A Robust Method of Facial Feature Tracking for Moving Images

Yuka Nomura*
Graduate School of Interdisciplinary Information Studies, The University of Tokyo

Takayuki Itoh †
Graduate School of Humanitics and Sciences, Ochanomizu University

Yasushi Yamaguchi ‡
The Interfaculty Initiative in Information Studies, The University of Tokyo

Abstract

Robustness of feature tracking is really important to recognize facial expressions and movements of human in the natural state. This paper presents a technique for a robust facial feature tracking for the human facial videos. Our technique first detects particular feature points by template matching, and then tracks the points in moving images by optical flow estimation. At the same time, the technique determines the reliabilities of the detected feature points. If the technique determines a feature point is not reliable, it executes template matching in the neighbor domain or estimates the position of the feature point according to the level of the reliability. By those steps, when the difficulty to detect the feature point is dissolved, we can detect and track feature points again appropriately and quickly. The technique can be applied to understanding of user's delicate feelings, conditions, or psychology.

Keywords: Facial feature tracking, template matching, optical flow estimation, video processing

1. Introduction

Various kinds of techniques for recognition of facial expressions have been recently presented. Those techniques realize better human and computer interaction. In most of the techniques, computers recognize facial changes by tracking face areas, facial parts and features. Those techniques will realize that computers understand human’s facial expression, psychology, gestures, and so on. Some researches have attempted to recognize facial changes by detection of particular facial feature points and tracking them.

However, those techniques have problems that elaboration and robustness for human’s motions are not enough. When we consider the human in the natural state, we also have to consider sudden occlusions and various human’s motions. Robustness of feature tracking is really important to recognize natural facial expressions and movements. Even though various robust techniques for feature tracking are proposed [1,2,3], we need to discuss for development of more efficient techniques.

This paper presents a technique for a robust facial feature tracking for the human facial videos, which realizes continuous tracking even while the features are occluded. Our technique applies template matching to the first frame of facial video for initialization of feature detection. While it determines the feature points are reliable, it tracks them by applying optical flow estimation since it is a fast feature tracking method. In addition, the technique supports an exception handling mechanism which is applied when the technique failed to track the points. The exception handling mechanism first executes template matching to a partial area. If the result of template matching is proper, it determines that it is a proper feature point, and returns to tracking by the optical flow estimation. Otherwise, it tracks other reliable points estimated by position measurement, so that it can quickly return to the step of partial template matching when it comes to discover the point again. By applying those steps, our technique realizes robust and fast method to detect and track facial feature points.

Our technique is somewhat similar to a recent technique [4] which has already realized robust and continuous tracking even while the features are under occlusions. We think our technique is different from the technique presented in [4] as follows:

- Our technique applies optical flow estimation while it can properly track the features, because it is faster than template matching.
- Our technique tracks estimated features while they are really occluded, and it makes quickly return to template matching when it can be detected again.

*e-mail: yuka@graco.c.u-tokyo.ac.jp
†e-mail: itot@computer.org
‡e-mail: yama@graco.c.u-tokyo.ac.jp

The reminder of the paper is organized as follows: Section 2 introduces related works in facial feature detection and tracking. Section 3 describes overview and detailed implementation of the presented technique. Sections 4 present the experiments of the technique. Section 5 concludes and suggests future works.

2. Related Works

Researches on faces and facial features detection and tracking have been hot topics. There are techniques focus on structures and pixels of features. One of the techniques focuses on features in line-structures of hair and eyebrow and to focus on features in the appearance of pixels corresponding to eye, nose and mouth [5]. Face color and dark color regions are also useful information for facial feature detection [6]. Head pose, face deformation, and combination of various facial movements are used as keys for recognizing the positions of facial feature points and analyzing various human's conditions [7,8]. Some other researches use infrared camera [9] to detect and track eyes, which are also useful for human's gaze estimation.

Some of several feature tracking techniques are somewhat common to our technique, since they use pattern matching for detection of features [1,2,3,4]. Especially, some techniques [2,3] apply combination method of pattern matching and shape extraction to realize robust detection and tracking. One other technique focuses on an optical-flow based approach for facial feature tracking that is sensitive to subtle changes in facial expression [11].

