AAM FITTING USING SHAPE PARAMETER DISTRIBUTION

Youhei Shiraishi, Shinya Fujie and Tetsunori Kobayashi

Waseda University, Okubo 3-4-1, Shinjuku, Tokyo, Japan

ABSTRACT

A novel constraint using shape parameter distribution into the AAM fitting method is proposed. Active appearance models (AAMs) are some of the most popular facial models. AAM-based face tracking delivers accurate alignment results. However, non-face-like shapes can also be estimated by AAMs, unlike by the conventional AAM fitting method, which only minimizes the matching error of the image. This is one of the causes for face tracking performance degradation in AAMs. A constraint using the shape parameter distribution is added in order to solve this problem.

Index Terms— Active appearance models, Inverse compositional image alignment, Face tracking

1. INTRODUCTION

In this paper, a novel face shape likelihood (FSL) constraint is proposed for the conventional fitting algorithm. FSL uses the likelihood of shape parameter distribution.

Facial feature point detection or tracking has been extensively studied as a basic technology for a variety of techniques, including personal identification and facial expression estimation. Methods of detecting and tracking facial feature points can be classified into two categories. The first is a technique using local features. Weighted vector concentration [1] using histograms of oriented gradients (HOG) [2] and Cosar's method [3] using Gabor features are examples in which local features are used. However, methods using local features detect or track feature points by moving feature points independently; therefore, they tend to be unstable for tracking. The second is a technique using a facial appearance model. Typical models are 3D morphable models [4] and active appearance models (AAMs) [5]. The former is a 3D face model while the latter is a 2D face model. Detection and tracking of facial feature points can be performed with high stability and accuracy using these models.

AAMs include shape and appearance models. A shape model uses shape parameters to represent facial shapes. The goal of AAM fitting is to estimate the shape parameter fitting an input facial image. Inverse compositional image alignment (ICIA) [5] is a typical AAM fitting technique. ICIA uses the gradient method to estimate the shape parameters that minimize matching errors between the input image and the mean appearance. Non-face-like shapes can be estimated as there are no shape parameter constraints in ICIA. In this paper, the novel FSL constraint, which uses shape parameter distribution, is proposed.

2. ACTIVE APPEARANCE MODELS

2.1. Model structure

AAMs [5] represent an object's shape with a shape model. $\boldsymbol{s}$ is a vector arranged coordinate of N feature points (x_i, y_i) as follows,

$$\boldsymbol{s} = [\, x_1, y_1, x_2, y_2, \cdots, x_N, y_N \,]^T \tag{1}$$

The mean shape, $\boldsymbol{s}_0$, and eigen vector, $\boldsymbol{s}_j$, are calculated by a principal component analysis (PCA) of the training data, which are facial image labeled feature points. $\boldsymbol{s}$ is represented by $\boldsymbol{s}_0$, and shape parameter $\boldsymbol{p} = (p_1, \cdots, p_j)$ is a weighted sum of $\boldsymbol{s}_j$ as follows,

$$\boldsymbol{s} = \boldsymbol{s}_0 + \sum_{j=1}^{m} p_j \boldsymbol{s}_j \tag{2}$$

Various shapes are generated by changing shape parameter $\boldsymbol{p}$. Figure 1 (a) shows examples of the shapes.

Facial appearance is represented by the appearance model. $A(\boldsymbol{x})$ is the facial image normalized to the mean shape $\boldsymbol{s}_0$. The mean appearance, $A_0(\boldsymbol{x})$, and eigen vector, $A_j(\boldsymbol{x})$, are calculated using PCA of the normalized training data. $A(\boldsymbol{x})$ is represented by $A_0(\boldsymbol{x})$ and appearance parameter $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_j)$ is a weighted sum of $A_j(\boldsymbol{x})$ as follows,

$$A(\boldsymbol{x}) = A_0(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j A_j(\boldsymbol{x}) \tag{3}$$

The mean appearance, $A_0(\boldsymbol{x})$, is called the template image. Figure 1 (b) shows an example of $A_0(\boldsymbol{x})$.

