Motion Pattern Detection for Dynamic Facial Expression Understanding

Hiroshi Mizoguchi[1], Seiyo Hiramatsu[2], Kazuyuki Hiraoka[2], Masaru Tanaka[2],
Takaomi Shigehara[2], and Taketoshi Mishima[2]
[1] Department of Mechanical Engineering, Faculty of Science and Technology,
Tokyo University of Science
2641 Yamazaki, Noda 278-8510, Japan
Tel. +81-4-7124-1501 (ext.3928), Fax.: +81-4-7123-9814, e-mail: hm@rs.noda.tus.ac.jp
[2] Department of Information and Computer Science, Faculty of Engineering,
Saitama University
255 Shimo Okubo, Saitama 338-8570, Japan
e-mail : {seiyo | hira | mtanaka | sigehara | mishima}@me.ics.saitama-u.ac.jp

Abstract: In this paper the authors present their attempt to realize a motion pattern detector that finds specified sequence of image from input motion image. The detector is intended to be used for time-varying facial expression understanding. Needless to say, facial expression understanding by machine is crucial and enriches quality of human machine interaction. Among various facial expressions, like blinking, there must be such expressions that can not be recognized if input expression image is static. Still image of blinking can not be distinguished from sleeping. In this paper, the authors discuss implementation of their motion pattern detector and describe experiments using the detector. Experimental results confirm the feasibility of the idea behind the implemented detector.

1. Introduction

In this paper, the authors present their attempt to realize a motion pattern detector that finds specified sequence of motion image. The authors study human collaborative machines. As a part of the study, they work on facial expression understanding.

Most research on facial expression understanding treats still image. However, among various facial expressions, there must be such expressions that can be hardly recognized if input image is static or partial sequence of motion. For example, a still image of closing eye cannot be distinguished whether it is in the middle of sleeping or blinking. Similar examples are smiling, nodding, cocking one's head in skepticism, and so forth. In this paper, the authors call such facial expression that the meaning appears only in motion image as *dynamic facial expression*. If computers could understand such dynamic facial expressions, it would contribute to improvement of computer human interface.

To understand the dynamic facial expression, at first its detection is obviously needed. Detecting the dynamic facial expression is equivalent to a specific motion pattern detection from a motion image sequence. Of course duration and velocity of a dynamic facial expression are not fixed and vary for various instances. Thus the authors modify their face detection method for still image [1][2] to detect the specific motion pattern. The face detection method utilizes correlation based pattern matching and is robust against variation of face position and size. The authors extend the method to time domain and apply to motion image. Original face template for the pattern matching is replaced by the specific motion pattern. The duration and velocity fluctuation of the dynamic facial expression can be absorbed by reducing the input motion image along time axis.

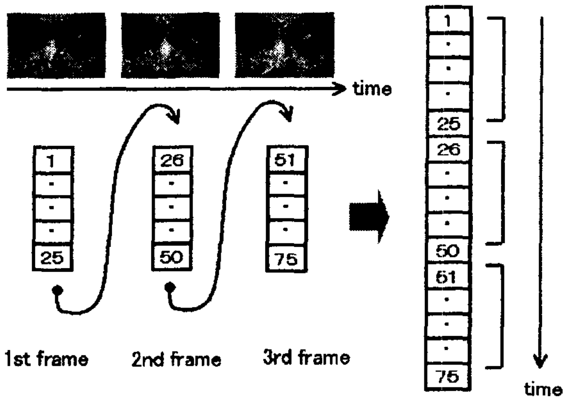

Figure 1. Vector Representation of Motion Image Frames

2. Motion Pattern Detection

To detect a dynamic facial expression, it is necessary to prepare a motion pattern, which represents the dynamic facial expression, as the template for the matching. As shown in fig. 1, a sequence of motion image frames is treated as a higher dimensional vector. Since the method is literally view-based, the template can be easily specified by simply extracting desired partial sequence from any motion image. The input motion image sequence to be searched should also be prepared. Since non-face image is meaningless "noise" for the facial expression detection, face part is extracted from the input image by using our face detection system [1][2]. For the convenience of later processing, the extracted face images are resized to the same size and formed a face image sequence.

