Background Robust Face Tracking Using Active Contour Technique Combined Active Appearance Model*

Jaewon Sung and Daijin Kim

Biometrics Engineering Research Center (BERC),
Pohang University of Science and Technology
{jwsung, dkim}@postech.ac.kr

Abstract. This paper proposes a two stage AAM fitting algorithm that is robust to the cluttered background and a large motion. The proposed AAM fitting algorithm consists of two alternative procedures: the active contour fitting to find the contour sample that best fits the face image and then the active appearance model fitting over the best selected contour. Experimental results show that the proposed active contour based AAM provides better accuracy and convergence characteristics in terms of RMS error and convergence rate, respectively, than the existing robust AAM.

1 Introduction

Active Appearance Models (AAMs) [1] are generative, parametric models of certain visual phenomena that show both shape and appearance variations. These variations are represented by linear models such as Principal Component Analysis (PCA), which finds a subspace reserving maximum variance of given data. The most common application of AAMs has been face modeling [1], [2], [3], [4].

Although the structure of the AAM is simple, fitting an AAM to an target image is a complex task that requires a non-linear optimization technique that requires a huge amount of computation when the standard non-linear optimization techniques such as the gradient descent method are used. Recently, a gradient based efficient AAM fitting algorithm, which is extended from an inverse compositional LK image matching algorithm [5], has been introduced by Matthews et. al. [4]. The AAM fitting problem is treated as an image matching problem that includes both shape and appearance variations with a piece-wise affine warping function. Other AAM fitting algorithms can be found in [6].

We propose a novel AAM fitting method that pre-estimates the change of the shape (motion) of an object using the active contour technique and then begins

* This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University.

D. Zhang and A.K. Jain (Eds.): ICB 2006, LNCS 3832, pp. 159–165, 2005.

existing AAM fitting algorithm using the motion compensated parameters. In this work, a CONDENSATION-like [7] active contour technique has been used to estimate the object contour effectively, thus accurately estimating the motion of the object in the image sequence.

The remainder of this paper is organized as follows. In section 2, we briefly review the original AAM fitting algorithm and active contour technique. In Section 3, we explain how the active contour technique can be incorporated into the AAM fitting algorithm to make it robust to the large motion. In section 4, experimental results are presented. Finally, we draw a conclusion.

2 Theoretical Backgrounds

2.1 Active Appearance Models

In 2D AAMs [1], [4], the 2D shape $\mathbf{s}$ of an object is represented by a triangulated 2D mesh and it is assumed that the varying shape can be approximated by a linear combination of a mean shape $\mathbf{s}_0$ and orthogonal shape bases $\mathbf{s}_i$ as

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

where p_i are the shape parameters and $\mathbf{s} = (x_i, y_1, ..., x_l, y_l)^T$. The appearance is defined in the mean shape $\mathbf{s}_0$ and the appearance variation is modeled by a linear combination of a mean appearance A_0 and orthogonal appearance bases A_i as

$$A = A_0 + \sum_{i=1}^{m} \alpha_i A_i, \tag{2}$$

where α_i are the appearance parameters and A_i represents the vectorized appearance. To build an AAM, we need a set of landmarked training images. The shape and appearance bases are computed by applying PCA to the shape and appearance data that are collected and normalized appropriately.

Using an 2D AAM, the shape-variable appearance of an object in the image can be represented by

$$M(W(\mathbf{x}; \mathbf{p}')) = \sum_{i=0}^{m} \alpha_i A_i(\mathbf{x}), \tag{3}$$

where W is a coordinate transformation function from the coordinate $\mathbf{x}$ in the template image frame to the coordinate of the synthesized image frame. The parameters of the warping function are represented by $\mathbf{p'}^T = (\mathbf{p}^T, \mathbf{q}^T) = (p_1, \ldots, p_n, q_1, \ldots, q_4)$, where $\mathbf{p}$ and $\mathbf{q}$ determine the varying 2D shape of the object and its similar transformation, respectively. Four similar transformation parameters q_1, q_2, q_3, and q_4 describe the scale, rotation, horizontal and vertical translation of the shape, respectively.

