.810	1 2 4 5 6 10 12	0.854

recognized the emotion "joy" with 97.1 %, while it reached less than 25 % for the emotion "anger". On the other hand, model 12 showed different results: for emotion "anger" a detection rate of 63.3 % was obtained, but the detection rate for "joy" was lower than with model 1. To combine the advantages of these models and neutralize the disadvantages as possible, the decisions of the models were fused. In the following, the results of the two fusion approaches will be presented:

Vote-Fusion. A simple fusion approach is to vote the detected emotions of all twelve classifiers. If a test sequence is classified incorrectly from one model, but the other eleven models detect it correctly, it will be outvoted. After the training of the individual classifiers, we investigated whether the classification rate could be improved, if not all twelve models would be combined. Therefore, all 2^{12} model combinations were computed. The results of the top ten combinations are shown in Table 5. The appearance of the models 1 and 6 in all displayed combinations is originated from their good individual recognition performance. Models using the eye regions occur only sporadically in this list. This shows that these models could reduce the classification rate because of their minor accuracy. The combination of all face and mouth regions led to a detection rate of 81.7 %. This means that, compared to the best single model, this combination has the ability to recognize 23 sequences more.

Probability-Fusion. This approach, as mentioned in Sect. 2.3, combines the logarithmic probabilities which are obtained from the six HMMs of each log-likelihood estimator. In many cases, when a sample was misclassified, the probability of the incorrect emotion was just slightly higher than the one of the correct emotion. The class-wise multiplication of the posterior probabilities has the potential to overcome the limitations of the early hardening in

the vote-fusion. The results are also shown in Table 5. Again, the face and mouth regions played an important role. The result, showing the highest classification rate of 86.1 % (372 out of 432), consisted of all models using the features of the regions face and mouth and the models 7 and 11. Table 6 shows the confusion matrix. This was the highest achieved recognition rate in this study. Thus, the probability-fusion could recognize 19 sequences more than the vote-fusion.

Table 6. Confusion matrix of the best result, obtained with probability-fusion: combination of all models with features of the regions face and mouth and the two models with orientation histograms and optical flow features in the right eye region. Total detection rate: 86.1 %.

det.emo\emo.	joy	ang.	sur.	disg.	sad.	fear
joy	**105**	0	0	11	2	4
ang.	0	**36**	0	3	2	0
sur.	0	1	**89**	1	0	8
disg.	0	7	1	**65**	1	10
sad.	0	5	1	1	**76**	2
fear	0	0	0	0	0	**1**
det. rate	1.000	0.735	0.978	0.803	0.938	0.040

5   Summary

The aim of this study was to investigate the behavior of HMMs in facial expression recognition and to demonstrate whether HMMs are capable of recognizing emotions in image sequences sufficiently well.

432 image sequences from the Cohn-Kanade database were used as dataset. Four different regions in each image were extracted (face, mouth, right and left eye) and three different methods for feature extraction were applied (PCA, orientation histograms and optical flow estimation). The data of each combination of region and feature was utilized to train a log-likelihood estimator which consisted of six HMMs (one for each emotion). The emotions were: "joy", "anger", "surprise", "disgust", "sadness" and "fear".

Not only the detection performance but also the architecture and the learning behavior was investigated. First, experiments with different number of states and mixtures per state were performed. It was observed that by increasing the number of states, the number of Gaussian distributions had to be reduced to achieve a good classification performance. As a result of this investigation, twelve single models for the feature-region pairs were identified. Also, experiments with the transition matrix, i.e. the structure of the HMMs, were conducted. We discovered that the forward models were slightly better than the fully connected models. The best single model with a detection rate of 76.9 % was the forward model with PCA-features in the region face (see Table 4). This model classified 332 out of 432 sequences correctly.

Furthermore, the outputs of the individual models were fused to combine the advantages of the models and possibly neutralize the shortcomings. Two fusion approaches were evaluated: vote-fusion and probability-fusion. These fusion approaches showed that the classification rates can be improved further by combining the results of the different models. The best fusion result with a classification rate of 86.1 % is shown in Sect. 4.4, including a confusion matrix in Table 6. This was the highest achieved recognition rate with 372 out of 432 detected sequences.

The experiments and results show that HMMs are very well capable of emotion recognition in image sequences. Time dependencies can simply be modeled by HMMs and the experiments also show that they play an important role.

Acknowledgments

References

1. Bayerl, P., Neumann, H.: Disambiguating visual motion through contextual feedback modulation. Neural Computation 16, 2041–2066 (2004)
2. Bayerl, P., Neumann, H.: A fast biologically inspired algorithm for recurrent motion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 246–260 (2007)
3. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2007)
4. Cohen, I., Garg, A., Huang, T.S.: Emotion recognition from facial expressions using multilevel HMM. In: Neural Information Processing Systems (2000)
5. Cohn, J.F., Kanade, T., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53 (2000)
6. Darwin, C.: The Expression of the Emotions in Man and Animals, 1st edn. Oxford University Press Inc., New York (1872)
7. Lin, D.T., Chen, J.: Facial expressions classification with hierarchical radial basis function networks. In: Proceedings of the 6th International Conference on Neural Information Processing, ICONIP, pp. 1202–1207 (1999)
8. Durbin, R., Eddy, S., Krogh, A., Mitchison, G.: Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge (1998)

9. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
10. Freeman, W.T., Roth, M.: Orientation histograms for hand gesture recognition. In: International Workshop on Automatic Face and Gesture Recognition, pp. 296–301 (1994)
11. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach. Learn. 51(2), 181–207 (2003)
12. Lang, P.J.: The emotion probe. studies of motivation and attention. The American psychologist 50(5), 372–385 (1995)
13. Lien, J.J.J., Kanade, T., Cohn, J., Li, C.: A multi-method approach for discriminating between similar facial expressions, including expression intensity estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1998) (June 1998)
14. Lisetti, C.L., Rumelhart, D.E.: Facial expression recognition using a neural network. In: Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference, pp. 328–332. AAAI Press, Menlo Park (1998)
15. Rabiner, L., Juang, B.H.: Fundamentals of Speech Recognition. Prentice Hall PTR, Englewood Cliffs (1993)
16. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. In: Proceedings of the IEEE, pp. 257–286 (1989)
17. Rosenblum, M., Yacoob, Y., Davis, L.: Human expression recognition from motion using a radial basis function network architecture. IEEE Transactions on Neural Networks 7(5), 1121–1138 (1996)
18. Schwenker, F., Sachs, A., Palm, G., Kestler, H.A.: Orientation histograms for face recognition. In: Schwenker, F., Marinai, S. (eds.) ANNPR 2006. LNCS (LNAI), vol. 4087, pp. 253–259. Springer, Heidelberg (2006)
19. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
20. Yeasin, M., Bullot, B., Sharma, R.: From facial expression to level of interest: A spatio-temporal approach. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 922–927 (2004)