Discrimination of Facial Expressions using Velocity Estimation from Image Sequence

Kazuyuki Miura*, Ryosuke Sadakane, Atsushi Osa, and Hidetoshi Miike

** Graduate School of Perceptual Sciences and Design Engineering, Yamaguchi University, 2-16-1, Tokiwadai, Ube-shi, Yamaguchi, 755-8611, Japan*

E-mail: c052fh@yamaguchi-u.ac.jp

We propose a method that classifies human facial expressions of positive, negative, and neutral from image sequence. We utilize two techniques of image sequence processing for the proposed method. The first one is a space-filtering velocimetry, which is a technique of velocity estimation from image sequence. The second technique is optical flow estimation utilizing gradient-based optimization. Optical flow provides local changes of motion field. We combine two techniques for the classification of facial expression. For the evaluation of the proposed method, we carried out two experiments. In the first experiment we confirmed recognition rates of volunteers who judged facial expression in an image sequence. In the second experiment we adopted the proposed method for a discrimination of the facial expression considering the result of the evaluation. Then, we get 60.9% of correctly classification on preliminary experiment.

Key words: *Facial Expressions Analysis, Space-filtering Velocimetry, Optical Flow, Image Sequence.*

Introduction

Visual information like a facial expression or human gesture plays an important role on human-to-human communication. Therefore we can regard that those visual information is also important for the man-machine interface system, which behave human-like response by considering user's state of emotions. This kind of studies aim to realize smooth communication tools on man-machine interface systems, but also aim to understand dynamic behaviors of human emotion.

There have been several studies on gesture recognition. Recently, we also proposed a method of extracting characteristics of human gestures (Miura et al., 2005). By means of the method we tested to 10 gestures and got a high recognition rate within a preliminary experiment. Meanwhile, we can find out a lot of reports about human facial expressions or facial action analysis. For example, using geometrical model of face and optical flow, Essa et al. proposed a method of extraction of facial muscle action (Essa & Pentland, 1994). And Tian et al. utilized facial feature like several feature points of mouth or eyes and wrinkles for analyzing facial expressions (Tian et al, 2001). The major limitation of these feature-based works is that there have to prepare a template of hole or each component of a face.

In this paper we propose a method of classifying human facial expressions as positive, negative and neutral one from an image sequence. This method contained two techniques of image sequence processing, space-filtering velocimetry and optical flow estimation. We carried out two experiments. The first one was an evaluation of facial expressions in image sequence. The second experiment was a discrimination of facial expression by adopting the proposed method. Comparing the experimental results, we confirmed that the proposed method works well within a classification into three facial expression of positive, negative and neutral.

Methods

Firstly we describe a procedure of the proposed method as follows:

Step 1: Acquiring an image sequence contained human facial expression by DV camera.

Step 2: Extracting an eye region (100 x 80 pixels) and a mouth region (180 x 120 pixels) from acquired image sequence (Fig.1, Fig.2).

Step 3: Calculating optical flow and converting the flow field into an image sequence (Fig.3).

Step 4: Executing the space-filtering velocimetry for the sequential flow image computed at step 3. This process means extraction of the characteristics of facial expression.

Step 5: Classifying the facial expression into positive, negative, or neutral one.

Here we mention a bit about *Step 2*. At the present stage we assume that locations of face parts are already detected. And we extract the facial parts in off-line processing. We define the eye region including left eye and eyebrow (see Fig.1) and the mouth region including mouth and nasolabial sulcus (Fig.2).

Secondly we give details of *Step 3* and optical flow estimation. In our method, optical flow calculated by the gradient-based method with spatio-temporal local optimization. After estimating optical flow, the flow fields converted to 8 bits gray scale image sequence as an input of the following step of processing, spatial-filtering velocimetry. Figure 3 shows examples of the flow vector-field image. This sequential flow image includes information of motions.

