MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

AN APPLICATION OF OPTICAL FLOW — EXTRACTION OF FACIAL EXPRESSION —

Kenji Mase

NTT Human Interface Laboratories
Nippon Telegraph and Telephone Corporation
1-2356, Take, Yokosuka 238-03 JAPAN
CSNET(e-mail): mase%nttcvg.ntt.jp@relay.cs.net

ABSTRACT

This report discusses a method that uses optical flow to estimate facial muscle movements which could then be recognized these movements as facial expressions. The human face has several features such as eyes, mouth and nose. Their movements and deformations are due to the contraction and/or relaxation of the facial muscles. Facial skin has the texture of a fine-grained organ and this helps in extracting the optical flow.

We estimate the movement of major facial muscles from the optical-flow field. We evaluate the muscle movement in each muscle window which defines one primary direction of muscle contraction. Finally, we get velocity patterns in terms of time. We will use the feature patterns to recognize facial expressions by means of pattern matching. Our experiments show that the features acquired by this approach are sufficient to recognize certain expressions.

1 Introduction

Just as humans use body language or non-verbal language such as gestures and facial expressions in daily communication, computers should also be able to communicate with humans visually in the future[1, 2]. This report presents a method that uses optical flow to estimate facial muscle movements[3] and then recognizes these movements as various facial expressions.

The human face has several features such as eyes, mouth and nose, whose movements and shape deformations may be extracted in various ways. However, these movements are the result of various combinations of contraction and/or relaxation of the facial muscles. It seems more straightforward to extract the muscle movement from skin displacement than to estimate each muscle action from feature point movement. Facial skin has the texture of a fine-grained organ which becomes a key for optical flow estimation.

Facial expression is a major media for non-verbal language in human to human communication. Mehrabian(1968) claimed that only 7 percent of messages in human communication was conveyed by verbal language while 55 percent was transmitted by facial expressions. Thus, we can say that at least facial expressions convey a lot of information. Advanced research on man-machine interfaces must tackle this issue.

This paper first briefly addresses research on facial expression as reported in various fields. Then a conventional optical flow algorithm is applied to image sequences of facial expressions. The methods to estimate facial muscle movement and interpretations of facial expression are presented. Two examples of estimated muscle motion are shown.

2 Facial Expression and Optical Flow

2.1 Facial Expression

Facial expression has been a research subject of psychology since Darwin's work on emotional expression in the 19th century[4]. Ekman & Friesen[5] developed a system named FACS(Facial Action Coding System) to analyze facial expressions systematically in the field of psychoanalysis. The system and their work made it possible for machine vision and computer graphics researcher to treat facial expressions as an engineering subject. FACS was soon used in generating facial animation(see [6] and its references.) Recently, a few trials to extract Action Units(AU: an element of expression in FACS) from feature deformations were reported.

Choi et. al.[7] used geometric parameters of facial features, such as lip height and angle of eyebrow, and assigned their change to appropriate AUs. It is very difficult to extract those facial features by image processing techniques and they extracted them interactively. Kass et.al.[8] used snakes to determine the deformation of facial features such as lips. Terzopoulos & Waters used that information to duplicate facial expressions in computer animation[9]. Still it seems difficult to fit snakes onto the facial parts having less explicit boundaries such as the lower lip.

Considering how an expression is formed by the contraction of facial muscles, we think deformation of skin provides a better estimation of muscle action than is possible with geometric feature parameters. A conventional gradient based optical flow algorithm can extract skin motion from its subtle texture. Mase & Pentland[10] exploited this concept to extract lip movements and applied it to computer lipreading.

2.2 Facial Muscles and Optical Flow

There are many optical flow estimation algorithms. Generally they are divided into three groups; gradient

Figure 1: Schematic of facial muscles[12]

