

Enhanced Resolution-Aware Fitting Algorithm using Interpolation Operator

Jongju Shin and Daijin Kim

Dept. of Computer Science and Engineering, POSTECH

San 31 Hyoja-dong, Pohang, Republic of Korea

{jjshin, dkim}@postech.ac.kr

Abstract

This paper proposes a new method to speed up the Resolution-Aware Fitting (RAF) algorithm. An interpolation operator is used instead of a blur operator in the RAF algorithm. The RAF algorithm with the interpolation operator is twice as fast as the RAF algorithm with the blur operator without losing fitting accuracy. It concluded that the RAF algorithm with the interpolation operator is superior to the RAF algorithm with the blur operator.

1. Introduction

The feature extraction is to extract the important components from images for further analysis. Active Appearance Model (AAM) [1] is widely used in feature extraction. AAM is a generative model that describes variations in shape and appearance of objects. For AAM, several fitting algorithms have been proposed such as the Simultaneous algorithm [3] and the Project Out algorithm [4]. These are iterative algorithms based on the gradient-descent method. The goal of these algorithms is to find the shape and the appearance parameters which minimize the difference between the model instance and the input image. The Simultaneous algorithm updates the shape and appearance parameters simultaneously in each iteration. The Project Out algorithm updates the shape parameter in each iteration, and computes the appearance parameter by projection. Of these algorithms, the Simultaneous algorithm is more accurate, but the Project Out algorithm is faster. This comparison is based on the assumption that the resolution of the input image and the model are similar. When the object's size in the input image is similar to the AAM model, conventional AAM algorithms like the Simultaneous algorithm and the Project Out algorithm work well. But if the object's size in the input image is much smaller than that of the AAM model of that ob-

ject, they work poorly because they don't consider the resolution of the input image.

The Resolution-Aware Fitting (RAF) algorithm [2] has been proposed to improve AAM's performance when the input image has lower resolution than the model. Unlike the conventional AAM algorithm, the difference between the model instance and the input image is computed in the coordinates of the input image using the blur operator [2]. Because the RAF algorithm considers the resolution of the input image, the RAF algorithm works better than the conventional AAM algorithm. In the conventional AAM algorithm, the jacobian and the gradients of the input image are precomputed, but in the RAF algorithm, the jacobian and the gradients are computed using blur operator in each iteration. The RAF algorithm takes about 10 times longer than the Simultaneous algorithm because the RAF algorithm doesn't have the benefit of the precomputations.

We propose an RAF algorithm with interpolation to speed up the RAF algorithm and show similar results. The interpolation operator is simpler than the blur operator, and operates similarly to the blur operator.

Section 2 describes the theoretical background, Section 3 describes the experiments, and Section 4 concludes this paper.

2. Theoretical Background

2.1 Active Appearance Model

AAM consists of two models, the shape and the appearance of an object like the face. The shape model is the combination of the basis shapes with the mean shape.

$$\mathbf{s}(\mathbf{p}) = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i \quad (1)$$

where \mathbf{s}_0 is the mean shape, \mathbf{s}_i is the basis shapes, and p_i is the parameter of the each basis shape. The shape \mathbf{s} is defined by the shape parameter $\mathbf{p} = (p_1, p_2, \dots, p_n)$.

The appearance model also consists of the mean appearance and the basis appearances. The appearance model is built on the coordinates of the mean shape s_0 .

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

where A_0 is the mean appearance, A_i is the basis appearances, and λ_i is the parameter of the each basis appearance. The appearance is defined by the appearance parameter $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$.

2.2 Conventional AAM equation

The model instance is generated by the shape and the appearance parameter. When the input image I is given, the shape and the appearance parameters are computed by equation (3):

$$\sum_{\mathbf{x} \in s_0} \left[I(W(\mathbf{x}; \mathbf{p})) - A(\mathbf{x}; \lambda) \right]^2, \quad (3)$$

where \mathbf{p} is the shape parameter, λ is the appearance parameter, \mathbf{x} is a pixel of the s_0 , and $W(\mathbf{x}; \mathbf{p})$ is the warping function. We use the similarity transform as the warping function. The similarity transform consists of the scale, the rotation, and two translations :

$$W(\mathbf{x}; \mathbf{p}) = s \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (4)$$

where s is the scale, a, b are the rotation, t_x is the translation of the x axis, and t_y is the translation of the y axis.

Equation (3) describes the difference between the appearance instance $A(\mathbf{x}; \lambda)$ and the warped input image $I(W(\mathbf{x}; \mathbf{p}))$ by the parameter \mathbf{p} . The input image is warped to the coordinates of the model s_0 . The shape and the appearance parameters can't be computed directly because equation (3) is highly nonlinear. The iterative method is used to update the parameters. At each iteration, $\Delta \mathbf{p}$ and $\Delta \lambda$ are updated to the current \mathbf{p} and λ .

