2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Multi-resolution-type AAM for Low-resolution Images

Satoru Suzuki[†1] and Yasue Mitsukura[†1]

†1 Faculty of Science and Technology, Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa, Japan
Email: 50009701291@st.tuat.ac.jp, mitsukura@sd.keio.ac.jp

Abstract—The purpose of this paper is to extract small faces from images. When a mobile robot tries to find a person from a far distance, the image resolution will be degraded largely, which makes face tracking difficult. Therefore, we present the multi-resolution-type active appearance model (AAM) robust to degraded facial images. In the proposed method, both high-resolution and downsampled low-resolution images are used together to train an AAM. The shape matching performance of multi-resolution-type AAM is evaluated by using images from low-resolution to high-resolution. The simulation results show that multi-resolution-type AAM has the possibility for coping with variety of image resolutions.

1. Introduction

Face tracking is an important task for human-machine interface (HMI), such as facial expression recognition, intention acquisition, and gesture recognition. However, face tracking is an even difficult problem. The changes of facial shape and texture caused by the movements of head and landmarks such as eyes and mouth, occlusion, and lighting conditions make it difficult to trace faces. If these problems are solved, HMI which enables successful interaction between robots and people will be achieved.

Skin color tracking [1] is the simplest and widely used approach to trace faces which have arbitrary sizes and shapes. However, it fails to trace faces in the cluttered scenes, and the accuracy to estimate facial positions are not good. On the other hand, template-based approach [2] which uses whole face is good at estimating the position of faces precisely. The drawback of this approach is that it requires a lot of templates to trace varied faces with rigid and non-rigid changes. Feature-based approach [2] which uses facial features such as eyes, nose, and mouth effectively traces faces regardless of the rigid and non-rigid changes. But it has some difficulties to extract each feature from images in the first place, and occlusion is one of the problems for feature extraction. Model-based approach like active appearance model (AAM) [3] [4] has a characteristic that it can handle rigid and non-rigid objects and localize the position precisely by matching the model with input images. AAM is a statistical model, and the shape and grayscale appearance of the objects are represented with low-dimensional features. The advantage of using AAM for face tracking is the adaptation to the different shape and texture of faces. However, the performance of AAM degrades when small faces which are represented with low-resolution appeares on the image. This is because that AAM is only trained with original size images.

Basic idea for tracking low-resolution faces is to use interpolation to increase the resolution of input images. Although the interpolation is thought as easy and useful solution to the low-resolution image, it involves calculation error. In order to overcome the problem of face tracking on low-resolution images, several methods have been introduced. Camera model for image formation proposed by Golse et al. can be used to trace faces by adjusting the resolution of AAM to be as same level as given images [5]. On the other hand, the pyramid-model proposed by Xiaomin et al. prepares several AAMs built by different resolution images [6]. Firstly, original images are used to build an AAM. Then, the original images are downsampled and they are utilized to build another AAM for lower resolution face. Every AAM is applied to an input image and their matching performances are compared with each other, and then the best AAM is selected from trained AAMs. The drawback of pyramid-model is that we require multiple AAMs, and applying all models to images is time consuming.

In this paper, we focus on the extracting small faces from images. In our previous study, face tracking is utilized to detect and trace a person by mobile robot [7]. Although face was traced well in the short distance from the robot to person, it became difficult in the long distance because of the degradation of image resolution. In order to overcome the problem of face tracking difficulty in the low-resolution, we present the multi-resolution-type AAM robust to degraded facial images. Our method contains of the following characteristics.

- Adaptation to arbitrary image resolutions.
- Requiring single AAM.

The multi-resolution-type AAM is modeled with the dataset which contains images which have several resolutions. Thereby, the model is expected to cover the variety of resolutions from low-resolution to high-resolution images. Comparing with pyramid-type AAM [6], our proposed method only requires single AAM. As a first step for face tracking in a video, we use images and evaluate the performance of the multi-resolution-type AAM. In the computer simulations of face matching, it was shown that

multi-resolution-type AAM has the possibility for coping with variety of image resolutions.

