in: Computer Vision Systems (ICVS 2009), Liege, LNCS 5815, pp. 364-373, Springer Verlag 2009

Increasing the Robustness of 2D Active Appearance Models for Real-World Applications

Ronny Stricker[1], Christian Martin[1,2], and Horst-Michael Gross[1⋆]

[1] Neuroinformatics and Cognitive Robotics Lab,
Ilmenau University of Technology, Germany
{ronny.stricker,christian.martin,horst-michael.gross}@tu-ilmenau.de
http://www.tu-ilmenau.de/neurob
[2] MetraLabs GmbH, Germany

Abstract. This paper presents an approach to increase the robustness of *Active Appearance Models* (AAMs) within the scope of human-robotinteraction. Due to unknown environments with changing illumination conditions and different users, which may perform unpredictable head movements, standard AAMs suffer from a lack of robustness. Therefore, this paper introduces several methods to increase the robustness of AAMs. In detail, we optimize the shape model to certain applications by using genetic algorithms. Furthermore, a modified retinex-filter to reduce the influence of illumination is presented. These approaches are finally combined with an adaptive parameter fitting approach, which can handle bad initializations. We obtain very promising results of experiments evaluating the IMM face database [1].

Key words: Active Appearance Model, Genetic Algorithm, Retinex-filter, Illumination, Optimization

1 Introduction

Within the scope of human-robot-interaction, it is often necessary to analyze the identity and the emotional state of the dialog partner. *Active Appearance Models* (AAM) have been established to characterize non-rigid objects, like human heads, and can be used to analyze the users state based on visual features. Therefore, the parameters of the AAM are adapted, so that the model fits to the current face. Afterwards, the parameters of the AAM can be utilized to determine the expression or gender of the users face. The main drawback of this approach is that it depends to a large extent on the current operational environment. Especially under real world conditions with uncontrolled observation constraints a mobile robot has to cope with different problems arising from these dependencies. This work suggests improvements of the robustness of the

⋆ The research leading to these results has received partial funding from the European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216487 (CompanionAble-Project).

adaption step in AAMs to gain a higher independence of the operational environment. This paper is organized as follows: After a brief description of the basics of AAMs, Sect. 3 gives an overview of the related work. Afterwards, we introduce our contribution to increase the robustness of the fitting process. Sect. 5 shows the results which can be achieved with the help of the proposed methods. The paper concludes with a summary and an outlook on ongoing work in Sect. 6.

2 Basics of Active Appearance Models

Active Appearance Models, first introduced in [2], provide a good possibility to model non rigid objects within the scope of image processing and are, therefore, very popular to model human faces or viscera. The AAM itself is a combination of two statistical models. First, the shape model represents the geometry of the object. Secondly, the appearance model allows the modeling of the object texture within the normalized mean shape of the model. The models are built by training images, which are labeled with landmark points on certain positions of the object. These n landmark locations build up the shape $\mathbf{s} = (x_1, y_1, ..., x_n, y_n)^T$ of an AAM instance. Using a Principle Component Analysis (PCA) for all training shapes, the resulting shape model can be represented by a set of shape parameters $\mathbf{p}$ combined with the basis shapes $\mathbf{s_i}$:

$$\mathbf{s}(\mathbf{p}) = \mathbf{s_0} + \sum_{i=1}^{n} p_i \mathbf{s_i}. \tag{1}$$

Afterwards, a triangulation of the mean shape $\mathbf{s_0}$ is used to establish a relation between the labeled points and the surface of the object. With the help of surface triangles, every single point on arbitrary shape $\mathbf{s_i}$ can be warped to a destination shape $\mathbf{s_j}$ using an affine transformation. With respect to [3] we can describe this transformation as $W(\mathbf{x}; \mathbf{p})$, which maps a point $\mathbf{x} = (x, y)^T$ within the model shape to the shape defined by the parameters $\mathbf{p}$. This transformation is used afterwards to build the appearance model, which is very similar to the shape model. The important difference is that each texture sample A_i, defined by the training images, is warped to the mean shape $\mathbf{s_0}$, using the described affine transformation. The texture parameters resulting from the subsequent PCA are denoted as λ. Therefore the texture object is very similar to the *Eigenface* approach:

$$\mathbf{A}(\lambda) = \mathbf{A_0} + \sum_{i=1}^{m} \lambda_i \mathbf{A_i} \quad , \forall \mathbf{x} \in \mathbf{s_0}. \tag{2}$$

