in: Proc. 8th IEEE Int. Conf. on Face and Gesture Recognition (FG'08), Amsterdam, paper no. 299, 6 pages, IEEE

A Real-time Facial Expression Recognition System based on Active Appearance Models using Gray Images and Edge Images [*]

Christian Martin
MetraLabs GmbH, Germany
christian.martin@tu-ilmenau.de

Uwe Werner, Horst-Michael Gross
Neuroinformatics and Cognitive Robotics Lab
Ilmenau Technical University, Germany

Abstract

In this paper, we present an approach for facial expression classification, based on Active Appearance Models. To be able to work in real-world, we applied the AAM framework on edge images, instead of gray images. This yields to more robustness against varying lighting conditions. Additionally, three different facial expression classifiers (AAM classifier set, MLP and SVM) are compared with each other. An essential advantage of the developed system is, that it is able to work in real-time - a prerequisite for the envisaged implementation on an interactive social robot. The real-time capability was achieved by a two-stage hierarchical AAM tracker and a very efficient implementation.

1. Introduction

In the last years, a growing number of applications appeared, which require detailed information about human faces from video data streams. Examples are interactive mobile service robots or other man-machine-systems, whose dialog are to be adapted to the current emotional state of the interaction partner. For that purpose, amongst other describing features, the interpretation of the facial expression of the user is necessary.

In [12] could be shown, that the *Active Appearance Model* (AAM) framework, which was originally introduced by [3], leads to better results in the field of facial expression analysis in comparison to *ICA* or *Graph Matching* techniques. A problem of their work was however the necessary computation time. In this paper, we present an advanced facial expression classification system, which also relies on the Active Appearance Model framework but is also able to work in real-time. In contrast to other systems, which mostly rely on gray images, our approach uses edges images. This helps us to overcome a major problem of many existing systems: different illumination conditions. Meanwhile, there exists a broad spectrum of facial expression recognition techniques, e.g. ICA, Graph Matching, or facial landmarks tracking, but in this paper, we only focus on the AAM framework because of the conceptual clearness and the advantages of this approach.

The paper is organized as follows: After the description of related work, the *Active Appearance Model* framework is briefly introduced in section 2. In section 3, the different analyzed facial expression classificators are presented. Finally, in section 4 the achieved results are shown. The paper ends with a short summary and conclusion in section 5.

1.1. Related work

In the last years, a number of approaches which use the Active Appearance Model framework for facial expression recognition, were presented.

In [8] Active Appearance Models are used to classify four basic emotions (happiness, sadness, anger and neutral) by means of a cascade of four *Support Vector Machines*. Overall, a classification rate between 64% and 94% was achieved.

In [9] only the appearance parameters of an AAM are used to classify the six basic emotions. The classification is done by means of a 3-layered feed-forward MLP with 94 input neurons and 7 outputs. This system reaches a classification rate between 85% and 97% on the test data set. A problem of this approach seems to be, that is might be very hard to estimate 94 appearance parameters robustly. Here, a modern information-theoretic feature selection technique could help thinning out the high dimensional input space.

In the work of Ratliff and Patterson [6], the FEEDTUM mimic database [11] was used. They employed an Active Appearance Model with 113 landmarks. For each of the six basic emotions a mean parameter vector was computed and the classification was done by means of a simple *Nearest Neighbour Classifier*. In their work, a classification rate between 63% and 93% for the different basic emotions was reached.

[*] The research leading to these results has received partial funding from the State Thuringia (TAB-Grant #2006-FE-0154) and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216487 (CompanionAble-Project).

978-1-4244-2154-1/08/$25.00 ©2008 IE

In [12] we compared different recognition techniques (AAM, ICA, Graph Matching) and different classifiers (Nearest Neighbour, MLP, RBF, LVQ). For facial expression recognition, the best results (a true positive rate of 72%) could be reached by means of a standard AAM approach in combination with a 3-layered MLP classifier.

2. Active Appearance Models

In the following section, the basic concepts of the *Active Appearance Model* framework and the extensions of our approach are described. For a detailed mathematical description, please refer to [1] and [2].