Robustness of facial feature tracking is an important problem. Some techniques have already realized the continuous detection of face under partial occlusions [4,10]. As discussed in Section 1, the technique described in [4] is somewhat similar to our technique.

There are various applications of facial feature tracking. For example, understanding of user's delicate feelings and psychology [10] is an interesting application.

3. Implementation

3.1 Overview

The technique presented in this paper tracks feature points in moving images taken by a digital video camera. The technique applies template matching for facial feature detection, and optical flow for the feature tracking, while it switches them according to the reliability of the detected points. The algorithm consists of the following steps.

1. Prepare the images of the subject's particular facial features as template images.
2. Detect particular features by template matching, in the first frame of the moving images.
3. Track the detected feature points by Lucas-Kanade methods, in the following frames.

If the detected points are reliable to track, the technique will continue to track them. However, if the detected points are not reliable, it executes the following steps.

1. Execute the template matching again in the near-field regions of points at the previous frame.
2. Estimate the appropriate position, if the similarity between the template and the area extracted by the template matching is low.

These two steps correspond to an exception handling on reliabilities of detected feature points. This is why our technique can realize a more robust facial feature tracking.

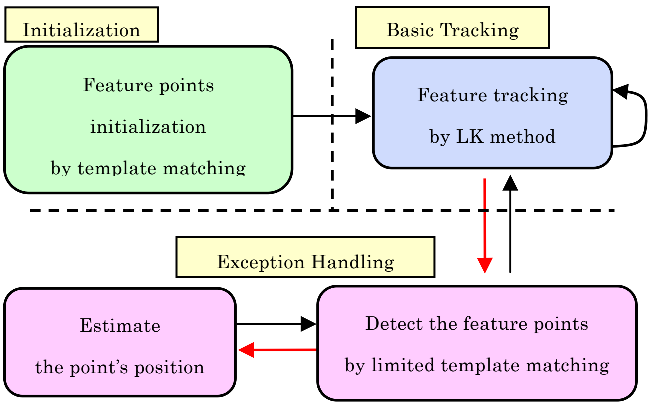

Figure 1. Overview flowchart

3.2 Template matching

The main loop of the presented technique tracks important facial features to track in the moving images. First, we prepare template images to detect features by template matching. Supposing that the template image is the size of M×N pixel, the technique overlaps it onto a partial M×N pixel area of the moving images. Here, let $I(i,j)$ as a value of pixel of i-th from the left and j-th from the top in the original image, and $T(i,j)$ as a value of pixel of i-th from the left and j-th from the top in the template image. This step calculates difference R applying the following equation (1) for the template matching.

$$R = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} | I(i,j) - T(i,j) | \quad \ldots(1)$$

The partial area, where the values minimize R, is considered to be the similar part, and we take it as the detection result of the facial features in the technique.

Currently, we manually extract template images from each subject's face image, and manually specify the position of the feature point on each template image.

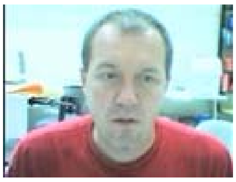

Figure 2. The first frame of a subject's facial video

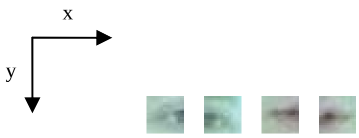

Figure 3. The template images

(right eye side, left eye side, right lip corner, left lip corner)

Table 1. Values of feature points and colors of template matching results' rectangles

	point values (x, y)	Color of rectangles
right eye side	(6, 10)	Blue
left eye side	(16, 12)	Red
right lip corner	(6, 5)	Cyan
left lip corner	(15, 8)	magenta

Figure 4. The result of template matching(rectangles) and feature points (green points)

We show an example of detection of the facial features by the template matching in the facial video provided in the database in [12]. Original image size is 320x240pixels, each template image has 18x18pixels. Figure 2 shows a subject in the first frame, and Figure 3 shows his template images. As shown in Table 1, we decided beforehand each feature point in each template images and each color of rectangles as the template matching results. As a result, it detects corners of the eyes and mouth from the first frame in the facial video of a subject as shown in Figure 4.

After executing template matching and detecting the similar partial area for each template image, the technique will track the point which has similar value as the feature point in the template image, as described in Section 3.2.