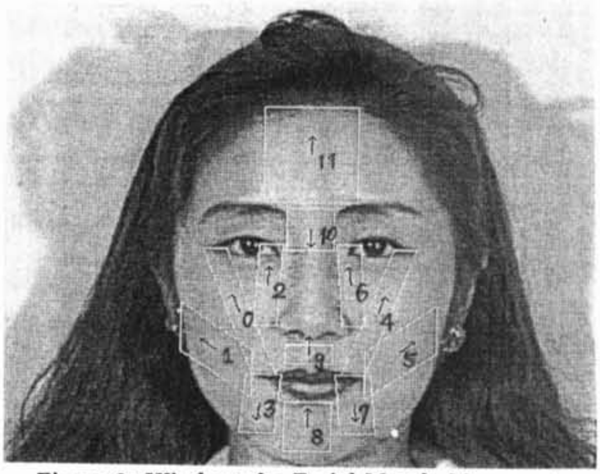


Figure 2: Windows for Facial Muscle Movement

based, correlation based, and filtering based. The conventional gradient based algorithm proposed by Horn & Schunk[11] is simple and works well for facial skin deformation except for the aperture problem and motion discontinuity. Although facial skin is not rigid, we can treat its motion field smooth locally. Thus we adopted Horn & Schunk's optical flow algorithm[11].

3 Extraction of Expression

The muscles of expression interact in a very complex way under the facial skin. The primary muscles involved in expressions are Obicularis Oris around the lips, Venter Frontalis in the forehead, and Buccinator, Zygomaticus and Angli Oris connected to the mouth. Figure 1[12] illustrates the placement of these muscles. Since several muscles contribute locally to an expression, it is of little use to measure all muscle movements. Mase & Pentland[10] analyzed optical flow patterns and extracted two principal motions around the lips for speech articulation.

In this paper, we use windows which define the major direction of muscle contraction(skin deformation). This allows the effect of each muscle to be determined. We calculate the average length of directional components in the major window axis from the optical flow vectors in each window.

$$m_i = \frac{1}{S_i} \int_{window_i} \mathbf{u}(\mathbf{x}) \cdot \mathbf{n}_i \, d\mathbf{x}, \qquad (1)$$

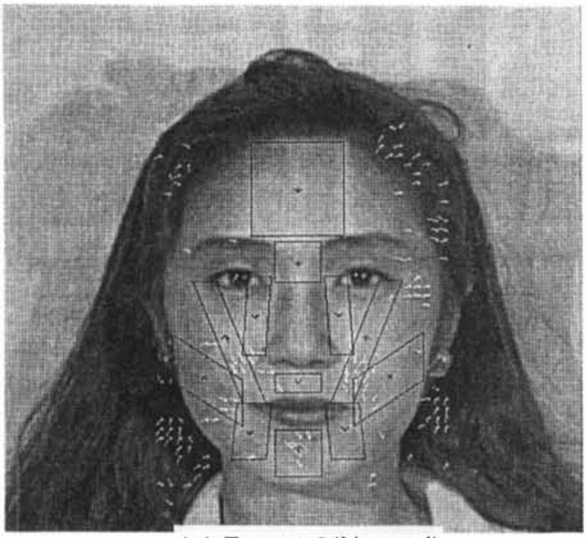
where $\mathbf{u}(\mathbf{x}) \equiv (u(x,y), v(x,y))$ is the optical flow field computed from two successive frames. S_i and $\mathbf{n}_i$ are the area size and the normal vector of the i-th window ($i = 0, \ldots, 11$) respectively. $\mathbf{n}_i$ is illustrated in figure 2 by arrows. Windows are interactively located using feature points as references. m_i is the estimated muscle movement in the i-th window ans is a function of time(frame). Although the movement is velocity, it is also regarded as displacement in a unit period. Figure 3 shows recorded image and estimated muscle velocities during one full cycle of a happy expression. The white and the black arrows are computed optical flow and estimated velocities, respectively. The optical flow field is computed for each pixel and is shown at every three pixels. Image size is 256 by 256 pixels. The estimated motions are functions of time. Figure 4 plots the velocities for each muscles. The facial region was manually clipped from each frame for processing. Figure 5 plots the velocities for another person who is smiling without opening mouth. While muscles m_0(pulling up lip extreme diagonally), m_1(pulling back lip extreme) and m_9(pulling up lip center) show similar velocity curve, muscles m_3(pulling down lip extreme) and m_8(push up lower lip) move differently from Figure 4. This is because of the different action of mouth opening.

4 Recognition of Expression

There are several possible ways to recognize expression from the data acquired by the above method. For instance; (I) If the expressions are simple(one expression), we can develop a vector consisting of muscle velocities at the time at which one arbitrary muscle reaches its maximum velocity. (II) Since the data consists of velocity, it becomes zero at maximum extent of the expression. The integral of the velocity is roughly regarded as the displacement. Then the integrals of the velocity until the time the velocity reaches zero are used for pattern matching. (III) The waveform pattern of each velocity may be used directly for pattern matching. (IV) We relate each muscle movement to a corresponding AU, then analyze it using FACS.

