A Proposal of Pattern Space Trajectory for Gesture Spotting Recognition

†Shigeki Nagaya, ††Susumu Seki, †Ryuichi Oka
†Tsukuba Research Center, ††Novel Functions Sharp Lab.
Real World Computing Partnership (RWCP)
1-6-1 Takezono, Tsukuba, Ibaraki, 305
Japan
nagaya@trc.rwcp.or.jp

Abstract

We propose a new appearance-based feature for real-time gesture recognition from motion images. The feature is the shape of the trajectory caused by human gestures, in the "Pattern Space" defined by the inner-product between patterns on frame images. It has three advantages, 1) it is invariant in term of the target human's position, size and lie, 2) it allows gesture recognition without interpreting frame image contents and 3) there is no costly statistical calculation involved. In this paper, we describe the properties of the gesture trajectory feature, and some experimental results in order to show its applicability to gesture recognition.

1 Introduction

We are researching a "gentle human interface" to realize a flexible information processing system. As one implementation of this, we are developing a CSCW (Computer Supported Cooperative Work) system with the multimodal interface shown in Fig. 1. Gesture recognition technology is very important to achieve this basic human interface.

Time-sequence matching [Takahashi 92, 94] [Darrell 93] is one of the most effective methods for gesture recognition. This method is very popular in speech recognition and has the advantage of recognizing gestures easily without interpreting the content of

Fig. 1 CSCW system with Multimodal Interface.

BMVC 1996 doi:10.5244/C.10.36

each frame image. However it does have several constraints. In this method, it is very important to avoid such constraints in selecting the time-sequence feature derived from frame images.

The following related research has taken the time difference between continuous frames [Takahashi 92] , the silhouette of the target human [Yamato 92] [Wilson 95], the eigenvalue [Turk 91] and the positions of human-body parts extracted by tracking [Bobick 96] into consideration. However these methods have several problems in that: 1) there are some constraints for the target human, position, size, and brightness, 2) they are not suitable with frame-wise calculation for real-time recognition and 3) they are unstable because they have to depend on interpreting the contents of each frame image.

We propose a new feature to resolve these problems together. The feature is in the shape of a trajectory in "Pattern Space", which is defined with the inner product between patterns on continuous frame images. It has three advantages: 1) it is shift, rotation and scale invariant, 2) it does not need to interpret frame image contents, or normalize the size and position of the target human area and 3) it does not need high cost statistical calculation. As a result, our system can recognize gestures without decision a human body area strictly. Furthermore the pattern space trajectory can be calculated frame-wise and gestures can be spotted without detecting their intervals in frame images. In this paper, we describe the pattern space trajectory which is a recognition method using this feature, and the results of examination.

2 Pattern Space Trajectory

2.1 Pattern Space Definition and its Characteristics

Consider the set P of the real-number valued functions that satisfy (1) and are defined on the bounded region $I = \{ (x, y) \mid 0 \le x < W, 0 \le y < H \}$ of a two-dimensional surface.

$$\int_I |f(\vec{x})|^2 d\vec{x} < \infty \qquad (1)$$

[Pattern Functions Definition]

If $f, g \in P$, α are real numbers and $f+g$ and αf are defined as $f(\vec{x}) + g(\vec{x})$, $\alpha f(\vec{x})$ respectively, then P becomes a vector space. The image pattern of a frame of video can be considered to be a point in P if we make $f(\vec{x})$ the brightness of the point at the frame coordinates defined by $\vec{x}$ in the $W \times H$ frame image. This means that patterns can be thought of as vectors. We will therefore refer to P as the "*Pattern Space*".

[Inner Product-Definition]

We can define the inner product of any two patterns in Pattern Space as

$$(f, \ g) = \int_I f(\vec{x})\, g(\vec{x})\, d\vec{x} \qquad (2)$$

and the norm of f naturally becomes $\|f\|=\sqrt{(f,\ f)}$. If we then consider the angle θ between f and g, we can use this definition of the inner product to produce Eq. (3).

$$\cos\theta=\frac{(f,\ g)}{\|f\|\|g\|} \tag{3}$$

Let us next consider the characteristics of this pattern space for situations in which the object of recognition does not extend beyond the boundaries of the image frame.