2.1. Building the Shape Model

In our work, we use a 2-dimensional shape model S consisting of $v = 58$ points. The points are placed in regions of the face, which typically have a lot of texture information. Each instance $\mathbf{s}$ of the shape model can be described as a vector consisting of $2v$ elements:

$$\mathbf{s} = (x_1, y_1, ..., x_v, y_v)^T \tag{1}$$

In a pre-processing step, all training shapes t_i are aligned by applying the *Generalized Orthogonal Procrustes Analysis* [7]. This algorithm removes all components from the data set, which are caused by scaling, translation and rotation. Furthermore, for a mimic recognition system, it is very useful to generate additional shapes by mirroring the training shapes horizontally. This leads to a new training data set $\mathbf{t}'$ of $N' = 2N$ examples.

Based on the training data set, the mean shape $\mathbf{s}_0$ can be computed as the mean of all N' training examples.

$$\mathbf{s}_0 = \frac{1}{N'} \sum_{i=1}^{N'} \mathbf{t}'_i \tag{2}$$

The main components of all shapes in the training data set can be computed by a *Principle Component Analysis (PCA)*. The m components with the m biggest eigenvalues will be selected as the shape components $\mathbf{s}_1$ to $\mathbf{s}_m$. Now it is possible to reconstruct the shapes of the training data set and also to generate new shapes, which are not part of the data set by means of the basic shape $\mathbf{s}_0$ and a linear combination of the components $\mathbf{s}_i$.

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i \mathbf{s}_i \tag{3}$$

The quality of the reconstruction of the shapes from the training data set depends on the number m of used components. The capability of generating new shapes, which are not represented in the training data set, however, depends on the diversity of the training data set.

Depending on the training data set, the main components $\mathbf{s}_i$ typically represent global variations of the face (like pitch and yaw), which are mostly invariant to facial mimics and also local changes (like opening and closing of the eyes or mouth), which are involved in mimics of the subjects. Figure 1 shows the basic shape $\mathbf{s}_0$.

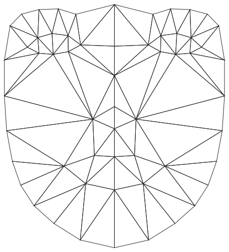

Figure 1. The basic shape $\mathbf{s}_0$ (consisting of $v = 58$ points) and the used triangulation.

2.2. Warp definition

The *Warp* $W(\mathbf{x}, \mathbf{p})$ defines a piecewise affine transformation between two shapes. In this work, the warp is used to transform an input image $I(\mathbf{x})$ with shape $\mathbf{s}$ and the parameters $\mathbf{p}$ (see equation (3)) to the basic shape $\mathbf{s}_0$. In this case, the transformation function is $x = W(x_0, \mathbf{p})$, where x_0 is a pixel in the basic shape and x is a pixel in the target image. Therefore, using the right shape parameters $\mathbf{p}$, the image $I(W(x_0, \mathbf{p}))$ defines the input image transformed on the basic shape $\mathbf{s}_0$.

2.3. Building the Appearance Model

Besides the *Shape Model*, the second important part of an AAM is the *Appearance Model*, which uses a transformation of the high dimensional input images to a linear subspace of *Eigenfaces*. This results in a drastic reduction of the dimension of the parameter space.

As a pre-processing step all input images $I(\mathbf{x})$ are filtered by a *Gauss Filter*, to remove the image noise. By means of the piecewise affine transformation $W(\mathbf{x}, \mathbf{p})$ the input image will be transformed to the basic shape $\mathbf{s}_0$ to $I(W(\mathbf{x}, \mathbf{p}))$. On this normalized images, a histogramm equalization is applied to reduce the lighting influences. On the input images, normalized this way, a PCA is applied. The k components with the largest eigenvalues are selected as the appearance components $A_1(\mathbf{x})$ to $A_k(\mathbf{x})$. The mean appearance components $A_0(\mathbf{x})$ can be computed by a simple mean of all normalized input images:

$$A_0(\mathbf{x}) = \frac{1}{N'} \sum_{i=1}^{N'} I(W(\mathbf{x}, \mathbf{p})) \tag{4}$$

In our work, for comparison purposes, we used two different types of appearance models. The first model is a standard *Gray Image Model* as introduced in [1], [2] and [3].