2.3 Blur operator equation

The conventional AAM algorithm shows lower fitting accuracy when the resolution of the input image is lower than the resolution of the model. We know the input image is warped into the coordinates of model s_0 during fitting. If the size of the face is smaller than that of the model, the interpolation is needed to enlarge the face of the input image to the coordinates of the model. This makes the error increase.

To solve this low resolution problem, Resolution-Aware Fitting (RAF) algorithm is proposed [2]. The fitting criterion of RAF algorithm is

$$\sum_{\mathbf{u} \in I} \left[I(\mathbf{u}) - B(A(W(\mathbf{x}; \mathbf{p}); \lambda); \mathbf{u}) \right]^2, \quad (5)$$

where \mathbf{p} is the shape parameter, λ is the appearance parameter, B is the blur operator, \mathbf{u} is a pixel in the shape area of the input image, and \mathbf{x} is a pixel of the model. Summation of equation (5) is computed for the coordinates of the input image. This is the main difference between the conventional AAM algorithm and the RAF algorithm. The blur operator transforms the model instance into the coordinates of the input image I :

$$B(W(A(\mathbf{x}; \lambda); \mathbf{p}); \mathbf{u}) = \frac{1}{\text{area}(\mathbf{u})} \sum_{\substack{\mathbf{x} \in s_0 \text{ s.t.} \\ \mathbf{u} - \begin{bmatrix} .5 \\ .5 \end{bmatrix} < W(\mathbf{x}; \mathbf{p}) < \mathbf{u} + \begin{bmatrix} .5 \\ .5 \end{bmatrix}}} A(\mathbf{x}; \lambda) |J(W(\mathbf{x}; \mathbf{p}))|. \quad (6)$$

Each pixel (\mathbf{x}) of the model instance is transformed to sub-pixel location within the coordinates of the input image using $W(\mathbf{x}; \mathbf{p})$. The transformed location is around \mathbf{u} and is not integer. Our goal is to estimate the intensity of the each pixel (\mathbf{u}) of the model instance on the coordinates of the input image. To estimate the intensity of the each pixel (\mathbf{u}), we need to determine which pixels of the model instance are transformed around each pixel of the transformed model instance:

$$\mathbf{u} - \begin{bmatrix} .5 \\ .5 \end{bmatrix} < W(\mathbf{x}; \mathbf{p}) < \mathbf{u} + \begin{bmatrix} .5 \\ .5 \end{bmatrix}. \quad (7)$$

After that, we compute the weighted mean of intensities of selected pixels. This result is the estimated intensity of \mathbf{u} . The steepest descent images by (5) are not only re-computed using blur operator in each iteration, and also Hessian is. The detail derivation is in [2].

2.4 Interpolation operator equation

The RAF algorithm computes the error term in the coordinates of the input image. To transform the model instance into the coordinates of the input image, the RAF algorithm uses the blur operator. We propose to use interpolation to transform the model instance into the coordinates of input image to reduce the fitting time. The fitting criterion of the RAF algorithm with interpolation is

$$\sum_{\mathbf{u} \in I} \left[I(\mathbf{u}) - A(W(\mathbf{u}; \mathbf{p}); \lambda) \right]^2. \quad (8)$$

Equation (8) is made by removing the blur operator from equation (5). The other difference between the equation (5) and the equation (8) is the direction of the warping function. The model instance A is warped back into the coordinates of the input image I using $A(W(\mathbf{u}; \mathbf{p}); \lambda)$. If each transformed location $W(\mathbf{u}; \mathbf{p})$ is not an integer, interpolation is needed to estimate the intensity at that location. If we use a bilinear interpolation, four corners around the transformed sub-pixel location are used. Because the interpolation operator is simpler than the blur operator, the execution time of the interpolation operator is shorter than that of the blur operator. To minimize the fitting criterion (8) with respect to \mathbf{p} and λ , we use gauss-newton gradient-descent algorithm. The appearance model A is represented as a weighted sum of appearance bases:

$$\sum_{\mathbf{u} \in I} \left[I(\mathbf{u}) - A_0(W(\mathbf{u}; \mathbf{p})) - \sum_i \lambda_i A_i(W(\mathbf{u}; \mathbf{p})) \right]^2.$$

The fitting criterion with first order Taylor expansion is

$$\sum_{\mathbf{u} \in I} \left[I(\mathbf{u}) - A_0(W(\mathbf{u}; \mathbf{p})) - \sum_i \lambda_i A_i(W(\mathbf{u}; \mathbf{p})) - SD \begin{pmatrix} \Delta \mathbf{p} \\ \Delta \lambda \end{pmatrix} \right]^2,$$

where

$$SD = \left[(\nabla A_0 + \sum_i \lambda_i \nabla A_i) \frac{\partial W}{\partial p_1}, \dots, (\nabla A_0 + \sum_i \lambda_i \nabla A_i) \frac{\partial W}{\partial p_n}, A_1(W(\mathbf{p})), \dots, A_m(W(\mathbf{p})) \right].$$