[Invariance for Congruent Transformation]

Motion in which both the shape and size of an object are maintained includes translation and rotation (including reflection). If we take congruent transformation of this type $\vec{x}\rightarrow\vec{x}'$, we can describe it in terms of a translation $\vec{t}$ and the orthogonal matrix O, as shown in Eq. (4):

$$x'=Ox+t \tag{4}$$

In this case, the transformation $f'=\Lambda(f)$ that takes place in pattern space is:

$$f'(\vec{x})=f(\vec{x}')=f(O\vec{x}+\vec{t}) \tag{5}$$

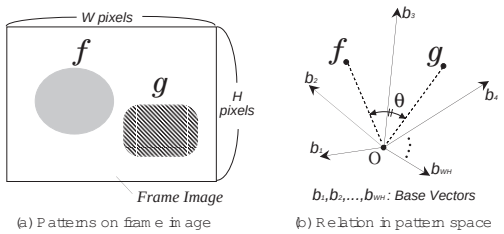
Further, for any real number α, $\Lambda(f+g)=\Lambda(f)+\Lambda(g)$ becomes $\Lambda(\alpha f)=\alpha\Lambda(f)$, which shows that Λ is a linear transformation in pattern space.

For any two patterns f and g, for motion of this type in which the objects do not move out of the frame, we have:

$$\begin{aligned}(f',\ g')&=\int_I f(O\vec{x}+\vec{t})g(O\vec{x}+\vec{t})d\vec{x}\\&=\int_I f(\vec{x}')g(\vec{x}')\|O^{-1}\|d\vec{x}'\\&=(f,\ g)\end{aligned} \tag{6}$$

which shows that the pattern-space inner product is preserved. In other words, this type of motion in patterns on the frame image corresponds to orthogonal transformations in pattern space.

We will now generalize on the discussion so far, and expand our motion to include affine transformations. In place of the orthogonal matrix O, we will use the nonzero



(a) Patterns on frame image (b) Relation in pattern space

Fig. 2 Definition and character of pattern space.

determinant two-dimensional matrix A, for which $\vec{x}' = A\vec{x} + \vec{t}$ In this case, the inner product becomes that shown in Eq. (7), and is generally not preserved.

$$(f',\ g') = \int_I f(\vec{x}')g(\vec{x}')\left\|A^{-1}\right\|d\vec{x}' = \frac{(f,\ g)}{\|A\|} \tag{7}$$

The angle θ' between f' and g' becomes that shown in Eq. (8), however, and the angle θ between f and g is preserved.

$$\cos\theta' = \frac{(f',\ g')}{\|f'\|\|g'\|} = \frac{(f,\ g)}{\|f\|\|g\|} = \cos\theta \tag{8}$$

In other words, for transformations that are similar on the frame images, the angle formed by the two patterns in pattern space is preserved (Fig. 2).

2.2 Trajectory in Pattern Space

[Trajectory of Motion Image]

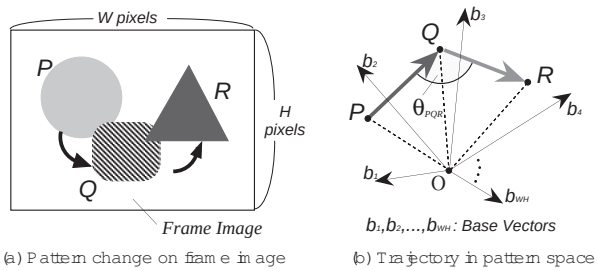
Let us now consider what happens when changes occur over time, as in full-motion video. We will define $f(\vec{x}, t)$ as the brightness at frame position $\vec{x}$ at time t. If t is held constant, brightness becomes a function of $\vec{x}$, and what we have can be thought of as the angle point in pattern space. Based on this, we can define $f(t)$ as the pattern at time t.

If the actual world is filmed with a sufficiently short time-resolution Δt, $f(\vec{x}, t)$ can be thought of as being temporally and spatially continuous. This means that, for a small positive value ε, we can be sure that $\left|f(\vec{x}, t+\Delta t) - f(\vec{x}, t)\right| < \varepsilon$. Thus, if ε is sufficiently small, the norm of $f(t+\Delta t) - f(t)$ in pattern space approaches zero:

$$\begin{aligned}\left\|f(t+\Delta t) - f(t)\right\| &= \sqrt{\int_I \left(f(\vec{x}, t+\Delta t) - f(\vec{x}, t)\right)^2 d\vec{x}} \\ &\leqq \varepsilon\sqrt{WH}\end{aligned} \tag{9}$$

In this way, temporal transitions in patterns can be treated as continuous trajectories traced in pattern space.

Consider a pattern (Fig. 3) that changes P→Q→R. Let us define the angle that is formed in pattern space by PQ, QR as θ_{PQR}.