The second type of appearance models used in this work employes edges information instead of gray values: The import features for mimic recognition like laugh lines or furrows in the brow are typically better extractable from edges images, than from gray value images. Furthermore, edge images are normally more robust to varying lighting conditions. For this type of appearance model, instead of the histogramm equalization, an edge filter is applied before the PCA is computed. The normalized input image $I(W(\mathbf{x},\mathbf{p}))$ is convolved by the filters G_x and G_y:

$$S_x = I(W(\mathbf{x},\mathbf{p})) * G_x, \quad S_y = I(W(\mathbf{x},\mathbf{p})) * G_y \qquad (5)$$

The gradient value matrix S is computed based on the absolute values of the filter results S_x and S_y:

$$S = \sqrt{S_x^2 + S_y^2} \qquad (6)$$

Figure 2 illustrates the functioning of this filter method.

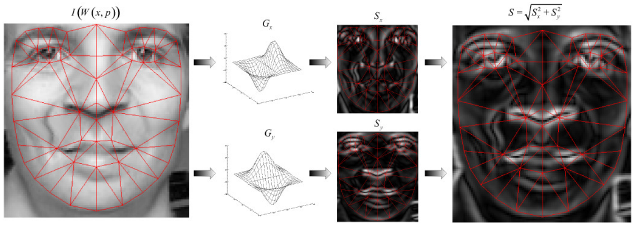

Figure 2. The edge filter: The edges filters G_x (x-direction) and G_y (y-direction) are applied to the normalized input image $I(W(\mathbf{x},\mathbf{p}))$. The gradient value S is computed on the results S_x and S_y of both filters.

For both types of model, based on the appearance components $A_i(\mathbf{x})$, now it is possible to generate an image $A(\mathbf{x})$ with the basic shape $\mathbf{s}_0$, as follows:

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) \qquad (7)$$

2.4. Model instances

Combining the shape model and the appearance models leads to a model instance. The instance $M(W(\mathbf{x},\mathbf{p}))$ describes a combination of an appearance model and its shape. For that, the appearance parameters $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)$ and the shape parameters $\mathbf{p} = (p_1, ..., p_n)$ are necessary. Using equation (7) the image $A(\mathbf{x})$ in the form of the basic shape $\mathbf{s}_0$ can be computed. After that, the image $A(\mathbf{x})$ can be transformed to the shape $\mathbf{s}$ by using the warp $W(\mathbf{x},\mathbf{p})$.

2.5. Model Adaptation

In the case for mimics recognition based on Active Appearance Models, it is necessary to adapt the trained model to an unknown input image $I(\mathbf{x})$. That means, that we have to find the optimal parameters $\mathbf{p}$ and $\boldsymbol{\lambda}$.

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \sum_{\mathbf{x}\in\mathbf{s}_0} \left[A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x},\mathbf{p})) \right]^2 \qquad (8)$$

For this optimization problem, the following error function $E(\mathbf{x})$ can be defined:

$$E(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x},\mathbf{p})). \qquad (9)$$

To solve this problem, in our work we use the a variant of the *Inverse Compositional Algorithm*, which was introduced in [1] and [2]. The problem of the original form of the adaptation algorithm is, that the appearance parameters λ_i are not part of the optimization. The Inverse Compositional Algorithm uses a projection algorithm, that was originally introduced in [5], which allow the optimization of the shape parameters $\mathbf{p}$ and the appearance parameters $\boldsymbol{\lambda}$ simultaneously. For more details, please refer to [1] and [2].

3. System Architecture

The developed mimic recognition system consists of four subsystems (see Figure 3). The first part is a face detector developed by Viola and Jones [10]. This detector is able to detect faces in real-time.