The resulting AAM can represent any object instance M covered by the training data using the shape parameter vector $\mathbf{p}$ and the appearance parameter vector λ using (3).

$$M(W(\mathbf{x}, \mathbf{p})) = \mathbf{A_0}(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \mathbf{A_i}(\mathbf{x}) \quad , \forall \mathbf{x} \in \mathbf{s_0}. \tag{3}$$

The goal of fitting an AAM to an unknown image, as defined by [2], is to minimize the squared difference between the synthesized model and the given image. Using

gradient descent to solve this problem leads to a very efficient fitting algorithm. To overcome the problem of simultaneous optimization of shape- and appearance parameters, Baker and Matthews introduced the *Project-Out* gradient descent image alignment algorithm [4]. As the exact formulation of the fitting algorithm lies beyond the scope of this paper, the reader is referred to [3, 4] for more detailed information.

3 Related Work

Within the last years, AAMs have become very popular for the purpose of face tracking [5, 6] or classification tasks, like facial expression recognition [7, 8]. In this context, the problems of illumination independence and robust fitting have been addressed by different approaches. A common method to cope with illumination changes is to model the illumination explicitly as shown in [9]. Besides the construction of the model, this methods add additional parameters to the AAM, which have to be determined during the fitting process and hence increase the complexity. A survey on different approaches dealing with illumination can be found in [10]. The problem of fitting robustness is addressed in [6] by using a hierarchy of models with different complexities. However, this approach involves the toggling between different models which complicates the combination with tracking algorithms. The problem of finding the optimal shape for an AAM, however, has been addressed significantly less in the literature. The only available work concentrates on optimizing the landmarks in terms of their salience as shown in [11]. To our knowledge, this is the only approach which tries to optimize the shape of an existing shape model.

4 Increasing the Robustness

Due to the principle of minimizing the difference between the input image and the synthesized model, the fitting process is very sensitive to differences between the training images and the images used during model fitting [12]. Furthermore, wrong initializations can lead the fitting process to local minima and, therefore, may cause a bad match. This problem increases with the number of model parameters growing as the complexity of the error surface increases as well. Therefore, the AAM shouldn't exceed the needed complexity for the desired application.

4.1 Optimization of the Shape Model

The construction of an AAM is based on training images which are labeled with specified landmark points. As a result, the model quality, defined as its ability to fit to unknown images, depends on the quality of the landmarks and the training images. Unfortunately, the process of adding landmarks to unknown images is quite complex and manual work is indispensable at least to refine the landmark positions. Yet, this process itself is very error-prone as well. Our tests have shown that the variance of the position of hand-labeled landmark points is very

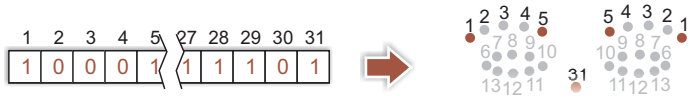

Fig. 1. Exploiting the symmetry constraint, the 58 landmarks of the IMM database shape [1] can be coded using a genome with 31 bits. Only the upper half of the face landmark points are displayed.

high. Furthermore, these errors are not equally distributed, so that landmarks in heavily textured regions can be reproduced very well. One way to find such reliable features using their salience is described in [11]. However, building a model based on the most salient features is not always equal to finding those landmarks optimal for the desired application purpose. Therefore, we present a new method to reduce the number of given landmarks to an explicit set in order to reduce the influence of badly labeled landmarks and to reduce the model complexity. Examples for such a reduced model can be found in [6] and [13], where some kind of *inner-face-model* is used.

Ideally, to reduce a given set of landmarks to an optimal set regarding the desired application involves the analysis of all combinations of different landmarks. Even if the symmetry of the human face it taken into account, the search domain is typically too large to be holistically analyzed. However, it can be expected that the adding and removing of several landmark points from the model have similar effects on different submodels. Therefore, it is a common way to use some kind of evolutionary search, e.g. genetic algorithm, to analyze the search domain in a sparse but purposive way. The different possible shape models are coded as a genome exploiting the symmetry of the human face (Fig. 1). To evaluate a genome, the corresponding AAM is generated and applied to a test dataset afterwards. The dataset should be designed in such a way that the desired application (e.g. emotion recognition) can be represented as good as possible. Therefore, it should contain the respective classification task of interest.