2.2 The AAM Fitting Algorithm

The problem of fitting a 2D AAM to a given image can be formulated as finding the appearance and shape parameters of an AAM that minimizes the following error

$$E = \sum_{\mathbf{x}\in\mathbf{s}_0} \left[\sum_{i=0}^{m} \alpha_i A_i(\mathbf{x}) - I(W(\mathbf{x};\mathbf{p}')) \right]^2 . \tag{4}$$

Among various gradient based fitting algorithms, we will briefly review the Inverse Compositional Simultaneous Update algorithm (SI), which is known to have the best convergence performance and the Inverse Compositional Normalization algorithm (NO), which is more efficient than the SI algorithm. The SI algorithm is derived by applying the Taylor expansion with respect to the both shape and appearance parameters. The update of model parameters $\Delta\boldsymbol{\theta}^T = \{\Delta\mathbf{p}'^T, \Delta\boldsymbol{\alpha}^T\}$ are computed as

$$\Delta\boldsymbol{\theta} = \left\{ \sum_{\mathbf{x}\in\mathbf{s}_0} SD^T(\mathbf{x}) SD(\mathbf{x}) \right\}^{-1} \sum_{\mathbf{x}\in\mathbf{s}_0} SD^T(\mathbf{x}) E(\mathbf{x}) \tag{5}$$

$$SD(\mathbf{x}) = \left[\nabla A(\mathbf{x};\boldsymbol{\alpha})^T \frac{\partial W}{\partial \mathbf{p}'}, A_1(\mathbf{x}), \ldots, A_m(\mathbf{x}) \right], \tag{6}$$

where $SD(\mathbf{x})$ represents the steepest descent vector of the model parameters θ. The warping parameters and appearance parameters are updated as $W(\mathbf{x};\mathbf{p}') \leftarrow W(\mathbf{x};\mathbf{p}') \circ W(\mathbf{x};\Delta\mathbf{p}')^{-1}$, and $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}$, respectively. The SI algorithm is inefficient because $SD(\mathbf{x})$ in (5) depends on varying parameters and must be recomputed at every iteration. The inverse compositional normalization algorithm (NO) makes use of the orthogonal property of appearance bases. This orthogonal property enables the error term in (4) to be decomposed into sums of two squared error terms:

$$\left\| A_0 + \sum_{i=1}^{m} \alpha_i A_i - I^W(\mathbf{p}') \right\|^2_{span(A_i)} + \left\| A_0 + \sum_{i=1}^{m} \alpha_i A_i - I^W(\mathbf{p}') \right\|^2_{span(A_i)^\perp}, \tag{7}$$

where $I^W(\mathbf{p}')$ means vector representation of backward warped image. The first term is defined in the subspace $span(A_i)$ that is spanned by the orthogonal appearance bases and the second term is defined in the subspace $span(A_i)^\perp$, orthogonal complement subspace. For any warping parameter $\mathbf{p}'$, the minimum value of the first term is always exactly 0. Since the norm in the second term only considers the component of the vector in the orthogonal complement of $span(A_i)$, any component in $span(Ai)$ can be dropped. As a result, the second error term can be optimized efficiently with respect to $\mathbf{p}'$ using an image matching algorithm such as the inverse compositional algorithm [6].

Robust fitting algorithms use the weighted least squares formulation that includes a weighing function into its error function. Weighted least squares formulation can be applied to NO algorithm to make it robust. Detailed derivations and explanations can be found in the [4].

2.3 Active Contour Techniques

In this paper, we locate the foreground object using a CONDENSATION-like contour-tracking technique which is based on probabilistic sampling.

A contour $\mathbf{c}$ of an object is represented by a set of boundary points $\mathbf{c} = (x_1, y_1, ..., x_v, y_v)^T$. We can represent all the possible contours within a specified contour space by a linear equation as

$$\mathbf{c} = \mathbf{c}_0 + S\mathbf{y}, \tag{8}$$

where $\mathbf{c}_0$ is the mean contour, S is a shape matrix that is dependent on the selected contour space and $\mathbf{y}$ is a contour parameter vector [8].