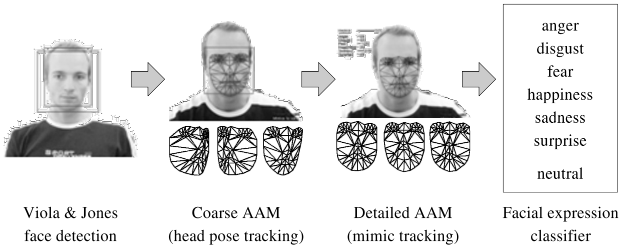

Figure 3. The main components of the developed facial expression classification system. The output of the face detector is used to initialize a *Coarse AAM*, which tracks the head pose. The *Detailed AAM* tracks the details of the face and generates the input for the facial expression classifier.

The position and size of the detected face is used to initialize a *Coarse AAM* for tracking the face in the image. Besides the trained shape components (sec section 2.1), we use four additional synthetic shape components:

$$\begin{aligned} \mathbf{s}_1^* &= \left(x_1^0, y_1^0, \ldots, x_v^0, y_v^0\right)^T \\ \mathbf{s}_2^* &= \left(-y_1^0, x_1^0, \ldots, -y_v^0, x_v^0\right)^T \\ \mathbf{s}_3^* &= (1, 0, \ldots, 1, 0)^T \\ \mathbf{s}_4^* &= (0, 1, \ldots, 0, 1)^T \end{aligned} \qquad (10)$$

These components describe the scaling ($\mathbf{s}_1^*$) of the shape, an approximation of the in-plane rotation ($\mathbf{s}_2^*$) and the translation on the x-axis and y-axis ($\mathbf{s}_3^*$, $\mathbf{s}_4^*$). The corresponding shape parameters p_i to these shape components, can be initialized from the output of the Viola and Jones face detector. In total, the Coarse AAM uses $n = 4 + 2$ shape components (four synthetic and the first two of the trained model) and $m = 6$ appearance components. This is sufficient for tracking the head and a coarse appearance estimation.

As soon as the error value $E(\mathbf{x})$ (see Eqn.(8)) for the Coarse AAM drops below a certain threshold, the more detailed AAM is initialized. This model consists of $n = 4+12$ shape parameters and $m = 16$ appearance components. The parameter values can be initialized from the parameter values of the Coarse AAM. This Detailed AAM is able to fit quiet good to the face in the input image. At time step $t+1$, the Detailed AAM is re-initialized by the known parameters of time t. If after some adaptation iterations, the error value exceeds a certain threshold, the system falls back to the Coarse AAM. If this also fails, the face detector is used to find a new hypothesis.

While the simple model is only able to do a coarse estimation of the input image, the detailed model is also able to estimate these details of the face, which are necessary for a mimic recognition. For the classification of the facial expression, we analyzed three different approaches:

- **AAM classifier set:** In this case, we generated particular models for each of the six basic emotions. Each model was trained by using the corresponding subset for the different basic emotions of the training data set. All models were applied simultaneously to the input image, the model with the lowest error value $E(\mathbf{x})$ is selected as winner.

- **MLP-based classifier:** For this classifier system, the estimated parameters p_i are used as input for a *Multi-Layer-Perceptron* (MLP). The MLP has two hidden layers with 15 respective 7 neurons. The output layer consists of 7 neurons (*Neutral* + 6 basic emotions). The MLP uses the $tanh$ activation function was trained with the standard *Backpropagation* algorithm.

- **SVM-based classifier:** The last classifier uses a standard *Support-Vector-Machine* with Gauss kernels for the mimics classification. This SVM also uses the parameters p_i as the input and has 7 output nodes, like the MLP. In the hidden layer, the SVM generates high dimensional hyperplanes, which should separate the features distribution of the six basic emotions.

4. Experimental Results

First, the model adaptation accuracy between the different models is shown in section 4.1. Section 4.2 presents the results of the facial expression classification. Finally, the overall performace of our system is shown in section 4.3.

