



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Next Generation Face Tracking Technology Using AAM Techniques
Author(s)	Corcoran, Peter; Ionita, Mircea C.; Bacivarov, Ioana
Publication Date	2007
Publication Information	Corcoran, P., Ionita, M. C., & Bacivarov, I. (2007). Next Generation Face Tracking Technology Using AAM Techniques. Paper presented at the Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on.
Publisher	IEEE
Link to publisher's version	http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4292639
Item record	http://hdl.handle.net/10379/1618

Downloaded 2024-03-13T08:26:09Z

Some rights reserved. For more information, please see the item record link above.



Next Generation Face Tracking Technology Using AAM Techniques

Peter Corcoran, Mircea C. Ionita, Ioana Bacivarov
Department of Electronic Engineering, NUI Galway, Ireland

Abstract—This paper realizes an investigation of the Active Appearance Model (AAM) techniques, with particular emphasis on problems related to the implementation of AAM in real-time face tracking applications for embedded systems. The paper includes (a) new training and model construction approaches for AAMs with increased robustness to head pose variations and changes in the illumination conditions, and (b) a corresponding AAM-based face tracking scheme.

I. INTRODUCTION

Arrived already at a stage of implementation and commercialization in digital cameras, face-tracking still remains an important and challenging task in Computer Vision. A major difficulty in face-tracking is the potential variability of human faces over time due to factors like: variation in pose or illumination and occlusions of the tracked object. When not controlled, any one of these sources of variability is enough to cause a tracking algorithm to lose its target.

Various methods have been proposed to overcome these challenges and the literature on face-tracking is abundant. According to [1], face tracking methods could be classified into three main groups: low level feature approaches, template matching approaches and statistical inference approaches. The low level feature approaches [2], [3], [4], [5], [6], [7] make use of low level face knowledge, such as skin color, background knowledge (background subtraction or rectangular features) or motion information to track faces. The template matching approaches involved tracking contours with snakes [8], 3D face model matching [9], [10], [11], shape and face template matching and wavelet networks matching. In [11], Mittrapiyanuruk *et al.* present a new method for tracking rigid objects using a modified version of the AAM, making the algorithm robust to partial and self occlusion of objects. AAM [12] is generally considered a complex deformable template model which has previously been applied to still images as well as for tracking [13], [14], [15], [16], [17]. In tracking, AAM has been successfully used as a stand-alone method [13], [14], as well as extended to 3D models. In [15], an efficient head and facial feature tracker combining a robust feature-based 3D pose estimator with the active model, allowing AAM to be more accurate and flexible, is proposed. In [17], Dornaika and J. Ahlberg address the 3D tracking of pose and animation of the human face in monocular image sequences using AAM, the main contribution being that for the search algorithm, CPU-time is not dependent on the dimension of the face space. The basic idea is to use the synthesis results associated with the old frame instead of carrying it out in the iterative search.

The third tracking category, statistical inference approaches, includes Kalman filtering techniques for uni-modal Gaussian representations [18], [19], Monte Carlo approaches for non-Gaussian nonlinear target tracking [20] and Bayesian Network inference approaches [21].

As technology progresses, the embedded imaging research starts to focus more and more on extending existing camera functionality to include embedded computer vision. It is already anticipated that machine vision will migrate from industrial applications to consumer and commercial domain. An embedded computer vision application should meet constraints on issues like real-time performance, power consumption, or memory requirements.

In this paper we propose an AAM-based face tracking technique, aimed to meet the requirements and constraints imposed by the embedded software systems. The technique highly relies on a statistical face model, specially designed so that to insure robustness to head pose variations and changes in the illumination conditions. The outline of this paper is as follows. Section II summarizes the main steps in generating and fitting a face appearance model of shape and texture. In Section III we investigate how a robust AAM-based face tracking application can be designed. We propose in this section a new training method for generating a face model, robust to head pose variations as well as to changes in the illumination conditions. The proposed training method allows us to further apply the fast AAM technique for optimizing the parameters of our improved model. The AAM-based face tracking scheme is then described in Section III-B. The robustness of the face tracking technique is then analysed in Section IV, based on preliminary tests on two standard video datasets. Finally, the conclusions of our research are drawn in Section V.