Cross-correlation value of the template and a partial sequence of the input motion image is calculated in the template matching process. After the cross-correlation value is calculated at a location, starting point of the partial sequence is shifted to the next frame. Thus the calculation of the cross-correlation is performed at frame by frame. Fig.

2 shows an example of the template matching. The figure plots the correlation value. We can find a peak of the correlation value at a point where a partial sequence similar to the template just starts. In other words, the motion pattern, i.e. the template, can be properly detected.

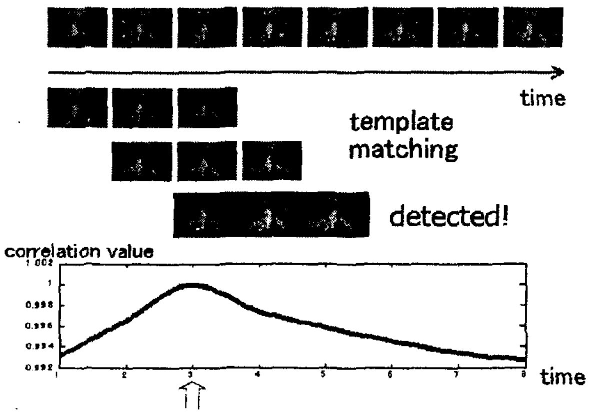

Figure 2. Example of Template Matching

As mentioned above, there must be the fluctuation of duration and velocity of the dynamic facial expression. To absorb the fluctuation, we generate temporally reduced motion image sequences, such as 9 tenth duration, 8 tenths, 7 tenths, and so forth as shown in fig. 3. Then we apply the template matching process to each sequences respectively. Among those results, we choose such sequence that corresponds to the highest correlation value as the detected partial sequence in the input motion image.

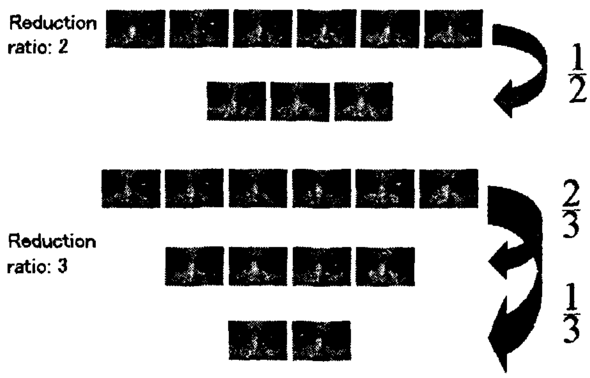

Figure 3. Temporally Reduced Motion Image Sequences

3. Experiments

To confirm feasibility of the idea described above, the authors conduct a experiment. The input motion image includes two chunks of blinking. They are different duration and velocity. The template is a simple four frames sequence of eye open-close-close-open. Fig. 4 shows the result of the template matching. As shown in the figure, there are two peaks at the blinks respectively. While the duration of the template is 4 frames, that of the detected partial sequences are 10 and 20 frames respectively. Thus the variation of the duration and velocity of the dynamic expression can be absorbed. This experimental result supports feasibility of the idea described above.

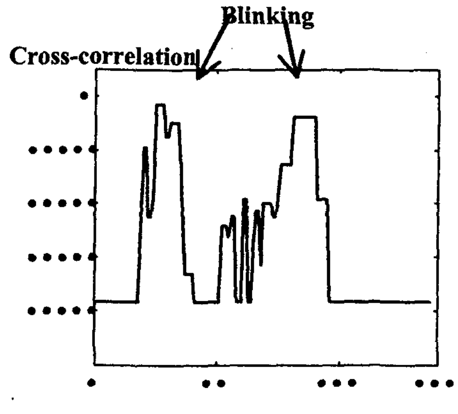

Figure 4. Result of Template Matching for Two Chunks of Blinking

Since this motion pattern detector is really view-based, the same source code of the program can be applied to different dynamic facial expressions, when the template is replaced by other motion pattern. To confirm it, the authors conduct another experiment. In the experiment, the blinking pattern is replaced as the template by a pattern of lip motion. Experimental result is shown in fig. 5. As shown in this figure, the lip motion can be detected without any modification to the original program.