4.1. Model Adaptation Accuracy

First, we compared how good the different Active Appearance Model types (Gray Image vs. Edge Image) are able to adapt to a given input image. To this purpose, the medium Euclidian distance between the $v = 58$ labeled data points of the input image and the data points of the generated output shape of the AAM are measured in pixels. Figure 4 shows the accuracy of both models. It is visible, that the Edge Image Model is able to fit the input image better than the Gray Image Model. Since a good adaptation accuracy is a prerequisite of a good facial expression classification, the subsequent facial expression classification was only tested for the Edge Image Model.

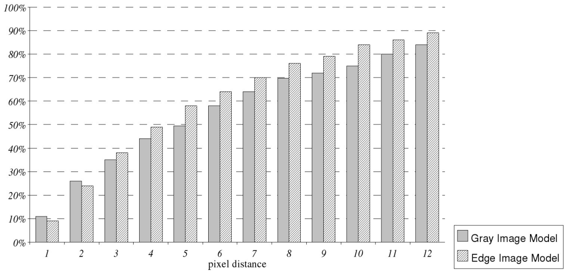

Figure 4. AAM model adaptation accuracy: For the two different AAM types (Gray Image Model and Edge Image Model), the percentage of the data is shown, whose medium data point distance is below a certain threshold. For example, the Gray Image Model is able to fit 73% while the Edge Image Model is able to fit 84% of the data with less than 10 pixels distance.

4.2. Facial Expression Classification

For our work, we used the FEEDTUM mimic database [11]. This database consists of 18 different persons (9 male and 9 female), each showing the six different basic emotions [4] (anger, disgust, fear, happiness, sadness, surprise and neutral) in a short video sequence. Figure 5 shows two typical examples of the over 50,000 images in the database. For this paper, we labeled a subset of 1,438 images by hand to build an Active Appearance Model. In this subset all 18 individuals are represented. In Table 1, the number of used images for each basic emotions is shown. Since we only selected each 10^{th} frame of the video sequences and the length of the sequences varies, the used number of images for the different emotions is not uniform.

For the training of the different classifier systems, this database was divided into three parts: The first part (60%, 862 images) was used as the training data set. Another 15% (216 images) are used as the validation data set for *cross validation* and the remaining 35% (360 images) are used as

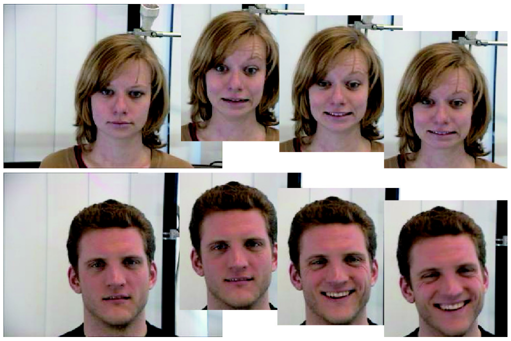

Figure 5. Examples from the FEEDTUM [11] database. Each data set consists of a short video sequence, where the mimic of the subjects changes from *Neutral* to one of the six basic emotions.

Ang.	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	*Neu.*
197	261	247	193	296	177	67

Table 1. Number of images of the different basic emotions in the used subset of the FEEDTUM data base. The numbers vary, due to the different length of the video sequences.

the test data set. For each of the images, the correct facial expression class is known from the database. This information was used as the ground truth for the different classifiers.

The used number of 1,438 images is not very large compared to other approaches. Although, our results still can be compared to other approaches, which use the whole FEEDTUM database, since we only selected each 10^{th} frame but the skipped frames are typically very similar to the selected one, due to the small time difference in the video.

4.2.1 Results of the AAM classifier

In the case of the AAM classifier set, six different models where trained with the data from the respective basic emotion. So in theory, this should result in six different models, where each of them could match very good to the corresponding basic emotion. For the testing, each model was applied to the input image and the Euclidian distance from the input image to the generated image was computed. The model with the lowest distance was selected as winner.

Table 2 shows the results of this approach. It is visible, that the basic emotion *Anger* could be matched very precisely (94,9%) from the corresponding model. But the results of the other expressions are very low (between 10% and 30%). Also the false positive rate is very high.