II. SHAPE AND TEXTURE MODELS OF APPEARANCE

Face appearance variability is modelled using separate shape and texture models. Shape is defined as the set of landmarks used to best describe the contour of the object of interest (*i.e.* the face), while texture is defined as the set of pixel values across the object of interest. A shape vector is obtained by concatenating the coordinates of all landmark points, as $(x_1, x_2, \dots, x_L, y_1, y_2, \dots, y_L)^T$, where L is the number of landmark points. A shape model is then obtained by applying PCA on the set of aligned shapes

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s, \quad (1)$$

where $\bar{\mathbf{s}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{s}_i$ is the mean shape vector, with N_s the number of shape observations; Φ_s is the matrix having as its columns the eigenvectors; \mathbf{b}_s defines the set of parameters of the shape model.

To obtain the texture model, face patches are first warped into the mean shape based on a triangulation algorithm. Then a texture vector $(t_1, t_2, \dots, t_P)^T$ is built for each training image by sampling the values across the warped (shape-normalized) patches, with P being the number of texture samples. The texture model is also derived by means of PCA on the texture vectors

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{b}_t, \quad (2)$$

where $\bar{\mathbf{t}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{t}_i$ is the mean texture vector, with N_t being the number of texture observations; Φ_t is again the matrix of eigenvectors, and \mathbf{b}_t the parameters for the texture model.

The sets of shape and texture parameters $\mathbf{c} = \begin{pmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{pmatrix}$ are used to describe the overall *appearance* variability of the modelled object, where \mathbf{W}_s is a vector of weights used to compensate for the differences in units between shape and texture parameters.

AAM is a fast technique used to optimize the parameters of a statistical model of appearance [12]. This is based on the approximation that a linear relationship exists between the residual texture error, which is given by the difference between the query and the modelled images, and the error in the model parameters. This approximation enables the use of a fixed gradient matrix for iteratively updating the parameters. The gradient matrix can be estimated offline from a set of training images. AAM appeared as an alternative to the standard optimization technique, which is computationally expensive, involving the numerical computation the gradient matrix with each iteration.

III. AN AAM-BASED FACE TRACKING TECHNIQUE

A. Appearance Model Design for Face Tracking

Firstly, a full appearance model, robust to head pose variations and changes in the lighting conditions, must be designed.

The shape model should be designed so that to include head pose variability. This is learnt from a special annotated face database where each individual is presented in several head poses. The shape parameters that encode the head pose variability are usually easy to identify. Normally, two of the shape parameters account for most of this variability and very little for other types of variability. Thus, the shape model in (1) can be explicitly written as

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_{s_pose} \mathbf{b}_{s_pose} + \Phi_{s_id} \mathbf{b}_{s_id}, \quad (3)$$

where $\Phi_{s_pose} \mathbf{b}_{s_pose}$ and $\Phi_{s_id} \mathbf{b}_{s_id}$ denote the separation between head pose and identity attributes of the shape model.

In order to obtain a robust face tracking application, we need to account also for the possible changes in the illumination conditions. Illumination conditions are a very important

attribute of a digital image. It is known that the variance introduced by changes in illumination can be higher than the variance introduced by identity changes. Thus, it is essential to find a way of dealing with illumination.

Our face tracking approach in this paper is based on modelling the appearance of the face. We thus try to model also the variability which can be introduced by directional lighting. Yet, it was shown that AAM cannot be successfully applied on images where important variations in the illumination conditions are present [22]. This is because the estimated gradient specializes around the mean of the dataset it is built from. On the other hand, by modelling only identity variability and assuming uniform illumination can drastically limit the applicability of an AAM.