(a) Pattern change on frame image (b) Trajectory in pattern space

Fig. 3 Trajectory along pattern transition.

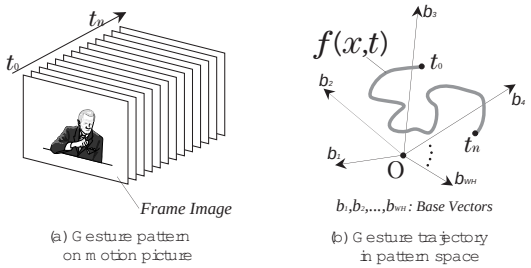


Fig. 4 Definition of Pattern Space trajectory.

Fig. 5 Human action "Side Extensions"

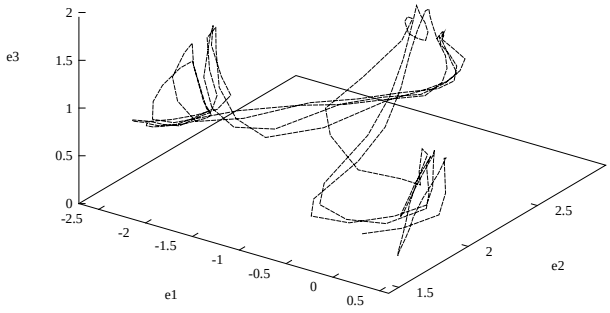


Fig. 6 Pattern Space Trajectory of "Side extensions" projected on 3D space.
(By using principal-component analysis. The sum of proportions for eigenvectors e1, e2, e3 are 0.853.)

When a pattern undergoes a linear transformation within the frame, a similar transformation occurs in the corresponding pattern space, so the angle θ'_{PQR} is equal to θ_{PQR}.

Supposing that the object of recognition does not extend beyond the boundaries of the image frame, the angle q between f and g is preserved for any congruent transformation [Nagaya 95]. In other words, for transformations that are similar on the frame images, the angle formed by the two patterns in pattern space is preserved.

[Pattern Space Trajectory by Human Gesture]

As shown in Fig. 4, if a sequence of temporal changes within the frame is due to a

gesture by a single human subject, the corresponding path are traced in pattern space. To preserve the angle formed by the two patterns, the trajectory in pattern space should be invariant across changes in that subject's position or apparent size and should have a shape that is characteristic of that particular type of gesture.

The human action "Side Extensions" is shown in Fig. 5, and the three-dimensional projection of the resulting pattern space trajectory is shown in Fig. 6. The projection was passed through principal-component analysis. Fig. 6 shows that the trajectories for repeated gesture almost agree regardless of minute variations in action.

3 Gesture Recognition Method

3.1 Variations in Single Gesture

Figure 6 also shows, minute variations are produced each time the gesture is repeated, causing slight corresponding changes in the trajectories. To absorb these real variations, we used polygonal approximations of the trajectory curves consisting of representative points to characterize the trajectory shape. These representative points are determined whether the curvature of the trajectory was local minimum or local maximum.

[Curvature of Pattern Space trajectory]

Let $f(s)$ be a trajectory in pattern space, where s is the length of the trajectory as measured from a specified base-point on that trajectory, and $f(s)$ is a function of time (i.e. t). The unit vector in a direction tangential to the trajectory's direction t, then becomes:

$$\tau = \frac{df(s)}{ds} \tag{10}$$

If we then rewrite this with K as the curvature, $\rho(\geq 0)$ as the radius of curvature, and $\boldsymbol{\upsilon}$ as the unit vector in a direction normal to the trajectory, we arrive at

$$K\nu = \frac{1}{\rho}\nu = \frac{d\tau}{ds} = \frac{d^2 f(s)}{ds^2} \tag{11}$$

The linear transformation A on the two-dimensional image then becomes a similar transformation in the pattern space, having a curvature K' of

$$K' = \left| \frac{d^2 f(s/\sqrt{\|\mathrm{A}\|})}{ds^2} \right| = \frac{K}{\|\mathrm{A}\|} \tag{12}$$

which is proportional to the original curvature.

Equation (12) indicates that for any gesture, regardless of the size of the subject on the frame image, the maximum and minimum curvatures will occur at the same positions on the trajectory curve. We have used these local maximums and local minimums as the bases to segment our trajectory.