2. Training of AAM

AAM is consisted of shape and grayscale appearance model, and they are trained with facial images. The training of AAM begins with collecting facial images and giving landmarks to them manually. Fig. 1 shows the example of facial image with landmarks. In this paper, 26 landmarks are used to describe a facial shape. The set of landmarks is represented by vector $\mathbf{x} = (x_1, y_1, x_2, y_2, \cdots, x_N, y_N)^T$ and (x_i, y_i) is corresponding to each landmark i. Then facial shape $\mathbf{x}$ is normalized with orientation, scale, and translation. The procedure of the shape normalization is as follows [8]:

1. Rotate, scale, and translate each shape to align with the first shape in the set.
2. Repeat (a)-(c) until the process converges.
   (a) Calculate the mean shape from the aligned shapes.
   (b) Normalize the orientation, scale and origin of the current mean to suitable defaults.
   (c) Realign every shape with the current mean shape.

In this paper, step 2 was repeated 10 times and finally the shape error converged to 0. Then, we apply principal component analysis (PCA) to normalized shapes. Any shape $\mathbf{x}$ is approximated as follows:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s, \tag{1}$$

where $\bar{\mathbf{x}}$ is a mean shape, $\mathbf{P}_s$ is a shape bases, and $\mathbf{b}_s$ is a shape parameters, respectively. After building shape model, we build grayscale appearance model. To build the appearance model, we transform each image so that its shape matches the mean shape. This processing is called warp. Firstly, each shape is divided into several triangular shapes generated from labelled landmarks using delaunay triangulation algorithm (Fig. 2). Then each image is warped by piecewise affine transform. After warping images, we extract grayscale appearance within the covered area by the mean shape from the warped image. We apply PCA to extracted images, and any appearance $\mathbf{g}$ is approximated as follows:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g, \tag{2}$$

where $\bar{\mathbf{g}}$ is a mean grayscale appearance, $\mathbf{P}_g$ is an appearance bases, and $\mathbf{b}_g$ is a grayscale appearance parameters, respectively. Now, achieving better matching with low-resolution image, we prepare donwsampled image of the original image and use them for building grayscale appearance model. The original image is downsampled with the ratio of 0.75, 0.5, and 0.25 to reduce its resolution. Downsampled images are then upsampled to arrange image size as same as original one. We use bilinear interpolation for donwsampling and upsampling. We show the example of the downsampled images in Fig. 3. This downsampling simulates the degradation of image resolution caused by the distance between person and camera. Original images (high-resolution images) are corresponding to the image obtained at a short distance, and downsampled images (low-resolution images) are corresponding to the image obtained at a long distance. The appearance model is calculated by applying PCA to original and downsampled images. On the other hand, shape model is trained with the set of landmarks of original images because facial shapes are invariant to image resolution.

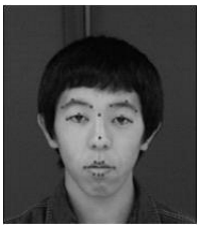

Figure 1: Facial landmarks labeled manually.

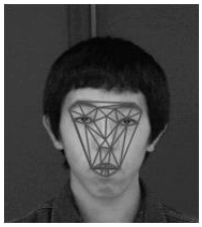

Figure 2: Delaunay triangulation of an image.

3. AAM Matching

The problem of face tracking is to minimize the grayscale appearance error between an AAM and an input image. The objective function is defined as following equations.

$$E = \delta \mathbf{I}^2 \tag{3}$$

$$\delta \mathbf{I} = I\Big(\mathbf{W}(\mathbf{x}; \mathbf{b}_s)\Big) - A(\mathbf{x}; \mathbf{b}_g) \tag{4}$$

Where $\mathbf{W}(\mathbf{x}; \mathbf{b}_s)$ represents the warp which translates the shape of input image to mean shape. $I\Big(\mathbf{W}(\mathbf{x}; \mathbf{b}_s)\Big)$ denotes a pixel value corresponding to warped image, and $A(\mathbf{x}; \mathbf{b}_g)$ denotes a pixel value of grayscale appearance model. When we minimize the objective function, the parameter of AAM should be arranged so as to match the

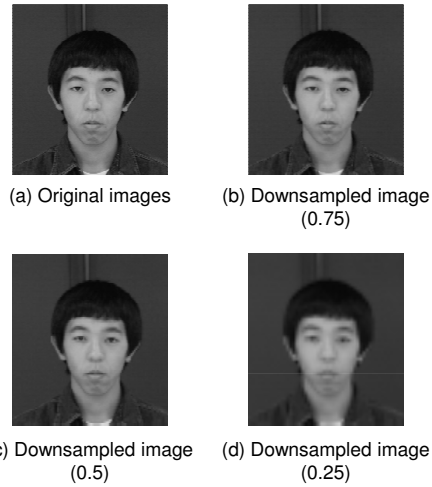

Figure 3: Original and downsampled images.