In order to solve this problem, we then consider a more elaborated way of modelling texture, which separates identity variability from the variability introduced by changes in the illumination conditions. The texture model in (2) is now be written as

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_{t_id} \mathbf{b}_{t_id} + \Phi_{t_illum} \mathbf{b}_{t_illum}, \quad (4)$$

where Φ_{t_illum} defines an added eigenspace for directional lighting, and \mathbf{b}_{t_illum} are the parameters which model this type of variation.

A special face database was used for extracting the directional lighting eigenspace and for modelling changes in the illumination conditions. This database shows a number of individuals in fixed frontal pose, while only the position of a light source is being altered.

Employing the texture model in (4) we are now able to solve the AAM illumination variability problem by projecting out the appearance variation introduced by directional lighting, as shown in (5), (6).

$$\mathbf{b}_{t_illum_opt} = \Phi_{t_illum}^T \mathbf{t}_{im}, \quad (5)$$

$$\mathbf{t}_{im_unif} = \mathbf{t}_{im} - \Phi_{t_illum} \mathbf{b}_{t_illum_opt}, \quad (6)$$

where \mathbf{t}_{im} represents the current image texture vector, and $\mathbf{b}_{t_illum_opt}$ is the set of optimal illumination parameters for the current image texture; \mathbf{t}_{im_unif} represents the filtered image texture vectored, corresponding to uniform illumination conditions.

By doing this, the non-uniform illumination component in the image is filtered out. We can now further apply the AAM technique for optimizing the remaining model parameters in the same way as it is done for uniform illumination.

B. The Video Face Tracking Scheme

A face detection stage is firstly employed in order to obtain an initial rough estimate of the face position, 2D rotation angle, and scale. These estimates are used to initialize the statistical model inside the image frame. We then distinguish between two separate stages of model fitting. As long as the model is able to model the face it means that the face tracking task is also being carried out. During the first stage of fitting the

model, all parameters need to be optimized, *i.e.* the shape parameters (including head pose) and texture parameters, as well as the four parameters which characterize the modelled patch inside the image frame. This stage is the most computationally expensive and it is to be performed over several frames. When the model parameters which define the person identity are successfully optimized, meaning that important changes in their values can no longer be recorded, they remain fixed to their current values. We then proceed to the second stage, in which we optimize only the model parameters that encode head pose variations, as well as the position, 2D rotation and scale parameters. As the computational requirements are substantially reduced, this operation is now performed on each separate frame.

IV. RESULTS

We tested the proposed technique on two different standard facial video databases, namely the VidTIMIT database [23], and the NRC-IIT facial video database [24], [25]. The first database shows different individuals while rotating their heads from right to left, up to full profiles, and then moving the head up and down. The second database presents a more variate set of movements including also forward and backward movements, which causes also changes in face illumination. In order to reduce the computational complexity during fitting the model parameters, we also use a predefined reduced resolution of the face patch. Applying the technique on the VidTIMIT database shows good tracking results for all individuals up to $\sim 30^\circ$ for horizontal face orientations and up to $\sim 45^\circ$ for vertical face orientations. The model is limited to these ranges because this 3D face variability is actually learnt using 2D shape models. A tracking example for this database is shown in Fig. 1. On the NRC-IIT database the model is able to



Fig. 1. Face tracking sequence example.

track the face also over the more complex movements of the head. Particularly, the tracking operation is not affected by the introduced variations in the illumination of the face. Weak points can though be noticed when the face becomes partially occluded (*e.g.* by moving a hand in front of the face).

V. CONCLUSIONS

We analysed in this paper the possibility of using AAM-based face tracking techniques in embedded systems like digital cameras. We proposed a model fitting method for the statistical model of the face, adapted for face tracking in video sequences. Furthermore, we consider that the reduced computational costs, associated with the proposed parameter optimization scheme, could meet the requirements of real-time applications in embedded (*e.g.* camera) systems.