3.2 Optical flow estimation

The technique then tracks the feature points in the following frames. We applied Lucas-Kanade method tracks them as an optical flow estimation method. Lucas-Kanade method is one of the spatial partial optimization methods assumed that an optical flow becomes fixed in the local domain of the same object in the dynamic scene [13]. Let the coordinates of a position on an image as *(x, y)*, and the gray value of the image in time *t* as *I(x, y)*. The optical flow *(u, v)* in the local domain *w* is defined as the following equation (2). Figure 5 shows an example of the results of Lucas-Kanade method using a subject' s facial video.

$$u = \frac{\sum_w \frac{\partial I}{\partial x} \cdot [J(p) - I(p)]}{\sum_w \left(\frac{\partial I}{\partial x}\right)^2}, v = \frac{\sum_w \frac{\partial I}{\partial y} \cdot [J(p) - I(p)]}{\sum_w \left(\frac{\partial I}{\partial y}\right)^2} \quad \text{...(2)}$$

$$I(p) = I(x, y, t), J(p) = I(x, y, t + \partial t)$$

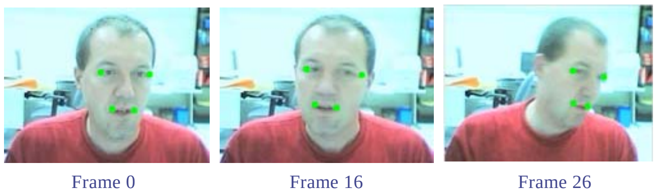

Figure 5. Results of feature tracking by Lucas-Kanade method

3.3 Exception handling

The presented technique estimates reliability of facial feature points while tracking features. It determines that the feature tracking result is unusual, when one of the following situations occurs:

・When the feature point is not detected.

・When the difference of the circumference pixel value of the original feature point and the point after movement exceeds the threshold.

In these cases, the proposal technique executes template matching limited to the small square domain enclosing the position of the feature point in the previous frame. Current our implementation defines the small domain about three times larger than the template images. Consequently, when the minimum value of R in equation (1) exceeds the threshold, it considers that the error of the detection result is large, and the presented technique therefore searches for the similar portion in a larger domain. If the minimum value of R becomes smaller than the threshold, it considers that this detection result is reliable, and the technique supposes it as a result of the proper detection. Figure 6 shows an example flowchart of the process. If the technique unusually tracks a feature point by Lucas-Kanade method, it defines the small domain (dotted frame) centering on the former feature point (blue point), as shown in Figure 6 (Right). Then, the technique executes template matching in the limited domain, and re-detects the feature points.

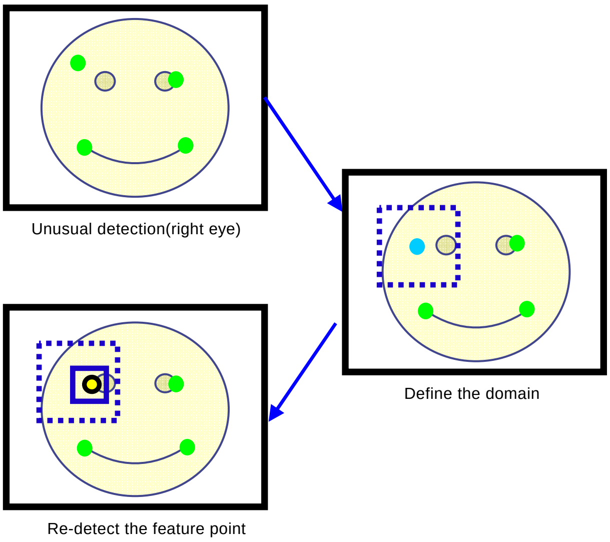

Figure 6. Template matching for the limited region

Figure 7. (Left) Original image. (Upper right) The result of template matching for whole region. (Lower right) The result of template matching for limited region.

The template matching for limited area has two advantages. One is for more proper detection, and the other is for reduction of computation time. As an example, we experiment on a still face image (266x315 pixels) using the same template images as shown in Figure 3. Figure 7 shows the original image and the results of the template matching for whole and limited region. It took 0.2180 seconds for template matching for whole region, and 0.0040 seconds for limited region. From the result, our technique is a more efficient way to detect proper features especially for real-time facial video.