4.2 Adding Robustness to Illumination Changes

Especially within the scope of face recognition, the effects of illumination have been examined very well [10]. The explicit modelling of the illumination can provide satisfying results, but is generally very complex and often not capable of real-time processing. Nevertheless, the model free *retinex filter* first introduced in [14] can achieve promising results within the scope of removing the influence of different kinds of illumination. This filter relates each pixel of the image to its local surroundings:

$$R(\mathbf{x}) = logI(\mathbf{x}) - log\,|F(\mathbf{x}) * I(\mathbf{x})| \tag{4}$$

where I is the input image and F denotes a function representing the surroundings of the pixel $\mathbf{x}$. Unfortunately, it is necessary to set the size of the surrounding area to an appropriate value to avoid problems of ghost shadows and the loss of detail or insufficient illumination normalization (Fig. 2). This problem has been

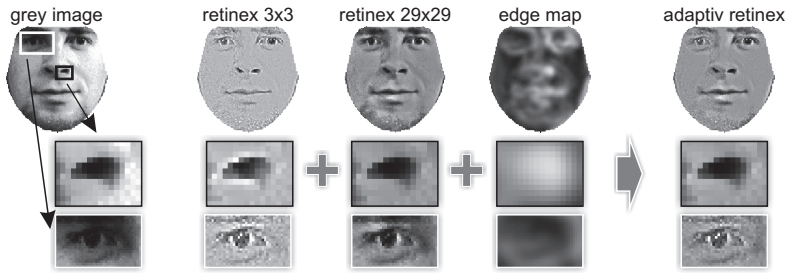

Fig. 2. Comparison of different retinex approaches - the original retinex approach with a surrounding of 3x3 Pixels generates ghost shadows (nostril) and reduces the detail (eye). The retinex filter with a surrounding of 29x29 pixels in turn shows only poor illumination normalization (eye). The adaptive retinex, however, combines the advantages of both filter sizes due to selective combination with the help of the edge map.

addressed with the *multiscale retinex* approach presented in [15], which combines retinex filters with different sizes of the surrounding area. Unfortunately, the described approach only diminishes the problems occurring from wrong parametrisation. Another approach can be found in [16], where the surrounding function is modeled using an anisotropic filter. Unfortunately, the filter is computationally intensive which is not desirable for real time applications. As a combination of the approaches described in [15] and [16], we introduce a kind of *adaptive retinex* filter. We combine the idea of the *multiscale retinex* approach with a dynamic combination function, which depends on the local edge strength. Therefore, we use two retinex filter $R_1(\mathbf{x})$ and $R_2(\mathbf{x})$ with different sizes of the surroundings and add an edge detector $E(\mathbf{x})$, which computes the local edge strength. If the surrounding size of R_2 is smaller than the size of R_1 the combination can be expressed by (5).

$$\mathcal{S}(\mathbf{x}) = \begin{cases} R_1(\mathbf{x}) & K(\mathbf{x}) < l_l \\ R_2(\mathbf{x}) & K(\mathbf{x}) > l_u \\ \frac{K(\mathbf{x})}{l_u} R_1 + (1 - \frac{K(\mathbf{x})}{l_u}) R_2 & l_l \leq K(\mathbf{x}) \leq l_u \end{cases} . \tag{5}$$

Where l_l and l_u denote the lower and upper bounds of the retinex filter with the bigger or smaller surroundings. For edge values between the lower and upper bound, a combination of the two different retinex filter outputs is taken. Due to the dynamic combination of different surrounding sizes, the presented filter is not that addicted to specific illumination conditions. Furthermore, the filter can be computed in a much more efficient manner than the anisotropic one [16].

4.3 Adaptive Parameter Fitting

The *Project-Out* fitting algorithm uses gradient descent and is, therefore, very sensitive to get stuck in local minima. One way to deal with this problem is to apply a hierarchy of models with an increasing number of parameters, as we have already shown in [17]. Nevertheless, it is hard to decide at which point of

time the fitting process should switch to a more detailed model. Furthermore, if applied to tracking purpose, the model parameters of a detailed model have to be refused if the fitting process has to switch back to a simple model. This paper introduces an approach for adaptive parameter fitting, which works with only one model of the object (in detail the face).

