Journal of Pattern Recognition Research 7 (2012) 72-79

Received Oct 24, 2011. Revised Jan 16, 2012. Accepted Mar 2, 2012.

Facial Feature Points Tracking Based on AAM with Optical Flow Constrained Initialization

Ying Cui *cuiying@patternrecognition.cn*

Zhong Jin *jinzhong@patternrecognition.cn*

School of Computer Science and Technology
Nanjing University of Science and Technology, Nanjing, China

Abstract

A facial feature points tracking method is proposed by adding Lucas-Kanade optical flow constraint on the face alignment algorithm, Active Appearance Model (AAM). The optical flow considers the inter-frame correspondence and uses it to estimate AAM initial shape more accurately with similarity preservation. Experiments show that the proposed method can successfully track the frames that general AAM tracks failed. The method achieves not only the accuracy but also the computation time improvement.

Keywords: Feature Point Tracking, Active Appearance Model, Lucas-Kanade optical flow.

1. Introduction

Through years of research, face recognition technology has made great strides and development. With the demand for development of many applications, such as video surveillance, video conferencing, information security and access control, the study based on videos has been one of the most active points in face research areas [1]. At the same time, accurate and quick face alignment is the foundation of face recognition, facial expression recognition and 3D face modeling [2]. Both of them make face tracking and face alignment in video sequences be an important problem that must be solved [3, 4].

Active appearance model (AAM) [5] is a fast and efficient method for face alignment in still images. It can be used directly in video tracking by searching on each subsequent frame with the initialization of the final result of the previous frame [6, 7]. However, this is only suitable for small movement between frames because AAM is sensitive to the initial shape and may easily be stuck in local minima due to its gradient decent optimization. To solve the problem, in this paper, the Lucas-Kanade optical flow [8] is used to improve AAM's efficiency in video face tracking by considering the inter-frame correspondences and a robust shape initialization method is proposed.

2. Active Appearance Model

AAM consists of two parts: AAM face modeling and AAM fitting. Face modeling constructs the shape model and the appearance model. AAM fitting is the process of finding the model parameters which best fit a given image.

The shape S of an AAM can be described by n feature points in the image:

$$S = (x_1, y_1, x_2, y_2, \cdots, x_n, y_n)^T$$

AAM allows linear shape variation. This means that the shape S can be expressed as a base shape S_0 plus a linear combination of k shape vectors $\{S_i\}$:

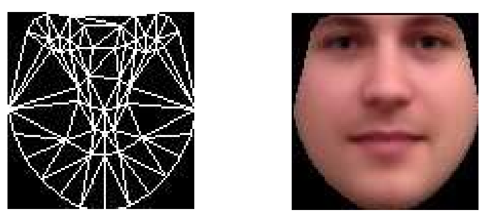
$$S(p) = S_0 + \sum_{i=1}^{k} p_i S_i \tag{1}$$

where $(p_1, p_2, \cdots, p_k)^T$ are the shape parameters. The base shape S_0 and the k shape vectors $\{S_i\}$ are computed by applying Principal Component Analysis (PCA) to the training shapes (from a hand labeled training set). The base shape S_0 is the mean shape and the shape vectors $\{S_i\}$ are the k eigenvectors corresponding to the k largest eigenvalues. Fig.1(a) shows the mesh of the base shape.

The appearance of an AAM is defined within the mesh of the base shape S_0. Let M_0 denote the set of pixels $\mathrm{x}=(x,y)^T$ that lie inside the mesh of S_0. The appearance of an AAM is then an image $A(\mathrm{x})$ defined over the pixels $\mathrm{x} \in M_0$. Similar to the shape S, the appearance $A(\mathrm{x})$ can be expressed as a base appearance $A_0(\mathrm{x})$ plus a linear combination of h appearance images $\{A_i(\mathrm{x})\}$:

$$A(\mathrm{x}) = A_0(\mathrm{x}) + \sum_{i=1}^{h} \lambda_i A_i(\mathrm{x}) \quad \forall \mathrm{x} \in M_0 \tag{2}$$

where the coefficients $\{\lambda_i\}$ are the appearance parameters. The base appearance $A_0(\mathrm{x})$ and appearance images $\{A_i(\mathrm{x})\}$ are computed by applying PCA to the shape-normalized training images. Each training image is shape-normalized by warping the hand labeled training shape onto the mesh of S_0. The base appearance A_0 is the mean image and the appearance images $\{A_i\}$ are the h eigenimages corresponding to the h largest eigenvalues. Fig.1(b) shows the base appearance according to the base shape.