4. Concluding Remarks

This paper describes an attempt to realize a motion pattern detector by the authors. The detector is to be the base for dynamic facial expression understanding. To realize the motion pattern detector, the authors modify their face pattern detector for still image to process motion image. Correlation based spatial pattern matching is extended to temporal pattern matching. Temporal reduction of the motion image sequence is possible to absorb the fluctuation of the duration and velocity of dynamic facial expression. The authors implement a motion pattern detector and conduct experiments using the implemented detector. Experimental results confirm the feasibility of the idea behind the implemented detector. By simply changing the template, different motion patterns can be detected without any modification to the detector. This feature is also confirmed experimentally. Real-time detection is a future work.

Acknowledgements

This work has been partly supported by CREST of JST(Japan Science and Technology) 279102. The work also has been supported in part by Grant-in-Aid for Scientific Research (C) of JSPS.

2 shows an example of the template matching. The figure plots the correlation value. We can find a peak of the correlation value at a point where a partial sequence similar to the template just starts. In other words, the motion pattern, i.e. the template, can be properly detected.

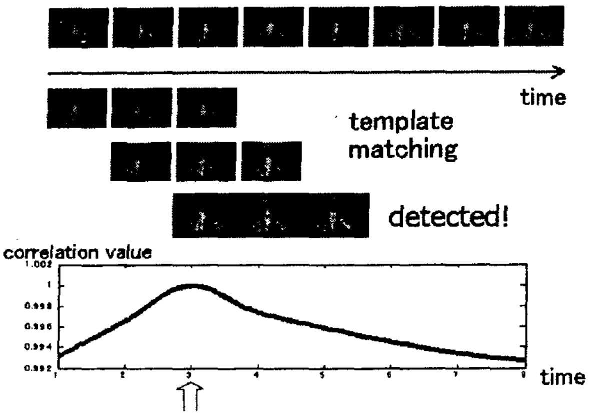

Figure 2. Example of Template Matching

As mentioned above, there must be the fluctuation of duration and velocity of the dynamic facial expression. To absorb the fluctuation, we generate temporally reduced motion image sequences, such as 9 tenth duration, 8 tenths, 7 tenths, and so forth as shown in fig. 3. Then we apply the template matching process to each sequences respectively. Among those results, we choose such sequence that corresponds to the highest correlation value as the detected partial sequence in the input motion image.

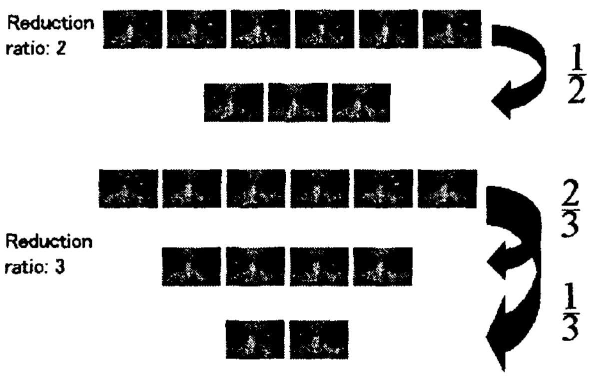

Figure 3. Temporally Reduced Motion Image Sequences

3. Experiments

To confirm feasibility of the idea described above, the authors conduct a experiment. The input motion image includes two chunks of blinking. They are different duration and velocity. The template is a simple four frames sequence of eye open-close-close-open. Fig. 4 shows the result of the template matching. As shown in the figure, there are two peaks at the blinks respectively. While the duration of the template is 4 frames, that of the detected partial sequences are 10 and 20 frames respectively. Thus the variation of the duration and velocity of the dynamic expression can be absorbed. This experimental result supports feasibility of the idea described above.