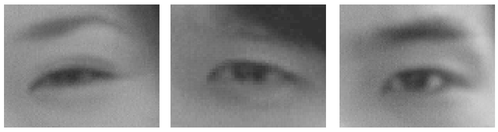
Next we discuss details of *Step 4* and the space-filtering velocimetry. This is a method of extracting speed information of moving objects in image sequence. Equations (1) to (3) give the main process of this velocimetry. We define S(x,y,t) is an acquired image sequence, I(x,y,t) is a filtered signal, and A(t) is a convoluted temporal signal.

$$A(t) = \sum_x \sum_y I(x, y, t), \tag{1}$$

$$I(x, y, t) = S(x, y, t) \cdot \sin\left\{\vec{K} \cdot (\vec{r} - \vec{v}_s t)\right\}, \tag{2}$$

$$\left|\vec{K}\right| = 2\pi / D, \tag{3}$$

where $\vec{K}$ is a wave number vector of a space-filter, $\vec{r}$ is a position vector of an attracted pixel, $\vec{v}_s$ is a translation velocity vector of the space-filter, and D is a wavelength of the space-filter. Two perpendicular directions (horizontal & vertical) are considered as $\vec{K}$ for detecting two-components of the velocity. After convolution calculation, we can get speed information as spectrum data by analyzing frequency of signal *A(t)*. We use a rigid window (32 points) for the First Fourier Transformation, and acquire a temporal change of the spectrum data. According to the sampling theory and sampling rate of the video signal (30 frames/sec), the range of frequency in the spectrum data spreads 0 to 15 Hz on 17 discrete points at each time step. This spectrum data expresses characteristics of facial expression. In both of horizontal and vertical data 7.5 Hz of data (the 9th point) express contributions of non-moving object.

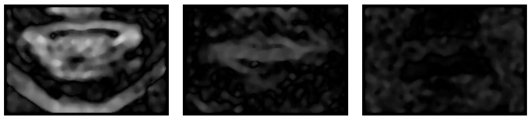


(a) Positive. (b) Negative. (c) Neutral.

Fig.1 The eye region of each facial expression. (100 x 80)

(a) Positive. (b) Negative. (c) Neutral.

Fig.2 The mouth region of each facial expression. (180 x 120)



(a) Positive. (b) Negative. (c) Neutral.

Fig.3 Optical flow images (corresponding Fig.2). Because of little movement of neutral face image (c), almost of pixel values of it are nearly zero.

And more than 7.5 Hz data express left-side (or upper) movements, and less than 7.5 Hz data express right-side (or lower) movements in horizontal (or in vertical) data.

After carrying out the space-filtering velocimetry, we still compress the characteristics of facial expression. At each time step of spectrum data, accumulation of 17 points of spectrum data is carried out with multiplying suitable weights. The weight values are defined as subtracting 7.5 from its frequency value (e.g. the weight value is –7.5 at 0 Hz, the value is +7.5 at 15Hz). This process is carried out to calculate respective spectrum components, horizontal and vertical. Brief outline of output signal in this process is as follows. In Fig.4, we can find out oscillatory areas. This kind of oscillation is caused by simultaneous movements of upper & lower or left & right motion (e.g. blink, mouth open and close). In addition, the upper-side of the horizontal data means left-side movement and the lower-side one means right-side movement (see Fig.4(a), (c)). Similarly, in the vertical data (Fig.4(b), (d)) upper-side of the graph means upper movement and lower-side means lower one.

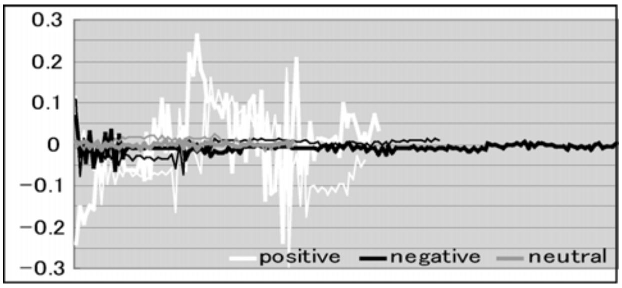
Finally, we mention rules of classification of facial expression. Here, we define four characteristic variables of MouthX(t), MouthY(t), EyeX(t), and EyeY(t). These correspond to horizontal characteristic of mouth region, vertical one of mouth region, horizontal and vertical characteristic of eye region. Rules are as follows:

Rule 1: Image sequence contains positive facial expression, if an absolute value of MouthY(t) is over a threshold T_{upper}.