(a) (b)

Fig. 1: The mesh of the base shape (a) and the base appearance (b)

To align the face shape on a given image $I(\mathrm{x})$, AAM aims to find the optimal shape parameters p and the optimal appearance parameters λ to minimize the difference between the image $I(\mathrm{x})$ and the model appearance $A(\mathrm{x})$. If x is a pixel in M_0, then the corresponding pixel in the given image I is W(x; p). At pixel x the AAM has the appearance $A(\mathrm{x})$ denoting by (2). At pixel W(x; p) the given image I has the intensity $I(\mathrm{W(x;\ p)})$. Then we can define the difference as

$$E(\mathrm{p}, \lambda) = \sum_{\mathrm{x} \in M_0} \left[A_0(\mathrm{x}) + \sum_{i=1}^{h} \lambda_i A_i(\mathrm{x}) - I(W(\mathrm{x;p})) \right]^2 \tag{3}$$

The function $W(\mathrm{x;\ p})$ is a piecewise affine warp defined by the pair of shapes S_0 and $S(\mathrm{p})$ [9].

The cost function (3) can be minimized by the inverse compositional algorithm, which is well known as its efficiency [9].

3. Feature Points Tracking

Optical flow is a kind of real-time object tracking algorithm. It can be integrated with other algorithms and used for face tracking and face recognition [10, 11]. In this paper, we integrated the Lucas-Kanade optical flow with AAM for facial feature points tracking.

3.1 Lucas-Kanade Optical Flow

Given two neighboring frames I_{t-1}, I_t, for a point $p = (x, y)^{\mathrm{T}}$ in I_{t-1}, if the optical flow is $d = (u, v)^{\mathrm{T}}$, then the corresponding point in I_t is $p + d$. Lucas-Kanade optical flow aims to find the offset d to minimize the match error between the local appearances of two corresponding points. That is, we can define a cost function upon $N(p)$, the local area of p:

$$e(d) = \sum_{\mathrm{x} \in N(p)} w(\mathrm{x})\big(I_t(\mathrm{x} + d) - I_{t-1}(\mathrm{x})\big)^2 \tag{4}$$

where $w(\mathrm{x})$ is the weight function. Optimize (4) and we can get the solution:

$$d = G^{-1}H \tag{5}$$

where $G = \sum\limits_{\mathrm{x} \in N(p)} w(\mathrm{x})\nabla I_t(\nabla I_t)^T$, $H = \sum\limits_{\mathrm{x} \in N(p)} w(\mathrm{x})\nabla I_t \Delta I)$, $\Delta I = I_{t-1} - I_t$, $\nabla I_t = \dfrac{dI_t}{d\mathrm{x}}$.

Lucas-Kanade optical flow demands feature points with salient local appearances, or it can not track accurately [12]. In this paper, Lucas-Kanade optical flow is used to track some key points with salient local appearances on the face. The key points are then used for constraining AAM shape initialization.

3.2 Inter-frame Correspondence

Human face can be broadly considered as a two-dimensional graphic which meets the Similarity Transformation constraint.

Let the current optical flow tracking set in frame t be $F^t = \begin{pmatrix} x^t_{f1}, x^t_{f2}, \cdots, x^t_{fl} \\ y^t_{f1}, y^t_{f2}, \cdots, y^t_{fl} \end{pmatrix}$, where $l \geq 3$, the optical flow tracking set in frame $t-1$ be F^{t-1}, then the Similarity Transformation between the two frames is:

$$F^t = CF^{t-1} + D \tag{6}$$

where $C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ is the scaling and rotation matrix, $D = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ is the translation. The detail of computing C and D can see Appendix B of [13].