model with input image. Therefore, we calculate the relationship between the difference of the grayscale appearance $\delta\mathbf{I}$, and parameter shift $\delta\mathbf{b}_s$ and $\delta\mathbf{b}_g$. This relationship is approximated by linear equation. For simplifying the equation, we combine $\delta\mathbf{b}_s$ with $\delta\mathbf{b}_g$ into $\delta\mathbf{c}$.

$$\begin{pmatrix} \delta\mathbf{b}_s \\ \delta\mathbf{b}_g \end{pmatrix} = \delta\mathbf{c} = \mathbf{T}\delta\mathbf{I} \tag{5}$$

To find $\mathbf{T}$, we apply multiple multivariable linear regression to the shift parameter $\delta\mathbf{c}$ and corresponding difference $\delta\mathbf{I}$. The AAM matching is conducted by calculating $\delta\mathbf{c}$ from $\delta\mathbf{I}$.

$$\mathbf{c} = \mathbf{c}_0 + k\delta\mathbf{c} \tag{6}$$

The parameter $\delta\mathbf{c}$ is used to update the current parameter $\mathbf{c}_0$ and k is a scaling coefficient.

4. Computer Simulations

4.1. Simulation Conditions

The shape matching performance of the multi-resolution-type AAM is evaluated by using facial dataset containing multi-resolution images. We pick out 40 frontal face images from the HOIP database, and reduce the resolution to 320x240. We call reduced images as original images. Then each original image is downsampled with the ratio of 0.75, 0.5, and 0.25. From these images, we create train and test dataset. We examine how the use of multi-resolution image for building AAM affects the matching performance to the input images from low-resolution to high-resolution. To build shape and grayscale appearance models, eigenvectors are selected from the highest eigenvalue until exceeding cumulative contribution value of 0.95. We evaluate the matching performance of the model based on the difference between the element of the shape model and that of the correct shape. In this paper, the hand labeled shape is considered as a correct shape. The shape difference is calculated by RMS.

$$E(x, x_{gt}) = \sum_{i=1}^{N} \sqrt{\frac{(x_i - x_{gt,i})^2 + (y_i - y_{gt,i})^2}{N}} \tag{7}$$

x_i and y_i are elements of shape model and $x_{gt,i}$ and $y_{gt,i}$ are that of correct shape. The matching between the model and input image starts in the condition of giving perturbation to the correct shape, and locating the AAM in that position. This simulation is performed using Visual Studio 2005 on a computer with Intel R Core TM 2 Duo CPU (1.20GHz) processor.

4.2. Simulation Results

We built 4 kinds of AAMs such as original image only(Original), original+downsampled image with the ratio of 0.75(Original+0.75), original+donwnsampled image with the ratio of 0.50(Original+0.50), and original+downsampled image with the ratio of 0.25(Original+0.25). We show the results of applying every AAM to the test dataset in Table 1. Top row indicates the test dataset, and left column indicates trained AAMs. This result shows the average and standard deviation of 10 trials. The AAM(Original) has a favorable accuracy to every test dataset. However, when we compare the performance of each model from the point of view of same test dataset, the best model varies depending on the test dataset. In the case of original and 0.75 dataset, AAM(Original+0.50) is the best, and other best results are written in bold type in Table 1. The problem we can see from Table 1 is that standard deviation tends to increase as the resolution of test images decrease. In addition, we obtained the RMSE of more than 50 to the 0.25 dataset, and these samples are removed when we compute the RMSE in Table 1. Then, we show the example of shape matching results in Fig. 4.

4.3. Discussions

We compare resolution specific models with AAM(Original+0.50) selected from the proposed models. Resolution specific models are generated by using each downsampled image only. We show the results of applying each AAM to the test dataset in Table 2. From the comparison of the RMSE in Table 2, AAM(Original+0.50) is better than others on the Original and 0.75 dataset. On the other hand, AAM(0.50) shows the good result on the 0.50 and 0.25 dataset. The performance of AAM(0.25) is quite low to every test dataset. We need to examine the cause of the degradation. From the Table 1 and Table 2, proposed model partially indicates good performance although more improvement is required to decrease the RMSE particularly in lower resolution images.