As future work we want to extend the technique so that to increase the robustness to a wider range of head pose variations, or to partial occlusions of the face region. This could be achieved by coupling the AAM-based approach with other techniques like for example skin detection.

REFERENCES

- [1] "The extension of statistical face detection to face tracking," in *CRV '04: Proceedings of the 1st Canadian Conference on Computer and Robot Vision (CRV'04)*, Washington, DC, USA, 2004, pp. 10–17, IEEE Computer Society.
- [2] V. Comaniciu and P. Meer, "Kernel-based object tracking," 2003.
- [3] P. Fieguth and D. Terzopoulos, "Color based tracking of heads and other mobile objects at video frame rates," 1997.
- [4] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.
- [5] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.
- [6] Paul Viola and Michael Jones, "Robust real-time object detection," *International Journal of Computer Vision - to appear*, 2002.
- [7] A. Jepson, D. Fleet, and T. Maraghi, "Robust online appearance models for visual tracking," 2001.
- [8] J. Sobottka and I. Pittas, "Segmentation and tracking of faces in color images," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 236–241.
- [9] D. M. Gavrilu and L. S. Davis, "3d model-based tracking of humans in action: a multi-view approach," Tech. Rep., College Park, MD, USA, 1995.
- [10] Marco La Cascia, Stan Sclaroff, and Vassilis Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 322–336, 2000.
- [11] Pradit Mittrapiyanuruk, Guilherme N. DeSouza, and Avinash C. Kak, "Accurate 3d tracking of rigid objects with occlusion using active appearance models," in *WACV-MOTION '05: Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION'05) - Volume 2*, Washington, DC, USA, 2005, pp. 90–95, IEEE Computer Society.
- [12] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *Lecture Notes in Computer Science*, vol. 1407, pp. 484–??, 1998.
- [13] M. B. Stegmann, "Object tracking using active appearance models," in *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis*, Søren I. Olsen, Ed., Copenhagen, Denmark, jul 2001, vol. 1, pp. 54–60, DIKU.
- [14] J. Ahlberg, "Using the active appearance algorithm for face and facial feature tracking," 2001.
- [15] F. Dornaika and J. Ahlberg, "Efficient active appearance model for real-time head and facial feature tracking," in *AMFG '03: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Washington, DC, USA, 2003, p. 173, IEEE Computer Society.
- [16] S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-based modeling in particle filters," 2003.
- [17] Fadi Dornaika and Jrgen Ahlberg, "Fast and reliable active appearance model search for 3d face tracking," .

- [18] J. Strom, T. Jebara, S. Basu, and A. Pentland, "Real time tracking and modeling of faces: An ekf-based analysis by synthesis approach," 1999.
- [19] Douglas DeCarlo and Dimitris N. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *International Journal of Computer Vision*, vol. 38, no. 2, pp. 99–127, 2000.
- [20] Michael Isard and Andrew Blake, "ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework," *Lecture Notes in Computer Science*, vol. 1406, pp. 893–908, 1998.
- [21] Fang Liu, Xueyin Lin, Stan Z Li, and Yuanchun Shi, "Multi-modal face tracking using bayesian network," *amfg*, vol. 00, pp. 135, 2003.
- [22] A.U. Batur and M.H. Hayes, "Adaptive active appearance models," vol. 14, no. 11, pp. 1707–1721, November 2005.
- [23] Conrad Sanderson, "The VidTIMIT Database," IDIAP-Com 06, IDIAP, 2002.
- [24] Dmitry O. Gorodnichy, "Video-based framework for face recognition in video," in *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision (CRV'05)*, Washington, DC, USA, 2005, pp. 330–338, IEEE Computer Society.
- [25] Dmitry O. Gorodnichy, "Associative neural networks as means for low-resolution video-based recognition," in *International Joint Conference on Neural Networks (IJCNN'05)*, 2005.