When the process cannot detect the proper points, we suppose that the feature is difficult to detect because it may be occluded by the front object, or the face looks away. In these cases, the technique calculates the estimated position of the feature point based on movements of other feature points.

The technique calculates errors between $\mathbf{p}_{ni}$, and $\mathbf{A}\mathbf{p}_{oi}$, where $\mathbf{p}_{ni}$ is i-th detected feature points in the n-th frame, A is a matrix for projection transformation, and $\mathbf{p}_{oi}$ is i-th detected feature point in the first frame, by the equation (3).

$$E = \sum_{i=0}^{N-1} w_i \,|\, \mathbf{A}\mathbf{p}_{0i} - \mathbf{p}_{ni} \,|^2 \quad ...(3)$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad \mathbf{p}_{ni} = (x_{ni}, y_{ni}, 1)$$

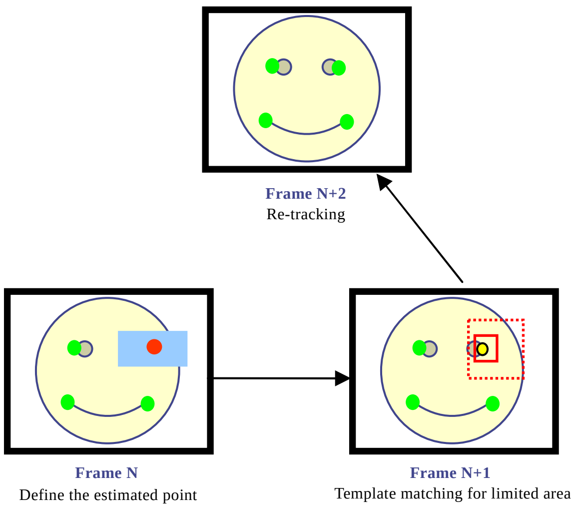

Figure 8. Exception handling in the case of occlusion

Here N is the number of feature points, and w_i is the weight proportional to the reliability of detection results, and E is the error value. Our technique calculates six values of **A** which minimizes E, and then calculates the estimated position of the lost feature points. Figure 8 shows one of the examples of the case that the features are difficult to detect, and how the exception handling works in the situation.

By those processing, when occlusions or some difficulties for feature detection are dissolved, we can detect feature points again appropriately and quickly by applying template matching at the near domains of the estimated facial feature points.

4. Experiments

Figure 9 shows the differences between conventional method and our method using the subject's facial video (320x240 pixels). We used template images as shown in Figure 3 (18x18 pixels). And in this case, we defined 60x60 pixels as a square domain size for the limited template matching. Figure 9 (Left) shows the result of tracking by Lucas-Kanade method, and Figure 9 (Right) shows the result of our technique. For clearness these figures show only the feature points unusually tracked by Lucas-Kanade method. In the left figures, green points show that features, the right corner of eye in frame 15 and the corners of mouth in frame, are detected incorrectly. On the other hand, we succeeded to detect the features correctly because the technique executes template matching again, as shown in Figure 9 (Right). Here colors of the result rectangles are defined in Table 1, and colors of the detected points are defined in Table 2.

Table 2. Colors of the detected points

	Color
Usual tracking points	Green
Points detected by the partial template matching	Yellow
Points detected by the position estimation	Red

Frame15

Frame53

Figure 9. Tracking by Lucas Kanade method(Left) and our method(Right)

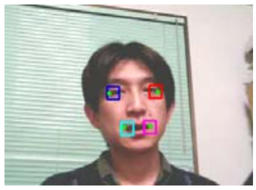

Figure 10. The initialization of another subject's video

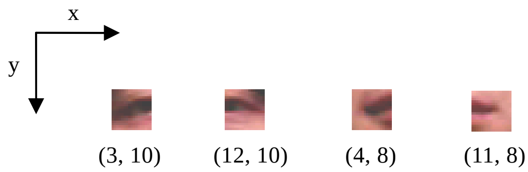

Figure 11. The template images and values of feature points (right eye side, left eye side, right lip corner, left lip corner)