4.2.2 Classification based on only shape parameters

In real-world it is sometimes very hard to a match an Active Appearance Model to an input image with a big number of shape and appearance parameters. Therefore, we tried to find out, how good a facial expression classification could be done with a very simple model. For this test, we used an AAM which only consists of $n = 10$ shape parameters and no appearance parameters. Table 3 shows the results of the MLP and the SVM classifier for this simple model. It is clearly visible, that the SVM approach has a considerable higher detection rate and also a lower false positive rate than the MLP approach.

	Ang.	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	FP
Ang.	186	128	89	100	172	72	47,9%
Dis.	0	28	11	1	0	0	1,1%
Fear	1	25	70	14	21	26	7,8%
Hap.	6	49	61	60	42	23	15,4%
Sad.	1	19	9	8	62	26	5,9%
Sur.	2	13	7	9	1	26	2,7%
sum	196	262	247	192	298	173	
%	94,9	10,7	28,3	31,3	20,8	15,0	

Table 2. Classification results of the AAM classifier set on the whole data set. The bottom row shows the detection rate and the right column shows the false positive rate (FP). This approach was able to classify the *Anger* emotion, but not the other basic emotions. Also the false positive rate is very high for some emotions.

MLP	*Ang.*	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	*Neu.*	FP
Ang.	14	1	0	1	0	0	1	0,9%
Dis.	6	38	7	1	6	3	13	11,8%
Fear	10	1	40	1	4	2	15	10,9%
Hap.	0	0	1	17	0	6	7	4,4%
Sad.	3	6	1	3	45	2	13	9,3%
Sur.	0	5	1	8	0	19	6	6,1%
Neu.	5	4	6	12	5	1	20	11,6%
sum	38	55	56	43	60	33	75	
%	36,8	69,1	71,4	39,5	75,0	57,6	26,7	

SVM	*Ang.*	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	*Neu.*	FP
Ang.	38	0	1	3	0	1	0	1,6%
Dis.	1	53	1	1	2	1	2	2,6%
Fear	0	0	52	0	0	1	0	0,3%
Hap.	2	1	0	34	2	3	0	2,6%
Sad.	2	0	4	5	76	0	6	6,2%
Sur.	4	3	2	5	5	31	1	6,2%
Neu.	0	0	2	0	0	1	14	0,9%
sum	47	57	62	48	85	38	23	
%	80,8	93,0	83,9	70,8	89,4	81,6	60,9	

Table 3. Results of the facial expression classification with only $n = 10$ shape parameters. The upper table shows the results achieved by the MLP classifier and the bottom table shows the results of the SVM classifier. The bottom row shows the detection rate and the right column shows the false positive rate (FP). The SVM approach outperformed the MLP approach in the detection rate and also in the false positive rate.

4.2.3 Classification based on shape and appearance

For a good facial expression classification, it is necessary to use also shape and appearance parameters as input for the classifier. Table 4 shows the results of the MLP and the SVM classifier based on $n = 10$ shape parameters and $m = 20$ appearance parameters. It is visible, that the additional appearance parameters clearly have improved the classification results. Besides the higher true positive rate, also the lower FP rate is visible. The best results could be achieved by the SVM classifier. On average, the SVM reaches a detection rate of 92%, while the MLP only reaches an average detection rate of 75%. A reason for this could be, that the high dimensional hyperplanes of the SVM are possibly able to separate the input space better than the MLP.

This results can directly compared to the results of the work of Ratliff and Patterson [6], since they used the same database. While their classification rate for the basic emotions varies between 63% and 93% percent (in average 82%), in our work we reached a detection rate always better than 90% (in average 92%) for the SVM.