3.3 Similarity Transformation Preservation

According to the similarity preservation of the same two frames, given the shape of frame $t-1$, we can pre-estimate the shape of frame t by the Similarity Transformation parameters C and D.

Let the current feature points set in frame t be $S^t = \begin{pmatrix} x_1^t, x_2^t, \cdots, x_n^t \\ y_1^t, y_2^t, \cdots, y_n^t \end{pmatrix}$, and the feature points set in frame $t-1$ be S^{t-1}, then with the similarity preservation, S^t can be estimated by parameters C and D:

$$\hat{S}^t = CS^{t-1} + D \tag{7}$$

3.4 AAM Tracking with Optical Flow Constraint

The tracking algorithm:

1. Find the key points with salient local appearances for Lucas-Kanade optical flow in the last frame, frame $t-1(t=2)$, and get F^{t-1}.
2. For frame t, compute the Lucas-Kanade optical flow tracking set F^t.
3. Compute the Similarity Transformation parameters C and D.
4. According to the AAM tracking result S^{t-1} of frame $t-1$, estimate the current corresponding set $\hat{S}^t$ by parameters C and D.
5. Let $\hat{S}^t$ be the initial shape of AAM, use AAM to track the feature points S^t.
6. Let $t = t+1$, go to 2).

4. Experiments

Firstly, we do experiments on still images to study the impact of initial shape on AAM fitting. Secondly, we do experiments on videos to verify the effectiveness of the proposed tracking method.

4.1 Impact of The Initial Shape

The AAM model is constructed on IMM face database. The IMM database consists of 240 images of 40 people, exhibiting variations in pose, expression and lighting. Two sets of images are used for testing, each with 40 images of 40 people. One is full frontal face, and the other is profile face rotated 30 degrees to the person's right.

The AAM search is initialized from a range of displacements (from -30 to +30 pixels) both in x and y, and for each displacement the point-to-point errors are evaluated by comparing the result with hand labeled ones. Some results are shown in Fig.2, Fig.3 and Fig.4. The large displacement leads to bad results, especially in profile faces. If the displacement is smaller than 10 pixels, the fitting result is good. If the displacement is larger than 10 pixels, the fitting result may lead to local optimization or even be a mess.

4.2 Face Feature Points Tracking

We hand labeled ten images with different poses and expressions for one person to perform the experiment. Each image is labeled with 58 feature points. The proposed method is tested on recorded videos. And the similar experiment is taken with general AAM tracking, by initializing with the final result of the previous frame. In the proposed method, 8 key points with salient local appearances are selected for Lucas-Kanade optical flow tracking, an example is shown in Fig.5. Then the optical flow tracking result is used for AAM shape initialization. Two video frame sizes are tested: size 480×640 and size 240×320.

As the test video segments are not hand labeled, we just subjectively judge the tracking is whether successful or not. The judgments are summarized in Table 1; the proposed

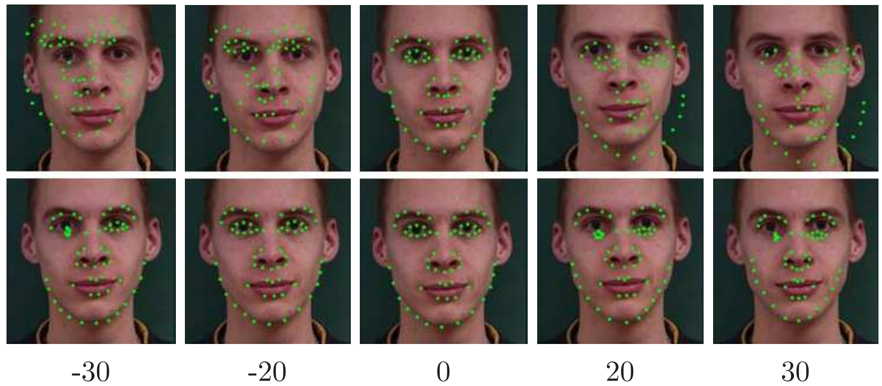


Fig. 2: Frontal face, the initial shape (top), the result shape (bottom)