5 Open Problem in Computer Vision

The extraction of facial expressions is regarded as an application of the general computer vision problem of non-rigid body motion analysis. In this paper, the head is roughly fixed so that head flotation was not consid-

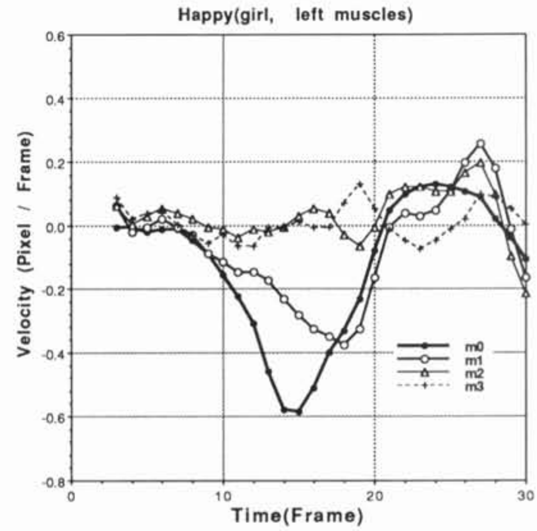


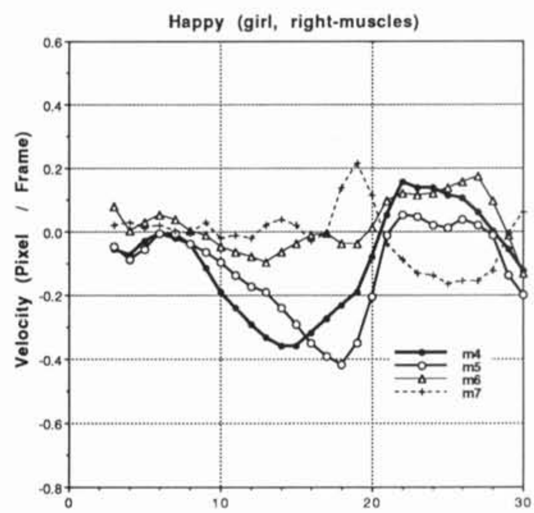
(a) Frame:8(Neutral)

(b) Frame:14(Maximum Speed)

(c) Frame:21(Maximum Position, Relaxing)

Figure 3: Estimated Muscle Motion(Happy)





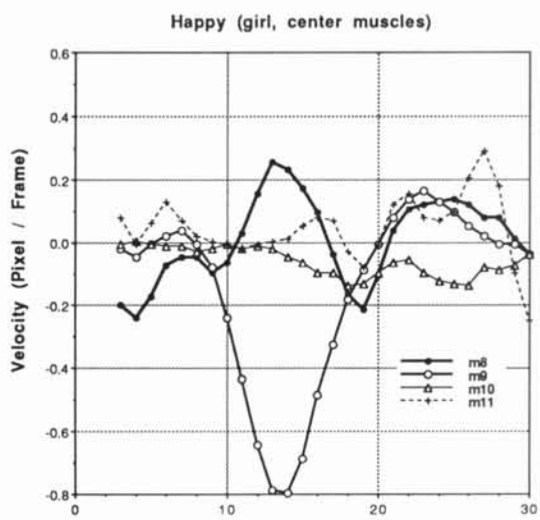
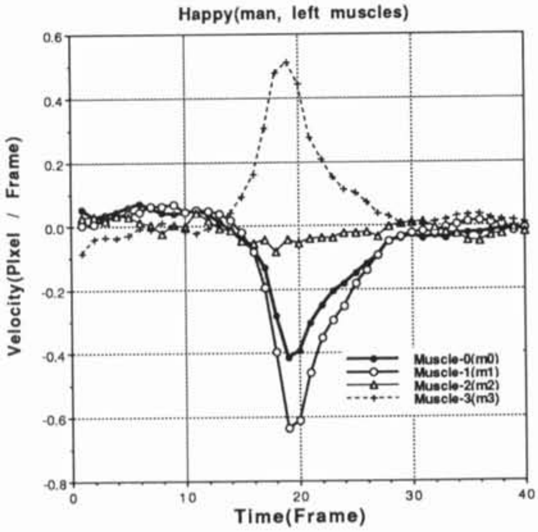


Figure 4: Plot of Estimated Motion(Happy)