2.2. Conventional fitting algorithm

The goal of fitting is to estimate the shape parameter, $\boldsymbol{p}$, which fits the input image. ICIA is a typical AAM fitting algorithm

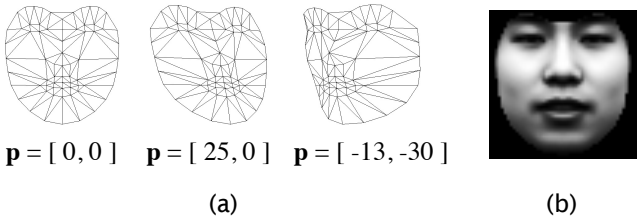

Fig. 1. Examples of AAMs. (a) Shape model with different shape parameters. (b) Template image of the appearance model.

that minimizes the matching error between the template image and input image as follows,

$$\sum_{\boldsymbol{x} \in s_0} [\, I(W(\boldsymbol{x}; \boldsymbol{p})) - A_0(\boldsymbol{x}) \,]^2 \tag{4}$$

Here, $\boldsymbol{x} = (x, y)$ is coordinate point which belongs to template image space, $W(\boldsymbol{x}; \boldsymbol{p})$ is a warp function with shape parameter $\boldsymbol{p}$, and $I(W(\boldsymbol{x}; \boldsymbol{p}))$ is a normalized input image. It is difficult to estimate $\boldsymbol{p}$ directly, and therefore $\Delta \boldsymbol{p}_T$, which is the difference of the shape parameter in template space, is estimated instead. The error function, E_I, is as follows,

$$E_I = \sum_{\boldsymbol{x} \in s_0} [\, I(W(\boldsymbol{x}; \boldsymbol{p})) - A_0(W(\boldsymbol{x}; \Delta \boldsymbol{p}_T)) \,]^2 \tag{5}$$

$A_0(W(\boldsymbol{x}; \Delta \boldsymbol{p}_T))$ is transformed to a linear function of $\Delta \boldsymbol{p}_T$ by Taylor expansion around $\Delta \boldsymbol{p}_T = 0$. Then, $\Delta \boldsymbol{p}_T$, which is minimized E_I , updates $\boldsymbol{p}$ iteratively up to convergence.

3. PROPOSED METHOD

3.1. Problem with ICIA

There are fitting algorithms other than ICIA for estimating shape parameter $\boldsymbol{p}$, including Lucas-Kanade image alignment [5], forwards compositional image alignment [5], and others. ICIA has more elements that can be calculated in advance, so its iteration process is faster than the other algorithms. Since ICIA only minimizes the image matching error, there is no constraint to shape parameter $\boldsymbol{p}$. In other words, equation (4) allows non-face-like shape parameters, provided the equation is minimized. Figure 2 shows an example of a result for an estimated non-face-like shape. The shape parameter is updated so that it can be seen distinctly from the correct face shape distribution. In addition, the image fitting result is shown in Figure 5 (a). Fitting of the next frame will most certainly fail if a non-face-like shape is estimated once when fitting the model to the videos. Therefore, it is thought that imposing a constraint on the shape parameters significantly contributes to fitting stability In this study, we aimed to improve ICIA by adding a constraint that uses face shape likelihood.

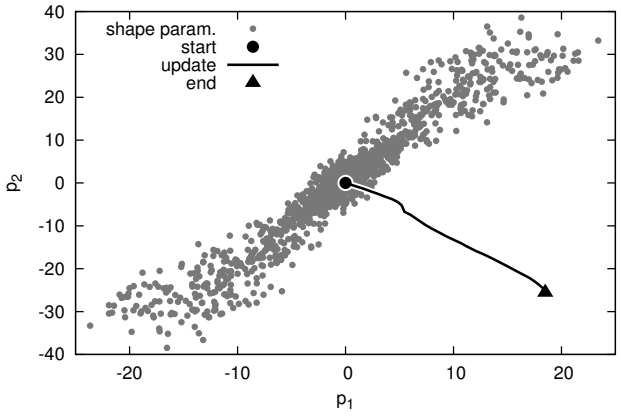

Fig. 2. The shape parameter distribution calculated from the correct shape. p_1 and p_2 indicate the first and second shape parameter dimensions, respectively. The line that starts at the black circle and ends at the triangle represents the ICIA shape parameter update state. The ICIA fitting result is shown in Figure 5 (a).