Due to the applied PCA, used to build the shape model of the face, the shape *Eigenvectors* can be sorted according to their *Eigenvalues*. The *Eigenvalues* in turn represent the variance of the training data in the direction of the associated *Eigenvector*. So, the first *Eigenvectors* have a higher importance to represent the given training data. Standard gradient descent fitting algorithms, like the *Project-Out* algorithm, are based on adapting all model parameters at the same time. However, this approach can force the model parameters, which are associated with *Eigenvectors* with lower importance, to diverge. The reason for this behaviour is that the first, and most important, parameters are not yet stabilized, so that the later parameters are likely to converge into the wrong direction (Fig. 3). We try to address this problem by dividing the model parameters into two different groups. First, the *primary* parameters which are important for the main head movements like pan and tilt, and the *secondary* parameters, responsible to code the shape variance of the inner face. Then, we can suppress changes of the *secondary* shape parameter during the fitting process as long as the *primary* shape parameters have not been stabilized. To detect the strength of the parameter changes, we compute the normalized parameter changes of the n *primary* parameters using the *Eigenvalues* EV:

$$E_p = \sum_{i=1}^{n} \left(\frac{\Delta p_i}{EV(p_i)} \right)^2 . \tag{6}$$

Afterwards, the parameter changes of the *secondary* parameters can be scaled using a logarithmic coefficient which equals zero if E_p equals the squared sum of all *Eigenvalues* of the *primary* parameters and is equal to 1 if E_p is equal to zero. As shown in Fig. 3 the introduced adaption of the *secondary* parameters can successfully smooth the parameter changes and, thus lead to a more purposive model fitting as we intend to show in Sect. 5.

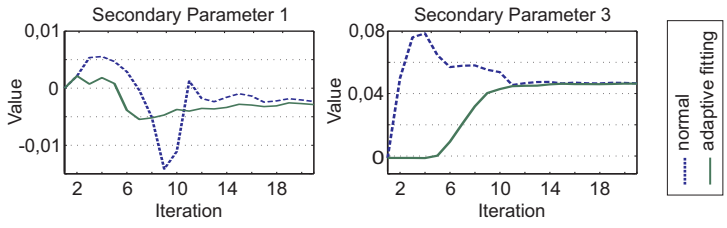

Fig. 3. Simultaneous fitting of all parameters can lead the *secondary* parameter into a wrong direction. Adaptive fitting can improve this behaviour by repressing changes of the *secondary* parameters as long as the *primary* parameters are not stabilized (until iteration 6).

5 Experimental Results

This section presents some experimental results we have achieved by using the described approaches. We decided to use the IMM face database [1] for our studies to produce consistent and meaningful results. The database consists of 6 different images for each of 40 different people. The images contain frontal and side views, as well as sidewise illuminated images and emotion images. To fit the built models to the images, we use the standard *Project-Out* fitting algorithm as described in [3] and start each fitting process with a frontal initialization to give consideration to common face detectors. As the AAMs are prone to initialization errors, we start the fitting process for each model and image for a certain amount of rounds, whereas the initialization is perturbed in every round with increasing variance. Afterwards, the quality of every fitting process is evaluated using a combined measure between the mean and the maximum distance between the ground truth shape, provided by the IMM database, and the fitted shape. This measure is able to distinguish between converged and diverged models using a threshold. Although this threshold appears to be seemingly at random and makes comparisons with other papers more difficult, we found it to be a good and meaningfull measure for quantitative comparisons of the suggested improvements. The fitting rates given below always refer to the declared images of every person within the IMM database (frontal images refer to the images 1 and 2; sidewise view images refer to the images 3 and 4; illumination image refers to image 5).

Optimization of the Shape This section presents the optimization of the shape, given by the IMM database, with respect to fitting accuracy. Therefore, we evaluate each computed genome with respect to its fitting accuracy, computed as the mean and maximum distance between the ground truth and the resulting shape. Thereby, the generated shapes show a significant improvement over the complete shape with respect to the distance values. Having a closer look at the

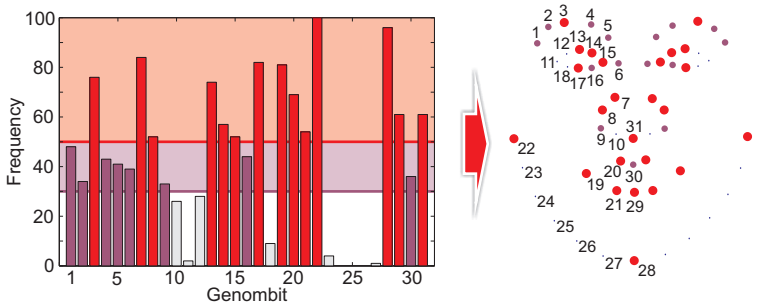

Fig. 4. The different landmarks of the 100 best genomes are color coded according to their frequency of employment. The lower surrounding of the face can be represented sufficiently by labels 22 and 28, whereas labels 23 to 27 can be ignored. For the sake of clarity only half of the labelpoints are annotated.