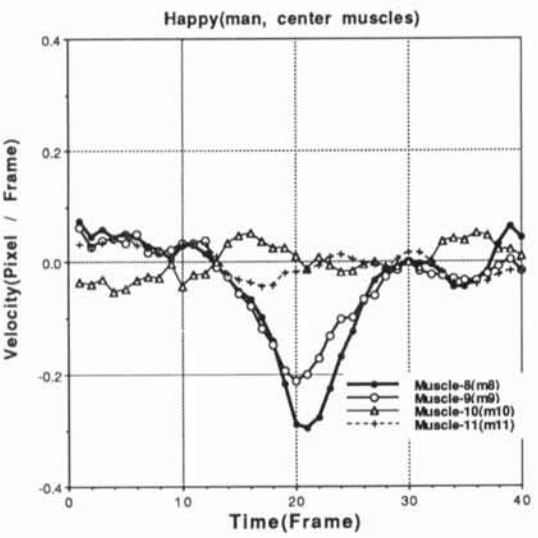


Figure 5: Plot of Estimated Motion(Smiling man)

ered in the optical flow calculations. The goal, however, is to extract global head motion and skin deformation simultaneously from freely taken pictures. Unfortunately, there is no stable facial point which could act as the key for global motion estimation. Even the tip of nose moves in some expressions. We cannot use any assumptions about the proportion of motion, i.e., some deformation exceeds global motion or vise versa.

6 Conclusion

This paper has presented a method to extract facial expressions with optical flow data. It is noted that deformation of facial muscles can be extracted surprisingly well with a conventional optical flow algorithm. One potential drawback of the algorithm is that it produces inconsistent flow at edges and occluding boundaries. The original algorithm should be modified to overcome this problem.

This method is based on motion analysis, rather than static shape analysis. On a general assumption that humans generally use the same muscles to make the same expression, this approach is person independent. The comparison of the example data supports this idea. However, racial or cultural influences must be considered. The experiment shows that even a small change in expression which is not perceivable visually can be extracted easily by this method. Even in a masked facial expression such as a pokerface, micro expressions[13] can be captured. From this point of view, this method can assist psychological science.

We are still researching on expression recognition. Further experimental results will be reported soon.

Acknowledgement

The author would like to thank Dr. Yukio Kobayashi, Dr. Yasuhito Suenaga and Dr. Ken-ichiro Ishii for their valuable advice on this work. Thanks also to the members of Visual Perception Laboratory for their help in the experiment.

References

[1] K. Mase, Y. Suenaga and T. Akimoto: "Head reader: A head motion understanding system for better man-machine interaction", IEEE proc. SMC, pp. 970–974(1987).

[2] K. Mase, Y. Watanabe and Y. Suenaga: "A real time head motion detection system", proc. SPIE 1260, pp. 262–269(1990).

[3] K. Mase: "Detection of facial muscle motion by optical-flow", tech. report IEICE of Japan, PRU89-128, (1990). (in Japanese)

[4] P. Ekman Ed.: "Darwin and Facial Expression", Academic Press, inc., New York, NY(1973).

[5] P. Ekman and W. V. Friesen: "The Facial Action Coding System", Consulting Psychologists Press, Inc., San Francisco, CA(1978).

[6] K. Waters: "A muscle model for animating three-dimensional facial expression", Computer Graphics, **21**, 4, pp. 17–24(1987).

[7] C. S. Choi, H. Harashima and T. Takebe: "Analysis of facial expressions using synthesis rules", PCSJ'89, 8–6, pp. 147–148(1989). (in Japanese)

[8] M. Kass, A. Witkin and D. Terzopoulos: "Snakes: Active contour models", Proc. ICCV-87, pp. 259–268(1987).

[9] D. Terzopoulos and K. Waters: "Analysis of facial images using physical and anatomical models", ICCV'90(1990). (to appear)

[10] K. Mase and A. Pentland: "Automatic lipreading by optical-flow analysis", trans. IEICE of Japan, **73–D–II**, 6, pp. 796–803(1990). (in Japanese, to be translated)

[11] B. K. P. Horn and B. G. Schunk: "Determining optical flow", Artificial Intelligence, **17**, pp. 185–203(1981).

[12] O. Mori et. al.: "Anatomy", Kinbara Publ., Japan(1950). (in Japanese)

[13] P. Ekman: "Telling Lies", Berkley Book(1985).