The CONDENSATION method [8] aims to estimate the posterior probability distribution $p(\mathbf{y}|\mathbf{z})$ of the parameter vector $\mathbf{y}$ in the contour space $S_\mathbf{y}$ using a factored sampling, where $\mathbf{z}$ denotes the observations from a sample set. The output of a factored sampling step in the CONDENSATION method is a set of samples with weights denoted as $\{(\mathbf{s}_1, \pi_1), (\mathbf{s}_2, \pi_2), ..., (\mathbf{s}_N, \pi_N)\}$, which approximates the conditional observation density $p(\mathbf{y}|\mathbf{z})$. In the factored sampling, a sample set $\{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N\}$ are randomly generated from the prior density $p(\mathbf{y})$ and then the weights π_i of the N generated samples are computed by

$$\pi_i = \frac{p_z(\mathbf{s}_i)}{\sum_{j=1}^{N} p_z(\mathbf{s}_j)}, \tag{9}$$

where $p_z(\mathbf{s}) = p(\mathbf{z}|\mathbf{y} = \mathbf{s})$ is the conditional observation density.

In this work, we measured $p(\mathbf{z}|\mathbf{y})$ using a fitness evaluation function that consider the quality of the image edge features found in the image and the distance between the contour sample and the image edge features as

$$p(\mathbf{z}|\mathbf{y}) \propto n_f \frac{\bar{s}_f}{\sigma_s \bar{d}_f}, \tag{10}$$

where n_f is the number of edge features that have found within a given search range along the normal direction of the contour, $\bar{s}_f$ and $\bar{d}_f$ are the mean magnitude of edge gradient and the mean distance of the n_f image edge features, and σ_s is used to compensate the different scales of the edge gradient and the distance.

3 Active Contour Based AAM

We apply the following two stages alternatively in order to track the face image. During stage I, we perform the active contour technique to find the contour sample that best fits the face image as follows,

1. Make the base shape c_0 and the shape matrix S in (8) using the fitted shape of the AAM at (t-1)-th image frame.
2. Generate N random samples $\{\mathbf{s}_1 \dots \mathbf{s}_N\}$ that are located near the computed contour $\mathbf{c}$

3. Evaluate the fitness of all generated samples using the conditional observation density function $p(\mathbf{z}|\mathbf{y})$ explained in section 2.3.
4. Choose the best sample $\mathbf{s}_{best}$ with the highest fitness value among N samples. We estimate the motion parameter $\hat{\mathbf{q}}^t$ at the next image frame t by composing two similar transformations $\mathbf{q}^{t-1}$ and $\Delta\hat{\mathbf{q}}^t$, where $\Delta\hat{\mathbf{q}}^t = \mathbf{s}_{best}$.

During stage II, we perform the active appearance model fitting algorithm over the best selected contour $\mathbf{s}_{best}$ as follows,

1. Run the AAM fitting algorithm using the shape parameter $\mathbf{p}^{t-1}$, the appearance parameter $\boldsymbol{\alpha}^{t-1}$, and the estimated motion parameter, $\hat{\mathbf{q}}^t$.
2. Obtain the optimal AAM model parameters $\mathbf{p}^t$, $\mathbf{q}^t$, and $\boldsymbol{\alpha}^t$.
3. Set the image frame index $t = t - 1$, and return to stage I until reaching the final frame.

4 Experimental Results

4.1 Comparison of Fitting Performances of Different AAM Methods

We compared the accuracy of three different AAM fitting methods such as the existing robust AAM (R-AAM), the proposed active contour based AAM (AC-AAM), and a combination of the two methods. For each methods, we measured the performances using two different types of parameter updates [6] such as the normalization method (NO-update) and the simultaneous update method (SI-update). The left and right figures of Fig 1 show the results from NO-update and SI-update, respectively. The top row of Fig. 1 shows the decreasing RMS error as the fitting algorithm is iterated, where the RMS error is defined as the mean distance between the ground truth shape points and the corresponding points of the current fitted shape. In each picture, the horizontal and vertical axis denotes the iteration index and the RMS error, respectively. Two curves are represented for each AAM method, corresponding to two differently perturbed AAM shapes, respectively. Each point over the curve is the average value of the RMS errors of 100 independent trials. Figure 1 shows that 1) the contour combined AAM fitting is converged within 5 iterations in most cases, 2) the fitting of the R-AAM method is not effective when the initial displacement is great, and 3) the proposed AC-AAM has a good convergence accuracy even if there is a great initial displacement. We also compared the convergence rate of the three different AAM fitting methods, where the convergence rate is defined by the ratio of convergence cases to all trials. The bottom row of Fig. 1 show the convergence rate where each point in the figure is the average convergence rate of 100 trials. Figure 1 shows that the difference of convergence rate between R-AAM and AC-AAM becomes larger as the initial displacement error increases, which implies that the proposed AC-AAM is more effective when the AAM shape is placed far from the target face. In the above experiments the combined AC-R-AAM shows the best convergence performance.