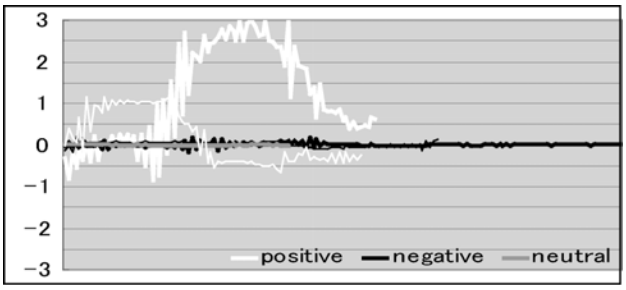
Rule 2: Even while data match to Rule 1, we have to judge negative, if an absolute value of EyeY(t) is over a threshold T_{lower}.

Rule 3: If data has no oscillation in MouthX(t) and match to Rule 1, judgment is neutral.

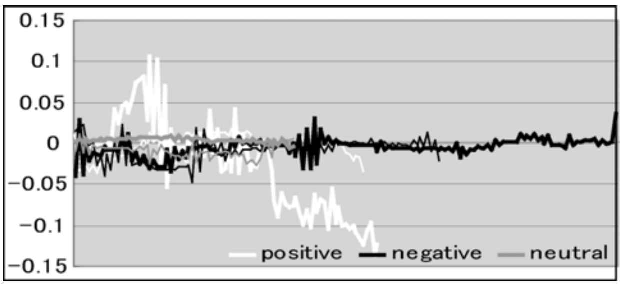
Rule 4: Even while data do not match Rule 1, positive facial expression is included if 3-time-step average value of MouthY(t) or EyeY(t) has a positive value.



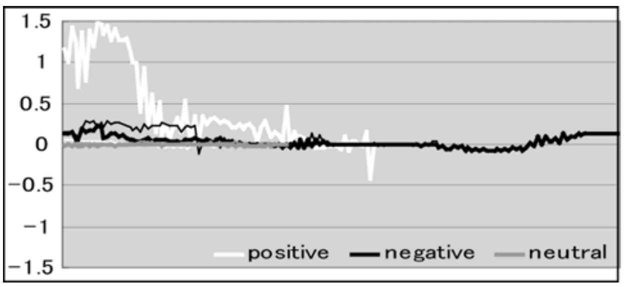
(a) Horizontal data of mouse region (MouthX(t)) ($x10^{12}$)



(b) Vertical data of mouse region (MouthY(t)) ($x10^{12}$)



(c) Horizontal data of eye region (EyeX(t)) ($x10^{12}$)



(d) Vertical data of eye region (EyeY(t)) ($x10^{12}$)

Fig.4 Temporal changes of motion characteristics from mouth and eye region. Couples of white, black, and gray lines show typical facial expressions of positive, negative, and neutral, respectively.

Experiments and results

In this section, we show two experimental results and vilify the proposed method. Image sequences we used in both experiments contained various facial expressions of four Japanese students in Yamaguchi University, Japan (two males and two females). The first experiment was an evaluation of facial expressions. Eight Japanese volunteers classified facial expressions from the image sequences (overall 42 data) into positive, negative or neutral. According to the results of evaluation, we decided 'actual' facial expression of each data. Figure 5 shows a part of experimental result of the evaluation. This shows a number of people who judged that facial expression is positive (or negative / neutral). Then, we adopted an image sequence, which is judged to contain positive (negative / neutral) facial expression by more than 6 volunteers among 8 ones.

The second experiment is a discrimination of the facial expression by the proposed method. According to the results of the evaluation, each actual facial expression of image sequence is decided any of positive, negative or neutral. We picked up 23 data as analysis targets that consist 9 positive facial expressions, 7 negative one and 7 neutral one. Figures 1 and 2 are snapshot of several image sequences for adopted proposed method. And the result of optical flow estimation for mouse region is in Fig.3. Extracted characteristics of typical data showed in Fig.4. The corresponding rate of the proposed method was 60.9% (14/23). (see Table.1 for details.)