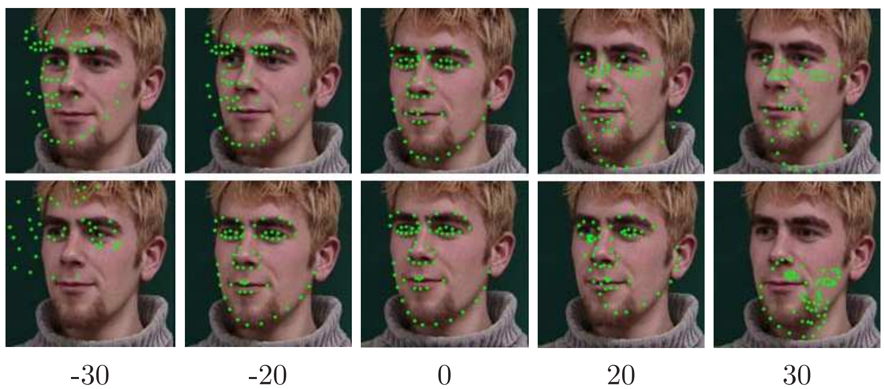


Fig. 3: Profile face, the initial shape (top), the result shape (bottom)

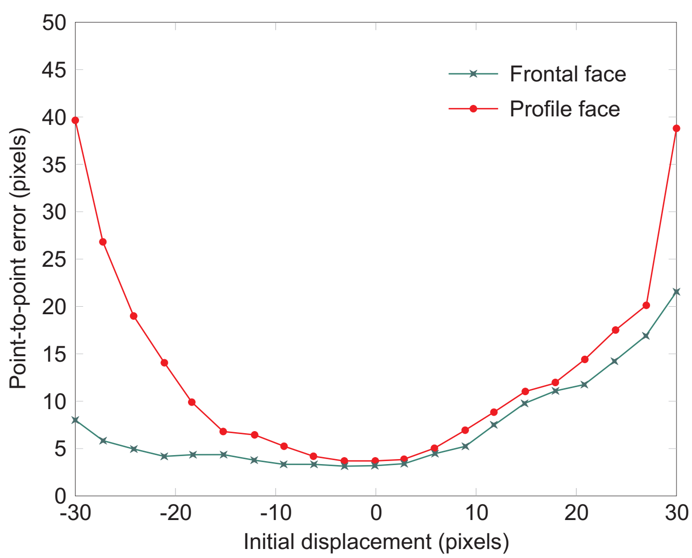


Fig. 4: Point-to-Point errors after search from displaced positions

method gives much more accurate, stable and robust results. Fig.6 shows some of the tracking results, the proposed method can successfully track the frames that general AAM tracks failed. The number of iterations of each frame is recorded during the process of tracking. Here we define the convergence rate as the average number of iterations of the successfully tracked frames. And the comparison of convergence rate is shown in Fig.7; the proposed method has a higher convergence efficiency compared with general AAM. Although optical flow computation takes some time, the proposed method is faster than general AAM tracking, owing to its better shape initialization. The average tracking time is shown in the last column of Table 1.

Fig. 5: Computed optical flow between frame 83 and 84

Table 1: Subjective Judgment of Tracking Results

Frame size	Method	Number of frames	Number of failed frames	Rate of failed frames	Mean tracking time per frame
480×640	Proposed method	400	6	1.5%	234.47 ms
	AAM	400	24	6%	304.11 ms
240×320	Proposed method	400	6	1.5%	43.23 ms
	AAM	400	22	5.5%	46.59 ms

Frame 80 Frame 103 Frame 293

Fig. 6: Tracking result comparison: general AAM (top), optical constraint AAM (bottom)