Frame170

Frame229

Frame235

Figure 12. Tracking by Lucas Kanade method(Left) and our method(Right) on another subject

Figures 10 to 12 show experiments on another subject's results. We took his facial video (352x240pixels) and Figure 11 shows template images (17x17pixels) and the coordinates of the feature points. And we defined the colors of the result rectangles same as the first subject, as described in Table 1. Around frame 170 of his facial video, his hand occludes his face. In this case, Figure 12 shows that Lucas-Kanade method fails to track the points from frame 170 to 235. On the other hand, our technique executes the exception handling. It shows the result of template matching for limited region (cyan rectangle) and the estimated point (red point). In this case, since the technique determined the result of template matching is not reliable, it estimated the proper point. In frame 229, since the occlusion was released, the technique discovered the proper feature point by the template matching, and defined the new feature point (yellow point). After this handling, we succeeded to keep tracking features in the following frames, as shown in frame 235.

5. Conclusion

This paper presented a robust technique for facial feature tracking which successfully recovers unusual detections. Our technique realizes a robust facial feature tracking under natural conditions by considering reliability of detected points. It applies fast Lucas-Kanade method to track features during usual situations, and applies template matching during exception handling. We think that this combination is good for the balance of computation time and reliability.

Our final goal on this study is application to the presented technique for understanding user's delicate feelings, or psychology. For example, we would like to develop following systems applying the presented technique:

- The system serving contents which suit user's condition.
- The avatar which supports user's communication.

Toward the above goal, we would like to experience with more facial feature points which are important to understand human's facial expressions. Moreover, we would like to measure the processing time of the unusual detection mentioned in Section 3.2 and analyze the accuracy of a tracking result, for the development of real-time face tracking system.

References

[1] D. Cristinacce and T. Cootes, "Feature Detection and Tracking with Constrained Local Models", Proc. British Machine Vision Conference, Vol. 3, pp. 929-938, 2006.

[2] K. Hotta et al. , “A Robust Face Detector under Partial Occlusion”, IEEE International Conference on Image Processing, pp. 597-600, 2004.10.

[3] Y. Matsumoto et al. , ”Development of Real-time Face and Gaze Measurement System and Applications to Intelligent User Interfaces” , Information Processing Society of Japan, Transactions on Computer Vision and Image Media, Vol. 47, No. SIG15(CVIM16), pp. 10-21, 2006.10. (in Japanese)

[4] Institute for Information Technology, “NRC-IIT Facial Video Database”, http://synapse.vit.iit.nrc.ca/db/video/faces/cvglab/

[5] B. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision” Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), April, 1981, pp. 674-679.

[6] K. Hayashi , M. Hashimoto and K. Sumi “Head Direction Estimation with Robust and Accurate Tracker to Facial Features”, The Institute of Electronics, Information and Communication Engineers, Vol. J84-D-Ⅱ No.8 pp1762-1771, 2001.8.

[7] K. Fukui and O. Yamaguchi, “Facial Feature Point Extraction Method Based on Combination of Shape Extraction”, The Institute of Electronics, Information and Communication Engineers, Vol. J80-D-Ⅱ No.8 pp2170-2177, 1997.8.

[8] H. Ikeda and K. Moriya, “High-precision Facial Region Detection Method and Its Utilization”, Meeting on Image Recognition and Understanding 2006(MIRU2006), pp.418-423, 2006.7.

[9] J. F. Cohn et al. , ” Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression”, Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), pp.396-401, 1998. 4.

[10] H. Shimoda et al. , “A Prototype of a Real-time Expression Recognition System from Dynamic Facial Image”, Human Interface Society Vol.1, No.2, 1999.

[11] K. Sugiyama et al. ,“Drowsiness Level Detection by Measuring Blinks Utilizing Image Processing”, R&D Review of Toyota CRDL, Vol. 31, No. 2, 1996. 6.

[12] J. F. Cohn et al. , “Multimodal Coordination of Facial Action, Head Rotation, and Eye Motion during Spontaneous Smiles”, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

[13] K. Oka and Youichi Sato, “Real-time Modeling of Face Deformation for 3D Head Pose Estimation”, Meeting on Image Recognition and Understanding 2005(MIRU2005), pp.167-173, 2005.7.