[Trajectory Segmentation Algorithm]

The algorithm used to segment pattern space trajectories follows. To determine curvature, second-order differences must be calculated. However, simply taking the differences between subsequent frame images produces results easily affected by noise, so instead we adopted a policy of waiting until the inter-frame difference reached a certain unit of length S_{th}.

First, we read in the latest frame image $INow$, and calculate its path length from the base image I_B. If $S_{In}>S_{th}$, we update the base image ($I_B \Leftarrow I_{Now}$) , and calculate the curvature K. Next, we determine the times in the history of the K time-series at which it reached its highest and lowest values, and store the frame images for those instants as motion elements ($N_{now} \Leftarrow I_{now}$). Finally, we calculate the angle θ that the last three motion elements define in pattern space, and go back to repeat the entire procedure.

Via this process, gestures can be broken up into time sequences that have motion elements (specific poses) as their elements. This can be thought of as the operation of segmenting single gestures into multiple components. In this paper, we will refer to this operation of using the polygonal approximation of the trajectories in order to break them up into their component elements as "segmentation."

3.2 Applying Continuous Dynamic Programming

In order to recognize human gestures, we have to judge the shapes of the polygonally approximated trajectories using CDP (Continuous Dynamic Programming) - whether the input frame image sequence matches the models formerly created or not.

We define the local distance between the model and the input image as follows:

$$d_{local} = \sqrt{\left(l_m \cos\theta_m - l_i \cos\theta_i\right)^2 + \left(l_m \sin\theta_m - l_i \sin\theta_i\right)^2} \qquad (13).$$

Here l is the ratio of the lengths of the two vectors linking three consecutive motion elements. Figure 7 plots the definition of the local distance d_{local}. For both the models and the input, the ratio of the Euclidean distances between the three most recent representative points and the angle formed by those points has been used to obtain two points in a polar coordinate system. The Euclidean distance between these two points is then calcu-

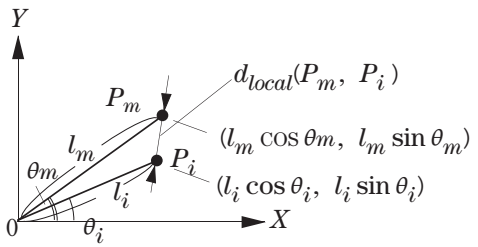


Fig. 7 Definition of local distance for CDP.

Fig. 8 Segmentation result of gesture "Bye"

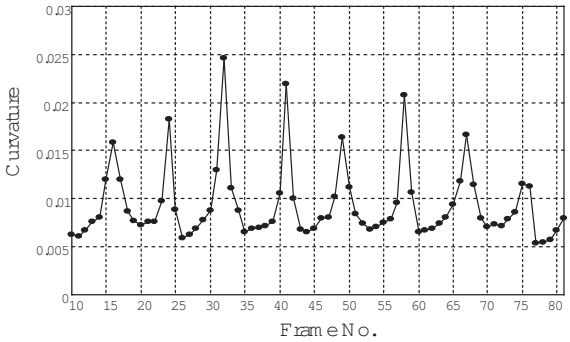


Fig. 9 Curvature sequence at gesture "Bye".

lated. Using this local distance, CDP calculates the total distance for every model and decide the nearest model for the input pattern space trajectory.

4 Experiments

We conducted recognition experiments for the gesture recognition method using the pattern space trajectory. We also analyzed the results using a motion image of the "Bye" gesture (waving good-bye) shown in Fig. 8. It also shows the results of polygonal approximations for the pattern space trajectory as a frame image when the frame-number was the vertex of polygonal approximations. Figure 9 shows the results of expressing the curvature of the gesture trajectory as a time series.

Local-minimums and local-maximums correspond to the frame numbers in Fig. 8 and Fig. 10. Also local-minimums and local-maximums correspond to the split second stop and the quickest point of human action. Namely, the curvature of a pattern space trajectory caused by human action indicates its speed.

Figures10 (a)(b)(c) show the projection of pattern space trajectory onto the local distance plane in Fig. 5. The vortex in Fig. 10 (a) was drawn by repetitive clockwise plotting, regarding the previous OP_i as the Y-axis. The CDP process of our recognition method could be regarded as searching for the nearest vortex shape corresponding to the