Discussion

In this section, we discuss about the proposed method and the rule of discrimination, comparing to another works. In facial expression/action analysis, most of researches apply Facial Action Coding System (FACS), which is human-observe-based system (Ekman & Friesen, 1978). In those researches facial movements are coded into an Action Unit (AU) or combinations of AUs. For example, an emotion of happiness coded AU6 (cheek raiser) and AU12 (lip corner puller), anger is AU2 (outer brow raiser), AU4 (brow lowerer), AU7 (lid tightener), AU24 (lip pressor) and the other AUs combination (Zhang & Ji, 2005). On the other hand, our method

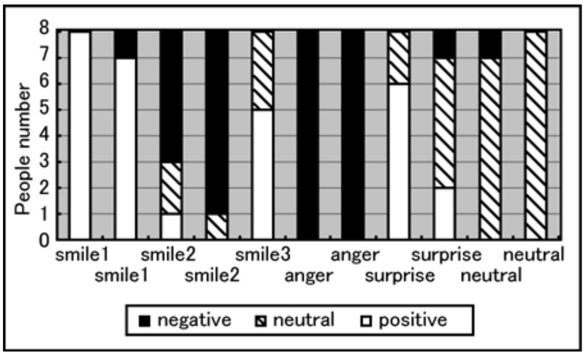


Fig.5 Result of evaluations. White, black and diagonal bars express numbers of people who judged positive, negative and neutral, respectively, for facial expression. ‘Smile1’ means true smile. ‘smile2’ is wry smile, and ‘smile3’ is forced smile.

follows the rules that we defined, however, these rules have some common criterions with FACS coding. Rule 1 simply judges vertical movement around mouth to put it briefly. Similarly, Rule 4 also simply estimates upper movement of mouth or eye region. In FACS coding, for example, these are regarded to correspond to AU6.

Next we describe on a structure of the proposed method. The main technologies of our method are optical flow estimation and the space-filtering velocimetry of sequential flow field image. The space-filtering velocimetry can detect speed information of moving objects in image sequence. However, it is considered that there is an impossibility of detecting correct speed information of an original acquired image, because it contains high pedestal level, which mainly depended on background brightness of the image. Therefore, we need some process to remove the background effect of image or detecting only moving object. The simplest process can be a pixel-based differentiation processing between frames. However, it seems too faint to detect moving object on a facial expression, because that the areas of moving objects are small in the face image, and the differences between frames are also small. According to these ideas we adopt optical flow estimation on the process detecting moving objects. It is required, however, to verify the usefulness of the proposed method. Especially, validity to use original image sequence and/or differential images has to be confirmed.

Another merit of utilizing optical flow is its possibility of tracking mouth and eye regions. At the present stage, we don’t consider the holistic head motion under an assumption that there is few head motion in acquired images. We think that we should take an automatic

Table.1 Results of classification of facial expressions.

		Result of classification		
		Positive	Neutral	Negative
Facial expression	Positive	**4/9**	4/9	1/9
	Neutral	0/7	**5/7**	2/7
	Negative	0/7	2/7	**5/7**

tracking of face parts in future, using the optical flow or more information. Additionally, we consider that it should be included in our facial classification method to utilize usual still image processing. To combine the proposed method and techniques of still image processing, it is expected to progress the proposed method for the practical uses.

Conclusions

We developed a method for classification of human facial expression into positive, negative and neutral. The main idea of our method is based on two motion estimation techniques from image sequences. And we confirmed that our method produced 60.9% of corresponding rate for classification on preliminary experiment. And then, in order to develop our method we’ll try to remove holistic head motion using optical flow or other information in the future works.

References

Irfan A. Essa., & Alex Pentland. (1994). A Vision System for Observing and Extracting Facial Action Parameters. *IEEE CVPR 1994 Conference*, 76-83.

Miura K., Jibu S., Osa A., & Miike H. (2005). Real-time Recognition of Human Gesture by Space-filtering Velocimetry for Sequential Stereo Images. *Meeting on Image Recognition and Understanding 2005 (CD-ROM)*, 1626-1627. (in Japanese)

P. Ekman., & W.V. Friesen. (1978). *The Facial Action Coding System: A Technique for The Measurement of Facial Movement.* San Francisco: Consulting Psychologists Press.

Ying-li Tian., T. Kanade., & Jeffrey F. Cohn. (2001). Recognizing Action Units for Facial Expression Analysis. *IEEE Trans. on PAMI*, **23**, 97-115.

Yongmian Zhang., & Qiang Ji. (2005). Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences. *IEEE Trans. on PAMI*, **27**, 699-714.