Table 1: The RMSE of shape matching about proposed models.

	Original	0.75	0.50	0.25
AAM(Original)	4.74 ± 1.28	4.20 ± 1.69	**3.59 ± 1.98**	3.77 ± 3.29
AAM(Original+0.75)	5.20 ± 1.41	4.79 ± 1.68	4.11 ± 1.99	**3.55 ± 2.33**
AAM(Original+0.50)	**4.64 ± 1.18**	**4.20 ± 1.32**	3.71 ± 1.67	3.88 ± 3.19
AAM(Original+0.25)	4.94 ± 1.88	4.62 ± 2.09	3.86 ± 2.79	3.59 ± 2.76

Table 2: The RMSE of shape matching about proposed model and resolution specific models.

	Original	0.75	0.50	0.25
AAM(Original+0.50)	**4.64 ± 1.18**	**4.20 ± 1.32**	3.71 ± 1.67	3.88 ± 3.19
AAM(0.75)	5.64 ± 1.10	4.51 ± 1.17	3.81 ± 1.41	3.58 ± 1.76
AAM(0.50)	4.91 ± 1.22	4.27 ± 1.37	**3.48 ± 1.28**	**3.16 ± 1.51**
AAM(0.25)	8.01 ± 1.24	6.24 ± 1.09	4.76 ± 1.30	4.52 ± 3.41

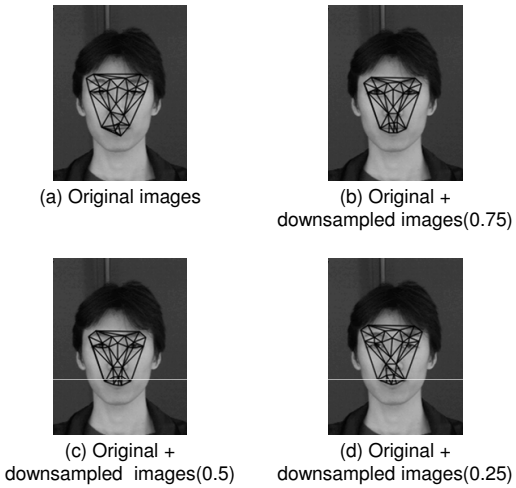

Figure 4: Applying several AAMs to original image.

5. Conclusions

The purpose of this paper is to extract small faces from images. The tracking performance of AAM degrades when the small size face which is represented with low-resolution appears on the image. To solve the matching problems of low-resolution image, we presented multi-resolution-type AAM robust to degraded facial images. We trained AAM by using variety of resolution images, and the matching performance was evaluated in terms of the difference between the shape model and correct shape. From the simulation results, it was shown that building AAM with low-resolution image contributed to improve the performance. Remaining works are improving shape matching performance, expanding to face tracking using image sequence, and head gesture recognition using AAM.

References

[1] Abdullah Bulbul, Zeynep Cipiloglu, and Tolga Capin, "A Color-based Face Tracking Algorithm for Enhancing Interaction with Mobile Devices", *The Visual Computer*, vol.26, No.5, pp.311-323, 2010.

[2] Erik Murphy-Chutorian, and Mohan Manubhai Trivedi, "Head Pose Estimation in Computer Vision: A Survey", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, No.4, pp.607-626, 2009.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models", *5th European Conference on Computer Vision*, vol.2, pp.484–498, 1998.

[4] Iain Matthews, and Simon Baker, "Active Appearance Models Revisited", *International Journal of Computer Vision*, vol.60, No.2, pp.135–164, 2004.

[5] Golsel Dedeoglu, Simon Baler, and Takeo Kanade, "Resolution-Aware Fitting of Active Appearance Models to Low Resolution Images", *Proc. of the 9th European Conference on Computer Vision*, pp.83–97, 2006.

[6] Xiaoming Liu, Peter H. Tu, and Frederick W. Wheeler, "Face Model Fitting on Low Resolution Images", *17th British Machine Vision Conference*, pp.1–10, 2006.

[7] Satoru Suzuki, Yasue Mitsukura, Shin-ichi Ito, Toshio Moriya, Nobutaka Kimura, and Takanari Tanabata, "A Human Tracking System on the Robot with Face Detection", *Journal of Information, International Information Institute*, vol.13, No.1, pp.137–143, 2010.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application", *Computer Vision and Image Understanding*. vol.61, No.1, pp.38–59, 1995.