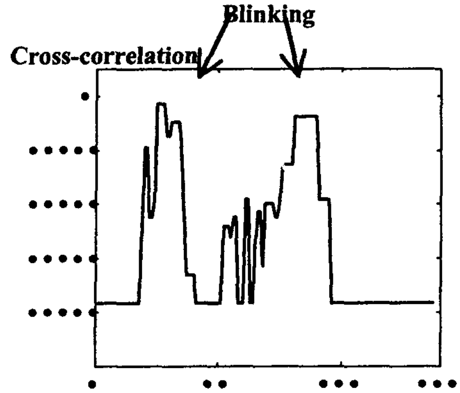

Figure 4. Result of Template Matching for Two Chunks of Blinking

Since this motion pattern detector is really view-based, the same source code of the program can be applied to different dynamic facial expressions, when the template is replaced by other motion pattern. To confirm it, the authors conduct another experiment. In the experiment, the blinking pattern is replaced as the template by a pattern of lip motion. Experimental result is shown in fig. 5. As shown in this figure, the lip motion can be detected without any modification to the original program.

4. Concluding Remarks

This paper describes an attempt to realize a motion pattern detector by the authors. The detector is to be the base for dynamic facial expression understanding. To realize the motion pattern detector, the authors modify their face pattern detector for still image to process motion image. Correlation based spatial pattern matching is extended to temporal pattern matching. Temporal reduction of the motion image sequence is possible to absorb the fluctuation of the duration and velocity of dynamic facial expression. The authors implement a motion pattern detector and conduct experiments using the implemented detector. Experimental results confirm the feasibility of the idea behind the implemented detector. By simply changing the template, different motion patterns can be detected without any modification to the detector. This feature is also confirmed experimentally. Real-time detection is a future work.

Acknowledgements

This work has been partly supported by CREST of JST(Japan Science and Technology) 279102. The work also has been supported in part by Grant-in-Aid for Scientific Research (C) of JSPS.

References

[1] K. Hidai, T. Kanamori, H. Mizoguchi, et al., Robust Face Detection Against Brightness Fluctuation and Size Variation, Proc. of IROS2000, pp. 1379-1384, 2000.

[2] H. Mizoguchi, et al., Implementing a Face Detection System Practically Robust against Size Variation and Brightness Fluctuation for Distributed Autonomous Human Supporting Robotic Environment, Proc. of DARS'02, 2002. (to appear)

[3] A. Pentland, Looking at People: Sensing for Ubiquitous and Wearable Computing, Trans. on PAMI, Vol.22, No.1, pp. 107-119, 2000.

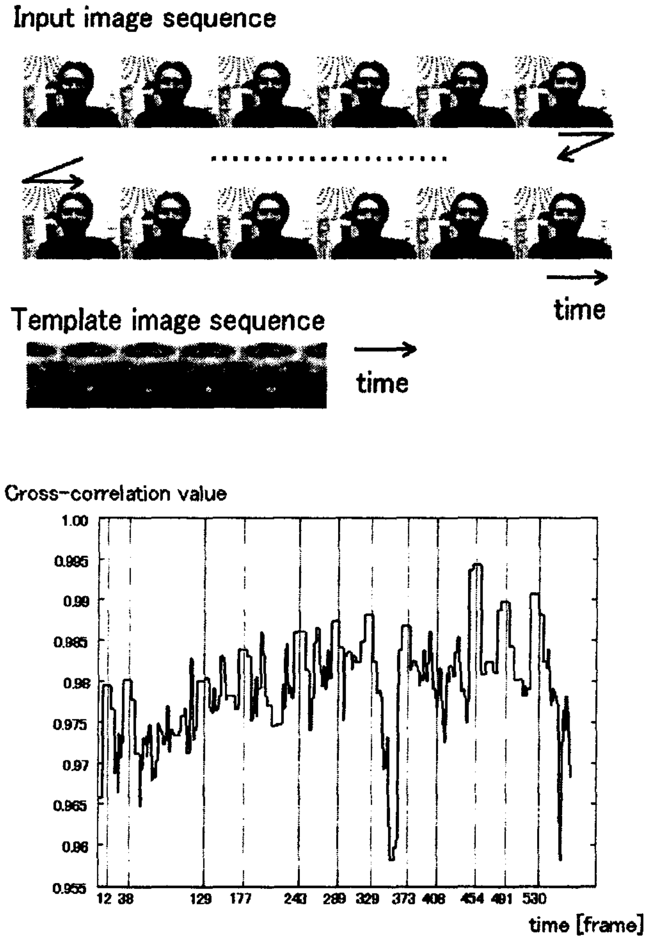

Figure 5. Detection of Lip Motion