Minimum is given by

$$\begin{pmatrix} \Delta \mathbf{p} \\ \Delta \lambda \end{pmatrix} = -H^{-1} \sum_{\mathbf{u} \in I} SD^T \left[I(\mathbf{u}) - A_0(W(\mathbf{u}; \mathbf{p})) - \sum_i \lambda_i A_i(W(\mathbf{u}; \mathbf{p})) \right]$$

, where H is Hessian matrix:

$$H = \sum_{\mathbf{u} \in I} SD^T SD.$$

3 Experiments

3.1 Experimental Conditions

3.1.1 Training images

We photographed four subjects, each under three kinds of illumination from nine directions(face-on, from directly above, below, right, and left, and from 15° above



Figure 1. The example of training images : nine directions of pose of the subject.

and below horizontal from the right and left)(Fig 1). The total number of photographs was 108. All were taken at a distance of 0.7m and were 640*480 pixels in size. All 108 face images are used to construct the face model. To test various sizes of face model, we constructed several models with different size. Models' widths are 100, 120, 140, 160, 180, and 200 pixels.

3.1.2 Test images

Tests were performed on 27 images, photographed at a distance of 0.7 m. All the images were 640*480 pixel in size. All test images are different from images used to build the AAM model. To reduce the resolution of the input images, all the test images were reduced to $\frac{1}{15}$ of their original size. Other resize factors should be tested in the future.

3.1.3 Metrics of Fitting Results

We compared the fitting accuracy and the speed of the conventional AAM algorithm, the original RAF algorithm that uses the blur operator, and the RAF algorithm that uses the interpolation operator. The fitting accuracy was measured by the tracking error and the speed was measured by the execution time.

- **Tracking error** : All the test images have 70 ground truth points on the face. The ground truth points are inserted by hand. After fitting the model on the low resolution test image, the shape is estimated using equation (1) with the the shape parameter \mathbf{p} . Because the estimated shape is on the coordinates of the input image which is resized by $\frac{1}{15}$ of

their original size, and the ground truth points are on the coordinates of the original input image, the estimated shape should be enlarged by a factor of 15. The RMS error between the estimated shape and the ground truth points is computed. Lower RMS error corresponds to better fitting accuracy.

- **Execution time** : Execution time is the time required 40 iterations. It is measured five times for each image using each model, and averaged.

3.2 Experimental Result

The average RMS error of RAF algorithm with interpolation is less clearly than the conventional algorithm and the RAF algorithm (Fig 2). The RAF algorithm with interpolation has smaller RMS error than the RAF algorithm regardless of the model width.

The RAF algorithm with interpolation has faster average execution time than the conventional algorithm and the RAF algorithm (Table 1).

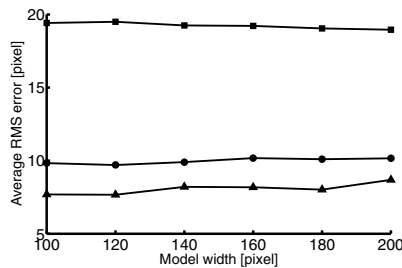


Figure 2. Average RMS error over the model width among the three algorithms : the conventional AAM algorithm(■), the RAF algorithm(●), and the RAF algorithm with interpolation (▲)

Table 1. The average±s.d. execution time (n=10) for the three algorithms. Computer: Pentium D 3.2GHZ with 2GB memory. The number of iterations : 40.

Algorithm	Average ±s.d. execution time (sec)
AAM	4.549±0.214
RAF	53.80758±11.111
RAF with interpolation	21.540±0.598

4 Conclusion

In this study, we improved the performance of the RAF algorithm by using an interpolation operator. When images have low resolutions, the RAF algorithm with interpolation is faster than the RAF algorithm and the conventional AAM algorithm (Table 1). The average RMS error of the RAF algorithm with interpolation is less than those of the conventional AAM algorithm and the RAF algorithm (Fig 2). This result indicates that the RAF algorithm with interpolation is more able to process low resolution images than the conventional algorithm and the RAF algorithm. The reason for this result is that in low resolution images, the RAF algorithm with interpolation creates better model instances than the RAF algorithm. Because the test images are some of the training images, experiments using other test images are required. The experiments in this paper are performed at downsampling factor 15 to simulate low resolution images. Experiments which compare the performance of the algorithms under conditions of : 1) the various downsampling factors; and 2) the various distances in the real situation are required.

5 Acknowledgment

This work was partially supported by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy (MKE). Also, it was partially supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University (R112002105070030(2008)).

References

- [1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407:484–498, 1998.
- [2] G. Dedeoglu, T. Kanade, and S. Baker. The asymmetry of image registration and its application to face tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):807–823, May 2007.
- [3] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(11):1080–1093, November 2005.
- [4] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135 – 164, November 2004.