3.2. Conventional constraint : eigen value constraint

Conventionally, a constraint which reduces the norm of the shape parameter $\boldsymbol{p}$ has been used [6]. This constraint is as follows,

$$E_E = \boldsymbol{p}^T \Sigma^{-1} \boldsymbol{p} \tag{6}$$

Here, Σ is the diagonal matrix whose entries are the eigen values of the shape model. In this paper, it is called the E_E eigen value constraint (EVC). The error function with EVC is as follows,

$$E = E_I + w_E E_E \tag{7}$$

Here, w_E is the EVC weight. EVC has the effect of suppressing shape parameter $\boldsymbol{p}$ from receding from around 0. Figure 3 shows the EVC cost contour and the updated state of $\boldsymbol{p}$ by ICIA with EVC in shape parameter space. In addition, the resulting image is shown in Figure 5 (b). EVC is a very simple constraint. In Figure 3, the update is successful by chance, but there is the potential that $\boldsymbol{p}$ is updated away from the distribution, depending on the matching error of the image, E_I.

3.3. Proposed constraint : face shape likelihood

To more precisely represent the correct shape parameter distribution, the distribution was modeled using the Gaussian mixture model (GMM). Our proposed constraint is the likelihood of the GMM as follows,

$$E_F = -\log \sum_{m=1}^{M} w_m \, \mathcal{N}\,(\, \boldsymbol{p} \mid \mu_m,\ \Sigma_m \,) \tag{8}$$

Here, M is the mixture number of the GMM, $\mathcal{N}(\cdot)$ is the normal distribution probability density function, $\boldsymbol{\mu}_m$ and Σ_m are

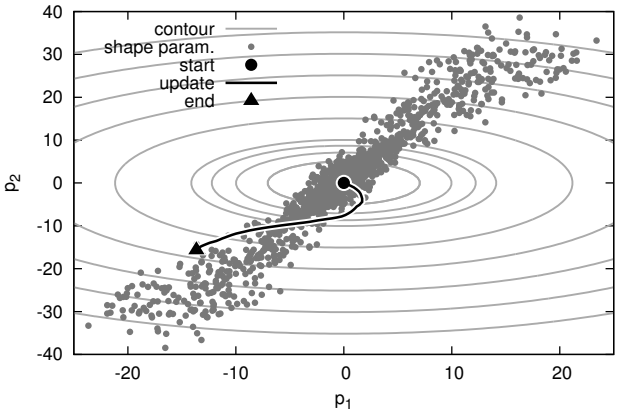

Fig. 3. The EVC contour and the state of the update by ICIA with EVC. The resulting image is shown in Figure 5 (b).

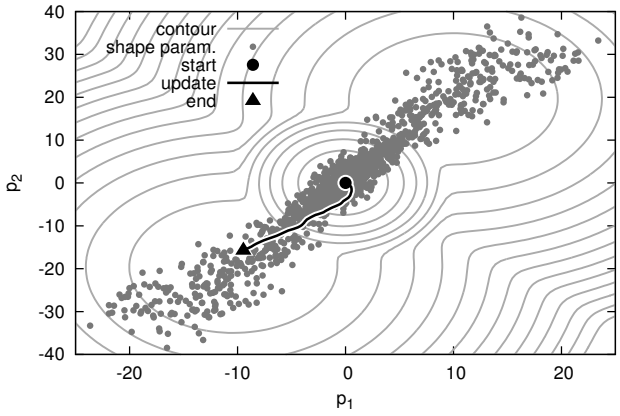

Fig. 4. The FSL contour and the state of the update by ICIA with FSL. The GMM mixture number is 4. The resulting image is shown in Figure 5 (c).

mean and diagonal covariance matrices respectively, and w_m is the normal distribution weight. E_F means the logarithm of the reciprocal likelihood which indicates how face-like the shape represented by shape parameter $\boldsymbol{p}$ is. E_F is called the face shape likelihood. The error function to be minimized with FSL is as follows,

$$E = E_I + w_F E_F \tag{9}$$

Here, w_F is the FSL weight. The FSL cost contour and the updated state of $\boldsymbol{p}$ by ICIA with FSL in shape parameter space are shown in Figure 4. In addition, the resulting image is shown in Figure 5 (c). The FSL contour better represents the correct shape parameter distribution than EVC. Therefore, it can be expected that the result is less likely to deviate from the distribution.