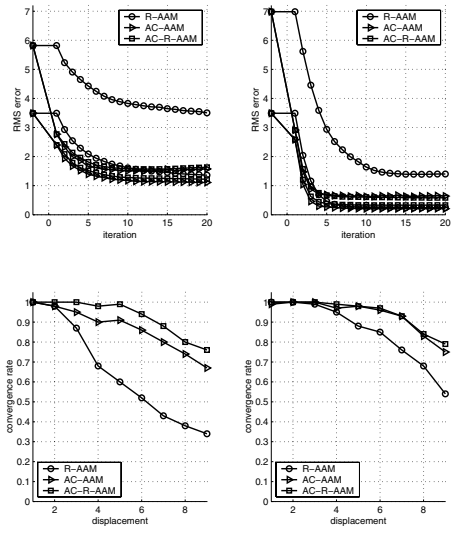

Fig. 1. Convergence characteristics of two different updates

4.2 Comparison of Execution Times Between Difference AAM Methods

Figure 2 shows the average number of iterations of the different methods, where the horizontal and the vertical axes denote the average number of iterations and the displacement σ, respectively. Each point represents the average number of iterations of independent successfully converged trails when the same stop condition if applied. From the Fig. 2, we note that the average number of iterations of AC-AAM and AC-R-AAM are almost constant even the displacement σ increases, while those of R-AAM increased rapidly as the displacement σ increases.

We measured the execution time of the different methods in our C implementation. It took about 5 msec for the active contour fitting when 50 samples, 51 contour points, and 10 pixels of search range were considered. Also, it took

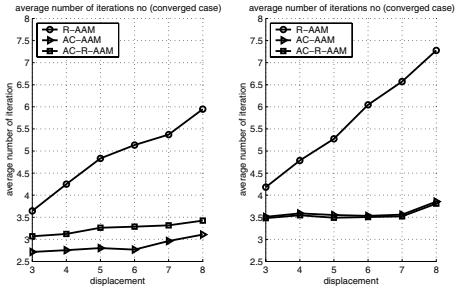

Fig. 2. Comparison of the number of iterations of three different AAM methods

about 8 msec and 26 msec for the NO-update and SI-update, respectively, in the robust AAM and it takes about 4 msec and 23 msec for the the NO-update and SI-update, respectively, in the the proposed AC-AAM.

5 Conclusion

In this paper, we proposed an active contour combined AAM fitting algorithm that is robust to a large motion of an object. Although the existing robust AAM can cope with the mismatch between currently estimated AAM instance and an input image, it does not converge well when the motion of the face is large. This comes from the fact that only the small part of the backward warped image may be used to estimate the update of parameters and it is not sufficient for correct estimation. The proposed AAM fitting method was robust to a large motion of the face because it rapidly locates the AAM instance to an area close to the correct face position. The proposed AAM fitting method was also fast because the active contour technique can estimate the large motion of the face more chiefly than the AAM fitting algorithm.

We performed many experiments to evaluate the accuracy and convergence characteristics in terms of RMS error and convergence rate, respectively. The combination of the existing robust AAM and the proposed active contour based AAM (AC-R-AAM) showed the best accuracy and convergence performance.

References

1. T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, issue 6, pp. 681–685, 2001.
2. G.J. Edwards, C.J. Taylor, T.F. Cootes, "Interpreting Face Images Using Active Appearance Models," *Proc. of IEEE 3rd International Conference on Automatic Face and Gesture Recognition*, vol.0, pp. 300, 1998.
3. G.J. Edwards, T.F. Cootes, C.J. Taylor, "Face Recognition Using Active Appearance Models,", *Proc. of 5th European Conference on Computer Vision*, vol. 2, pp. 581, June 1998.
4. S. Baker and I. Matthews, "Active Appearance Models Revisited," CMU-RI-TR-03-01, CMU, Apr 2003.
5. B.D. Lucas T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. of International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.
6. I. Matthews, R. Gross, and S. Baker, "Lucas-Kanade 20 Years on: A Unifying Framework: Part 3," CMU-RI-TR-03-05, CMU, Nov 2003.
7. M. Isard and A. Blake, "CONDENSATION-Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, vol. 29, pp. 5–28, 1998.
8. M. Isard and A. Blake, *Active Contours*, Springer, 1998.