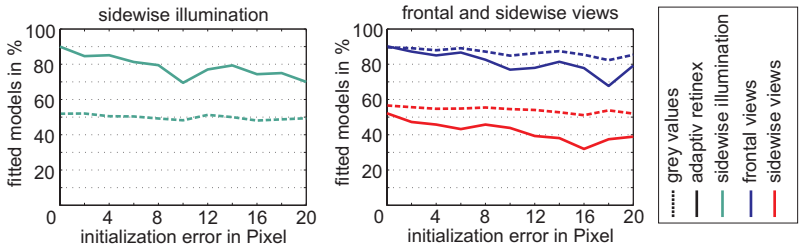

Fig. 5. Using the *adaptive retinex* filter significantly improves the fitting performance for sidewise illuminated images. Unfortunately filter tends to be more sensitive to bad initialization.

used landmark points of the 100 best shapes, it can be seen that especially the surrounding of the face is not necessary for accurate model fitting (Fig. 4). This is an interesting finding in contrast to commonly used AAM labeling instructions. The landmarks in the inner face region are least affected by the shape reduction. This points seem to be necessary for reliable model fitting given different poses and emotions.

Robustness to Illumination Changes To show the benefit of the proposed *adaptive retinex* filter we build the AAMs based on the images with frontal illumination. The models are applied afterwards to the images with sidewise illumination (Fig. 5). Although the fitting can be significantly improved for images with sidewise illumination, the image preprocessing seems to be more sensitive to bad initialization. The *adaptive retinex* filter also removes slight illumination changes occurring from the three-dimensional structure of the head – for example the illumination on the cheek. This seems to complicate the fitting process, especially for rotated heads. Nevertheless, this disadvantage seems to be uncritical in most cases due to the great benefit achieved for sidewise illumination.

Adaptive Fitting The effect of applying adaptive fitting to the *Project-Out* algorithm on frontal and sidewise views is illustrated in Fig. 6. While the fitting

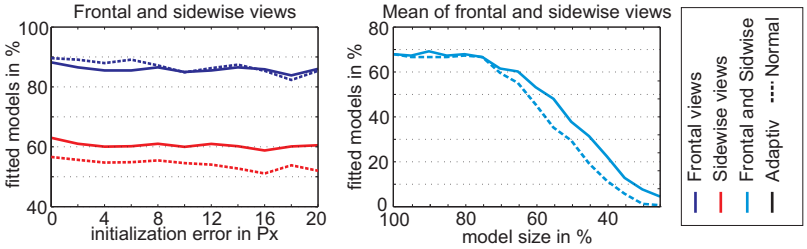

Fig. 6. *Adaptive fitting* improves the fitting performance for images with bad initialization. Therefore sidewise views and models with wrong scaling initialization can be improved. Left: Shape perturbed in x and y direction. Right: Shape with perturbed scaling.

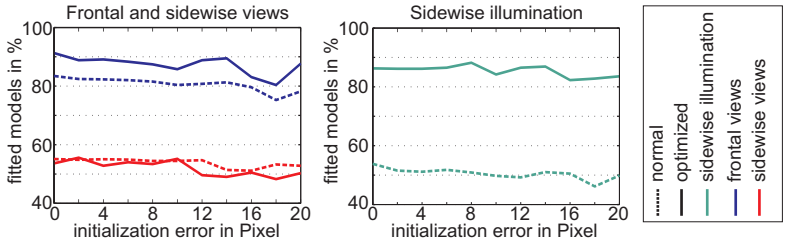

Fig. 7. The combined approach achieves better fitting performance. Especially the performace for sidewise view can be significantly improved from 50 % to 85 %.

performance remains almost the same for frontal views, it can be improved for sidewise views. This result indicates that the adaptive fitting successfully enables the algorithm to rotate the model, before the adaption of the *secondary* parameters is carried out. The same improvement can be observed for a model initialization with perturbed size parameters. Again the model is able to scale the main shape before the *secondary* parameters were fitted and, therefore, is able to produce significantly better results.