4. EXPERIMENT

An experiment was conducted to evaluate fitting stability using ICIA with the proposed FSL constraint. Fitting was performed to test videos and lost frames (frames failing fitting) were counted.

4.1. Setting

Two types of videos were prepared. The first was a controlled set, where subjects were instructed to move their faces (CTRL set) The second set was a conversation scene (CONV set). Both of sets included 14 persons' videos (13 male, 1 female). Each video was shot at 30 fps with a length of around 900 frames (30 s).

An augmented AAM, called a multi-band AAM, that is robust to illumination variation was used in this experiment. The multi-band AAM has a gray scale facial image appearance model and two x-axis and y-axis gradient image direction models. Hence, ICIA error functions include gradient

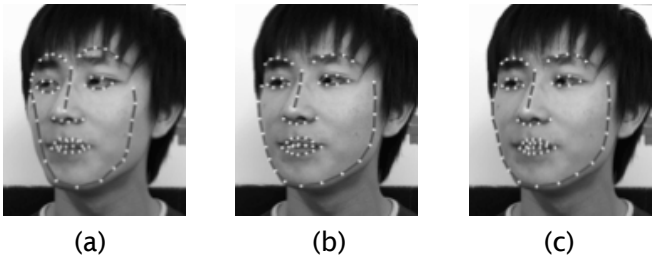

Fig. 5. Fitting result examples: (a) ICIA (b) ICIA+EVC (c) ICIA+FSL

image E_X and E_Y matching errors. In addition, robust shape initialization (RSI) was used to perform more stable fittings. RSI estimates the initial shape using optical flow. The multi-band AAM and RSI are detailed in paper [7].

The following three fitting methods were compared: (a) ICIA (basic), (b) ICIA+EVC (conventional), (c) ICIA+FSL (proposed). Table 1 shows the error function of each method. The experiment was conducted as the shape parameter dimensions changed from 7 to 30. The weights for each constraint and the GMM mixture number for FSL were decided by a preliminary experiment with respect to each shape parameter's dimension. Each method began with face detection [8], and the initial shape parameter was then calculated from the face region. If a lost frame was detected, the face detection process was restarted.

4.2. Lost frame detection

If even one of the following conditions is met, a lost frame is detected.

- No face is detected in the frame
- The normalized norm of the shape parameter, $\boldsymbol{p}$, is

Table 1. Error function E of each method.

Method	Error function E
ICIA	$E_I + w_X E_X + w_Y E_Y$
ICIA+EVC	$E_I + w_X E_X + w_Y E_Y + w_E E_E$
ICIA+FSL	$E_I + w_X E_X + w_Y E_Y + w_F E_F$

larger than threshold th_n

- The appearance reconstruction error (ARE) is larger than threshold th_a

The normalized norm of $\boldsymbol{p}$ has the same value as EVC. When the normalized norm of $\boldsymbol{p}$ is large, the shape is considered to be non-face-like, and hence, a lost frame is deemed to have occurred. th_n is set to 40. ARE is the weighted mean squared error as follows:

$$ARE = \sqrt{\frac{R(\boldsymbol{p})}{\sum_{\boldsymbol{x} \in s_0} 1}} \tag{10}$$

$$R(\boldsymbol{p}) = \sum_{\boldsymbol{x} \in s_0} \left[\left\{ A_0(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j A_j(\boldsymbol{x}) - I(W(\boldsymbol{x}; \boldsymbol{p})) \right\} M(\boldsymbol{x}) \right]^2 \tag{11}$$

Here, $M(\boldsymbol{x})$ is the weighted mask image, $I(W(\boldsymbol{x}; \boldsymbol{p}))$ is the fitting result image, and $A_0(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j A_j(\boldsymbol{x})$ is the image reconstructed using the appearance model. Examples of these images are shown in Figure 6. $M(\boldsymbol{x})$ was used to emphasize the error on the contour. If ARE is large, it is thought that the appearance model cannot explain the fitting result well, and hence, a lost frame is deemed to have occurred. The lost frame is easily detected as the th_a is smaller. In this experiment, th_a was changed by two increments from 26 to 30.