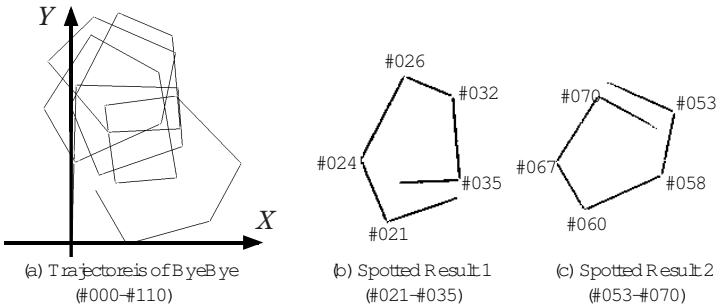


Fig. 10 Projection of Pattern space trajectory on the Local distance plane (Fig. 7)

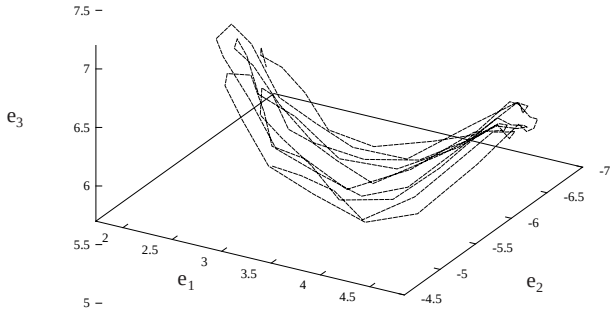


Fig. 11 Pattern space trajectory of "Bye" projected on 3D space
(The sum of proportions for eigenvectors e1, e2, e3 are 0.892.)

particular gesture from input vortex. Figures10 (b)(c) show the search results. This pentagonoid shape expresses the sequence for "Bye". We found that the two sequence images #021-#035 and #053-#070 in Fig. 7 corresponding to Fig. 10 (b) and Fig. 10 (c) are almost the same as human gestures, and that each frame image of the vertices agreed with the other one. From these result, we consider that the pattern space trajectory is effective for human gesture recognition.

5 Conclusion

We described the pattern space trajectory as a new feature for gesture recognition. It has three characteristics: 1) it is invariant for position, lie and size of the target human in the frame images, 2) it allows gesture recognition without interpreting frame image contents, or normalizing the size and position of the target human area and 3) no costly statistical calculation is required. Also we showed its effectiveness in gesture recognition using the results of several experiments.

As a future work, we plan to make recognition tests for large volumes of actual video data, and to evaluate the respective advantages and disadvantages. We also plan to give further consideration to the applications of gesture segmentation functions.

References

[Turk 91] Turk and Pentland : "Eigenfaces for recognition", Journal of Cognitive Neuroscience, No. 3, 1991, pp 71-86.

[Takahashi 92] Takahashi, Seki, Kojima and Oka : "Spotting Recognition of Human Gesture ", Technical Report of IEICE, IE92-136, 1992.

[Takahashi 94] Takahashi, Seki, Kojima and Oka : "Spotting Recognition of Human Gesture from Time-Varying Images", Trans. of IEICE (D-II), J77-DII, 8, 1994, pp 1552-1561.

[Seki 95] Seki, Kojima, Nagaya and Oka : "Efficient gesture recognition algorithm based of Continuous Dynamic Programming", Proc. of RWC Symposium Technical Report, 1995, pp 47-48.

[Nagaya 95] Nagaya, Seki and Oka : "Gesture Recognition Using Multiple Resolution Feature", Technical Report of IEICE, PRU95-99, 1995, pp121-126.

[Yamato 92] Yamato, Ohya and Ishii : "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Models", Proc. of CVPR, 1992, pp 379-385.

[Darrell 93] Darrell and Pentland : "Recognition of Space-Time Gesture using a Distributed Representation", M.I.T. Media Laboratory Vision and Modeling Group Technical Report, No.197,1993.

[Baudel 93] Baudel and Beaudouin-Lafon : "Charade : Remote Control of Object Using Free-Hand Gestures", CACM, Vol. 36, No.7,1993.

[Murase 94] Murase and Nayar : "Learning and Recognition of 3D Object from Appearance", Technical Report of IEICE, PRU93-120, 1994, pp 31-38.

[Sakaguchi 95] Sakaguchi, Ohya and Kishino : "Facial Expression Recognition from Image Sequence Using Hidden Markov Model", The Journal of the Institute of Television Engineers of Japan, 49-8, 1995, pp 1060-1067.

[Wilson 95] Wilson and Bobick : "Using Configration States for the Representation and Recognition of Gesture", Proc. Of the fifth International Conference of Computer Vision, 1995.

[Bobick 95] Bobick and Davis : "An Appearance-based Representation of Action", M.I.T. Media Laboratory Perceptual Computing Section Technical Report, 1995, No.369.