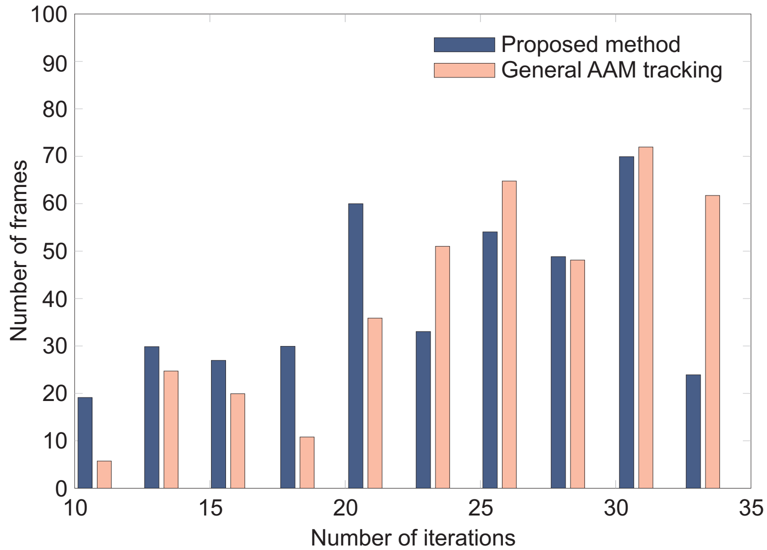


Fig. 7: Comparison of convergence rate

5. Conclusions and Future Works

By adding Lucas-Kanade optical flow object tracking constraint, we have gotten a more robust AAM-based facial feature points tracking method. The inadequacy is that the transformation constraint is only with scale, translation and rotation in plane. In the future we will think about the depth rotation constraint, and make the tracker more robustly track profile views with large angles. The good tracked feature points can be used for virtual mouse controlling, face recognition and 3D face modeling.

Acknowledgments

This work is partially supported by the Natural Science Foundation of China under grant Nos. 60973098, 61005005

References

[1] Yan Yan and Zhang Yu-Jin. "State-of-the-Art on Video-Base Face Recognition". *Chinese Journal of Computers*, 2009, Vol.32, No.5, pp. 878-886.

[2] Song Gang, Ai Haizhou and Xu Guangyou. "Texture constrained facial feature point tracking". *Journal of Software*, 2004, 15(11), pp. 1607-1615.

[3] Minyoung Kim, Kumar, S., Pavlovic, V.,and Rowley, H.. "Face tracking and recognition with visual constraints in real-world videos". *Computer Vision and Pattern Recognition (CVPR)*, 2008 IEEE Conference on, pp. 1-8.

[4] Mingcai Zhou, Lin Liang, Jian Sun, and Yangsheng Wang. "AAM based face tracking with temporal matching and face segmentation". *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on, pp. 701-708.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6), pp. 681-685, 2001.

[6] Ahlberg J. "Real-Time facial feature tracking using an active model with fast image warping". *In: Kollias S, Maragos P, eds. Proc. Of the Intl Workshop on Very Low Bitrate Video (VLBV)*. Athens, 2001, pp. 39-43. (http://www.icg.isy.liu.se/ ahlberg/papers/conf/ VLBV-01.pdf)

[7] T.F.Cootes and C.J. Taylor. "An Algorithm for Tuning an Active Appearance Model to New Data". *British Machine Vision Conference (BMVC)*, pp. 919-928, 2006.

[8] Andres Bruhn, Joachim Weichert and Christoph Schnorr. "Lucas/Kanade meets Horn/Schunck: Combining Local and Global Optic Flow Methods". *International Journal of Computer Vision*, 61(3), pp. 211-231, 2005.

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

[10] Douglas Decarlo and Dimitris Metaxas. "Optical Flow Constraints on Deformable Models with Applications to Face Tracking". *International Journal of Computer Vision*, vol.38, no.2, pp.99-127, July 2000.

[11] Hsieh CK, Lai SH and Chen YC. "An optical flow-based approach to robust face recognition under expression variations". *IEEE Transactions on Image Processing*, 19(1), pp. 233-240, Jan 2010.

[12] Krishnan Ramnath, Simon Baker, Iain Matthews and Deva Ramanan. "Increasing the density of Active Appearance Models". *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[13] T.F. Cootes and C.J. Taylor. "Statistical Models of Appearance for Computer Vision". *Technical Report, University of Manchester*, 125 pages, 2004.
(www.face-rec.org/algorithms/AAM/app_models.pdf)