4.3. Results

The results are shown in Figures 7-10. The CTRL set result when th_a was fixed as 28 and the dimension of the shape parameter was changed is shown in Figure 7. The CTRL set result when the dimension of the shape parameter was fixed as 20 and th_a was changed is shown in Figure 8. Both ICIA+EVC and ICIA+FSL had better results than basic ICIA, regardless of the dimension of the shape parameter or th_a. Comparing ICIA+EVC and ICIA+FSL, there was no significant difference in the result. The CONV set result when th_a was fixed as 28 and the dimension of the shape parameter was changed is shown in Figure 9. The CONV set result when the dimension of the shape parameter was fixed as 20 and th_a was changed is shown in Figure 10. Similar to the CTRL set result, ICIA+EVC and ICIA+FSL had better results than basic ICIA. Comparing ICIA+EVC and ICIA+FSL,

Fig. 6. Examples of images for calculating ARE: (a) fitting result (b) normalized result image (c) reconstructed image (d) weighted mask image. Black pixels are 0.0, gray pixels are 0.5, and white pixels are 1.0.

ICIA+FSL performed better than ICIA+EVC when the dimension of the shape parameter was greater than 15 (Figure 9), and ICIA+FSL was better regardless of th_a (Figure 10). The CONV set videos had larger facial movements (especially mouth movements) than those of the CTRL set. Therefore, the proposed ICIA+FSL method is thought to be robust against large facial movements.

5. CONCLUSION

A novel FSL constraint was proposed for the typical AAM fitting method, ICIA. Experiments showed that ICIA with FSL has better fitting stability than ICIA with conventional EVC constraints, especially for large facial movements.

In the future, we plan to evaluate the fitting accuracy of the proposed method. FSL represents the correct shape parameter distribution better than EVC, and hence, it is thought that ICIA with FSL performs considerably better than ICIA with EVC.

6. REFERENCES

[1] T. Kozakaya, M. Yuasa T. Shibata, and O. Yamaguchi, "Facial feature localization using weighted vector concentration approach," *Image and Vision Computing*, vol. 28(5), pp. 772–780, 2010.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.

[3] S. Cosar and M. Cetin, "A graphical model based solution to the facial feature point tracking problem," *Image and Vision Computing*, vol. 29(5), pp. 335–350, 2011.

[4] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Computer Graphics, Annual Conference Series (SIGGRAPH)*, 1999, pp. 187–194.

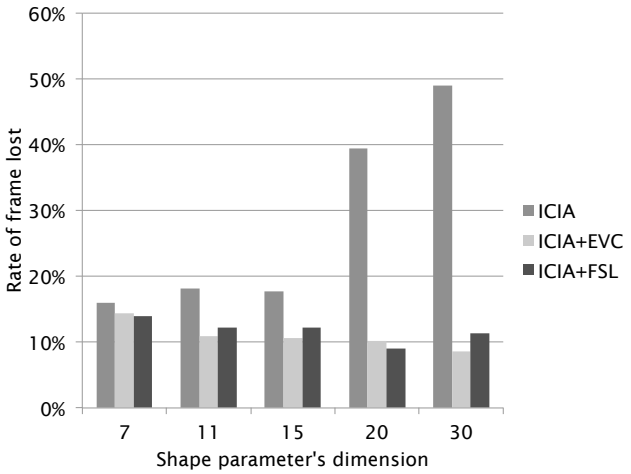

Fig. 7. The CTRL set result when th_a was fixed as 28 and the shape parameter dimension was changed.