MLP	*Ang.*	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	*Neu.*	FP
Ang.	38	1	0	3	1	0	0	1,6%
Dis.	3	46	0	1	1	3	3	3,7%
Fear	1	1	46	1	2	1	3	2,9%
Hap.	1	4	1	34	2	6	2	5,1%
Sad.	2	3	0	3	79	0	1	3,3%
Sur.	1	4	3	6	1	39	2	5,5%
Neu.	0	0	1	0	2	1	7	1,2%
sum	46	59	51	48	88	50	18	
%	82,6	78,0	90,2	70,8	89,8	78,0	38,9	

SVM	*Ang.*	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*	*Neu.*	FP
Ang.	41	0	0	0	1	1	0	0,6%
Dis.	0	66	1	5	0	2	0	2,7%
Fear	0	0	66	0	3	5	0	2,7%
Hap.	0	0	0	49	0	3	0	0,9%
Sad.	3	3	0	0	59	0	0	2,0%
Sur.	0	0	1	0	0	38	0	0,3%
Neu.	0	0	0	0	2	0	11	0,6%
sum	44	69	68	54	65	49	11	
%	93,2	95,7	97,1	90,7	90,8	77,6	100,0	

Table 4. Results of the facial expression classification with $n = 10$ shape components and $m = 20$ appearance parameters of the edge image model. The upper table shows the results achieved by the MLP classifier and the bottom table shows the results of the SVM classifier. The bottom row shows the detection rate and the right column shows the false positive rate (FP).

4.3. Real-time capabilities

A single iteration step of the AAM takes about $2ms$ in out implementation. During the computation in the different subsystems (see Sec. 3), a correct adaptation typically requires 5 to 12 interaction cycles for AAM parameter adaptation, which gives a total time of maximum $24ms$. The MLP or SVM classification takes less than $1ms$. Therefore, our system is able to work in real-time on a video sequence with 20 frames per second.

5. Summary and Conclusion

In this paper, we have presented a facial expression classification system based on Active Appearance Models. Besides the application of the AAM framework to gray value images, we have used the AAMs on edge images. It turned out, that AAMs on edge images are able to fit better on a given input image. Furthermore, we have compared three different systems for facial expression classification. The simple AAM classifier set only reached bad results. The MLP and the SVM classifiers reached reasonable good result, while the SVM classifier outperforms the MLP. The system presented in this work, was already successfully used on an interactive mobile service robot for an online facial expression classification of face images of people not included in the data base. This demonstrates the capability to re-use the trained models for unknown data. In the future, we want to benchmark our approach with different standard databases.

References

[1] S. Baker and I. Matthews. Equivalence and Efficiency of Image Alignment Algorithms. In *Proc. of the 2001 IEEE Conf. on CVPR*, volume 1, pages 1090–1097, 2001.

[2] S. Baker and I. Matthews. Lucas-Kanade 20 Years On: A Unifying Framework. *Int. Journal of Computer Vision*, 56(3):221–255, 2004.

[3] T. Cootes, G. Edwards, and C. Taylor. Active Appearance Models. In *Proc. ECCV*, volume 2, pages 484–498, 1998.

[4] P. Ekman and W. Friesen. *Unmasking the face. A guide to recognizing emotions from facial clues*. Prentice Hall, 1975.

[5] G. D. Hager and P. N. Belhumeur. Efficient Region Tracking With Parametric Models of Geometry and Illumination. *IEEE Trans. PAMI*, 20(10):1025–1039, 1998.

[6] M. S. Ratliff and E. Patterson. Emotion Recognition using Facial Expressions with Active Appearance Models. In *Proc. of HRI*, 2008.

[7] A. Ross. Procrustes analysis, 2005. http://www.cse.sc.edu.

[8] Y. Saatci and C. Town. Cascaded Classification of Gender and Facial Expression using Active Appearance Models. In *Proc. of the 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 393–400, 2006.

[9] H. van Kuilenburg, M. Wiering, and M. den Uyl. A Model Based Method for Automatic Facial Expression Recognition. In *Proc. of the ECML*, 2005.

[10] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of CVPR*, pages 511–518, 2001.

[11] F. Wallhoff. FEEDTUM: Facial Expressions and Emotion Database, Technical University of Munich, 2005. http://www.mmk.ei.tum.de/˜waf/fgnet/ feedtum.html.

[12] T. Wilhelm, H.-J. Boehme, and H.-M. Gross. Classification of Face Images for Gender, Age, Facial Expression, and Identity. In *Proc. ICANN*, pages 569–574, 2005.