Full Optimization After having discussed the effects of each of the introduced methods, now the results for the combined approach are presented in Fig. 7. Although the different methods degrade under certain circumstances, the combined approach produces better results for all kinds of images of the IMM database. Although the fitting performance is only slightly increased for neutrally illuminated images, it can be significantly improved for sidewise illumination conditions. Therefore, the fitting performance can be increased by approximately 10 percent for frontal views, and by 35 percent for sidewise illuminated images.

6 Conclusion and Future Work

In this paper, we have presented different methods to increase the robustness of AAMs. First of all, an approach to adapt the complexity of the shape model to certain applications is introduced. Besides the reduction of the shape complexity, this method may be used to find reduced sets of label points to speed up the labeling process. We have successfully shown that the shape model defined by the IMM database can be reduced from 58 points to 25 points in order to increase fitting accuracy. Furthermore, we have introduced an *adaptive retinex* filter, which is able to normalize different illumination conditions, which occur in uncontrolled environments. To cope with fast head movements and rotated faces, we applied an adaptive parameter fitting, to guide the model parameter within the high dimensional error surface. The different methods show promising results for different AAM specific problems. The combination of all methods leads to a robust and real-time capable approach which has been tested in our lab on the mobile robot SCITOS and performs significantly better than standard approaches. Continuing our work, we will cope with the problem of shape optimization within the scope of emotion classification.

References

1. M. B. Stegmann, B. K. Ersbøll, R. Larsen R.: FAME – A Flexible Appearance Modelling Environment. In: IEEE Trans. on Medical Imaging, pp. 1319–1331 (2003)
2. Cootes, T.F., Edwards, G., Taylor, C.J.: Active Appearance Models. In: Proc. of the European Conf. on Computer Vision (1998)
3. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. In: Int. Journal of Computer Vision, pp. 221–255 (2004)
4. Baker, S., Matthews, I.: Equivalence and efficiency of image alignment algorithms. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1090–1097 (2001)
5. Baker, S., Matthews, I., Xiao, J., Gross, R., Kanade, T.: Real-time non-rigid driver head tracking for driver mental state estimation. At: Robotics Institute, Carnegie Mellon University (2004)
6. Kobayashi, A., Satake, J., Hirayama, T., Kawashima, H., Matsuyama, T.: Person-Independent Face Tracking Based on Dynamic AAM Selection. In: 8th IEEE Int. Conf. on Automatic Face and Gesture Recognition (2008)
7. M. S. Ratliff and E. Patterson. Emotion Recognition using Facial Expressions with Active Appearance Models. In Proc. of HRI (2008)
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, pp. 394–400 (2006)
9. Kahraman, F., Gokmen, M., Darkner, S., Larsen, R.: An Active Illumination and Appearance (AIA) Model for Face Alignment. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
10. Zou, X., Kittler, J., Messer, K.: Illumination Invariant Face Recognition: A Survey. In: IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems, pp. 1–8 (2007)
11. Nguyen, M. H., la Torre Frade, F. D.: Facial Feature Detection with Optimal Pixel Reduction SVMs. In: 8th IEEE Int. Conf. on Automatic Face and Gesture Recognition (2008)
12. De la Torre, F., Collet, A., Quero, M., Cohn, J., Kanade, T. Filtered Component Analysis to Increase Robustness to Local Minima in Appearance Models. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
13. Kim, D., Kim, J., Cho, S., Jang, Y., Chung, S.-T., Kim, B. G.: Progressive AAM Based Robust Face Alignment. In: Proc. of world academy of science, engineering and technology, pp. 483–487 (2007)
14. Jobson, D., Rahman, Z., Woodell, G.: Properties and performance of a center/surround retinex. In: IEEE Transactions on Image Processing, pp. 451–462 (1997)
15. Jobson, D., Rahman, Z., Woodell, G.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. In: IEEE Transactions on Image Processing, pp. 965–976 (1997)
16. Wang, H., Li, S., Wang, Y.: Face recognition under varying lighting conditions using self quotient image. In: Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 819–824 (2004)
17. Martin, Ch., Gross, H.-M.: A Real-time Facial Expression Recognition System based on Active Appearance Models using Gray Images and Edge Images. In: Proc. of the 8th IEEE Int. Conf. on Face and Gesture Recognition, paper no. 299 (2008)