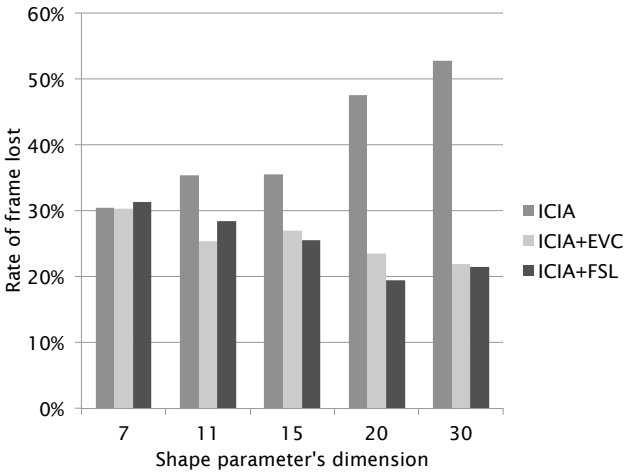

Fig. 9. The CONV set result when th_a was fixed as 28 the shape parameter dimension was changed.

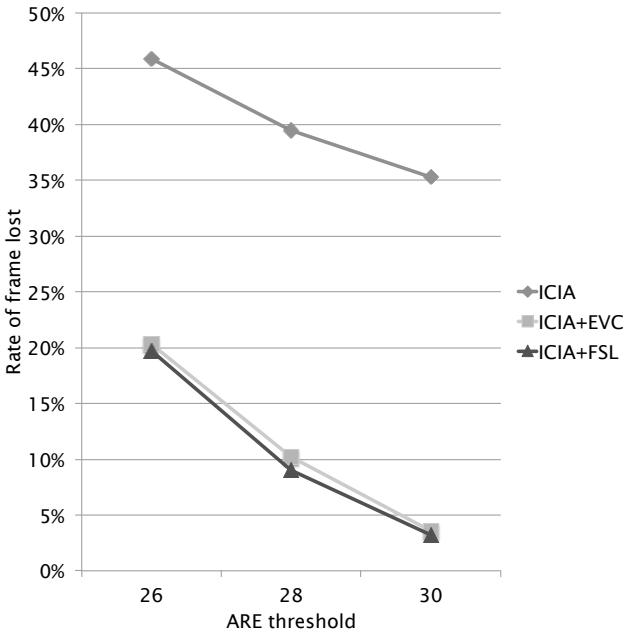

Fig. 8. The CTRL set result when the shape parameter dimension was fixed as 20 and th_a was changed.

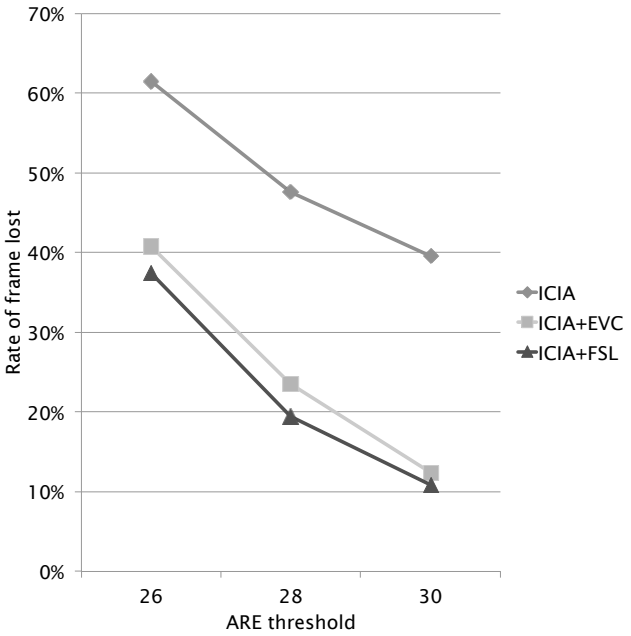

Fig. 10. The CONV set result when the shape parameter dimension was fixed as 20 and th_a was changed.

[5] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60(2), pp. 135–164, 2004.

[6] S. Baker, R. Gross, and I. Matthews, "Lucas-kanade 20 years on: A unifying framework: Part4," *Technical Report CMU-RI-TR-04-14*, 2004.

[7] M. Zhou, L. Liang, J. Sun, and Y. Wang, "AAM based face tracking with temporal matching and face segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 701–708.

[8] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57(2), pp. 137–154, 2004.