A Hilbert Warping Method for Camera-based Finger-writing Recognition

Hiroyuki Ishida[1,2], Tomokazu Takahashi[3], Ichiro Ide[1] and Hiroshi Murase[1]

[1] Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan
E-mail: hishi@murase.m.is.nagoya-u.ac.jp
[2]Japan Society for the Promotion of Science, Japan
[3] Gifu Shotoku Gakuen University, Nakauzura 1-38, Gifu-shi, Gifu, 501-6194 Japan

Abstract

We propose a time-warping algorithm for recognizing finger actions by a camera. In the proposed method, an input image sequence is aligned to the reference sequences by phase-synchronization of the analytic signals, and then classified by comparing the cumulative distances. A major benefit of this method is that over-fitting to sequences of incorrect categories is restricted. The proposed method exhibited high recognition accuracy in finger-writing character recognition.

1. Introduction

Camera-based analysis of human behavior has been studied for decades [1]. One of its applications is finger-writing recognition system [2] in which characters written in the air are identified. It has gained attention as a novel means of man-machine interaction [3] because: (1) users can operate computers just by simple finger-actions, and (2) it does not require extra equipments except for a camera.

In [2], finger-writing characters were recognized from trajectories of the finger position. Since the trajectories are nonlinearly warped with respect to the time axis, the dynamic time warping (DTW) method [4] is employed for the sequence alignment; an input sequence is classified to a reference sequence which gives the minimum cumulative distance.

However, the DTW has a drawback for the classification task. Because the DTW finds the best alignment for the reference sequences of all categories, misclassification can occur due to the over-fitting to incorrect categories. To cope with this problem, we propose a "Hilbert warping" method which finds the proper alignment only for the correct category. In the proposed method, the sequences are converted into the form of analytic signals [5]. An important property of the an-

alytic signal is that its instantaneous phase increases constantly. Using this property, both of the sequences are aligned by phase-synchronization of analytic signals. Undesirable over-fitting to incorrect categories is avoided if the sequence alignment is performed by the phase-synchronization.

In this paper, we apply the proposed method to camera-based recognition of finger-writing characters. Figure 1 shows the flow of the proposed method. Firstly, image sequences are converted to time-varying feature vectors by the eigenspace method [6], as proposed in a gesture recognition method [7]. Secondly, each feature value is transformed to an analytic signal. The empirical mode decomposition (EMD) [8] is introduced here to ensure that the phase of the analytic signal becomes monotonic. Finally, the cumulative distance between two sequences are calculated by synchronizing the phase of analytic signals.

This paper is organized as follows: Section 2 introduces the property of analytic signals. In Section 3, the proposed Hilbert warping method is described. Results are presented in Section 4.

2. Analytic signal

An image sequence is transformed to analytic signals [5] for sequence alignment. Let $f(t)$ be a feature value obtained from the t-th image in the sequence. An analytic signal $a(t)$ is composed of the original signal $f(t)$ as the real part and its Hilbert transform $\mathcal{H}[f(t)] = (1/\pi t) * f(t)$ as the imaginary part [5]. It is denoted as

$$a(t) = f(t) + j\mathcal{H}[f(t)] = |a(t)|e^{j\phi(t)}, \quad (1)$$

where $\phi(t)$ is defined as the instantaneous phase. In principle, $\phi(t)$ increases monotonically, which means that $a(t)$ rotates counter-clockwise in the complex plane as illustrated in Fig. 2.
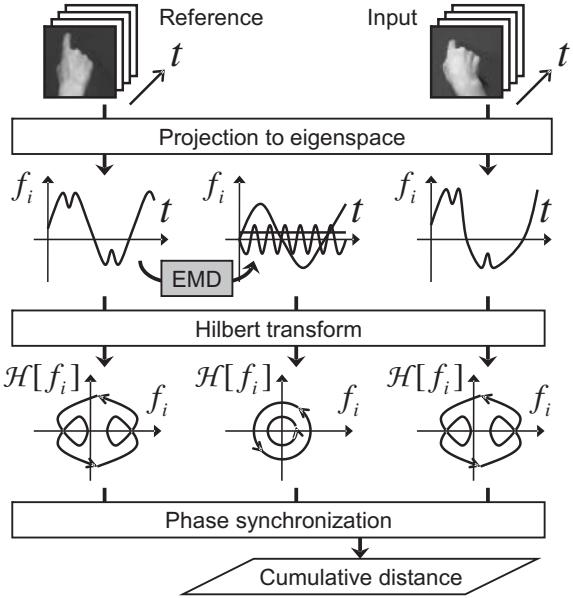
Figure 1. Proposed Hilbert warping method for finger-writing recognition.

Figure 2. Construction of analytic signal. $\mathcal{H}[f(t)]$ **is the Hilbert transform of** $f(t)$.

Figure 3. Feature vectors in eigenspace.

3. Hilbert warping method

The method for the sequence classification is described in this Section. Although a similar approach was proposed in [9], the over-fitting problem in the classification was not taken into consideration. Furthermore, the performance of the sequence alignment was not perfect, since neither EMD nor a feature vector was used. The proposed method ensures proper alignment for a correct category, but avoids over-fitting to incorrect categories.

3.1 Feature vector

Using the eigenspace method, feature vectors are obtained from images.

Initially, the mean vector $\boldsymbol{\mu}$ and an R-dimensional eigenspace $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_R\}$ are constructed from all reference images [6]. Let the t-th image in a sequence be represented by a normalized vector $\boldsymbol{x}(t)$. It is projected on the eigenspace as a point $\boldsymbol{g}(t)$ by

$$\boldsymbol{g}(t) = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_R \end{bmatrix}^\top (\boldsymbol{x}(t) - \boldsymbol{\mu}) \quad (2)$$
$$= \begin{bmatrix} f_1(t) & \cdots & f_R(t) \end{bmatrix}^\top, \quad (3)$$

as shown in Fig. 3. These $f_i(t)$ $(1 \leq i \leq R)$ are used as the feature values for sequence alignment.

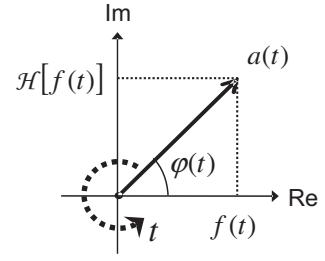### 3.2 Calculation of phase-shift

The feature vector $\boldsymbol{g}(t)$ is converted to an analytic signal vector (ASV) $\boldsymbol{\alpha}(t)$ by transforming each element $f_i(t)$ to an analytic signal $a_i(t)$ using Eq. (1). Thereby, $\boldsymbol{\alpha}(t)$ is represented by

$$\boldsymbol{\alpha}(t) = \begin{bmatrix} a_1(t) & \cdots & a_R(t) \end{bmatrix}^\top. \quad (4)$$

Let $\boldsymbol{\alpha}^{(c)}(t)$ be a reference ASV of category c, and $\boldsymbol{\alpha}^{in}(t)$ be an input ASV. Phase-shift is evaluated from the argument ($\angle$) of the Hermitian inner product $p^{(c)}(t_1, t_2)$ given by

$$p^{(c)}(t_1, t_2) = \begin{bmatrix} \boldsymbol{\alpha}^{(c)}(t_1) \end{bmatrix}^* \boldsymbol{\alpha}^{in}(t_2), \quad (5)$$

where the superscript $*$ denotes the complex conjugate transpose of a vector. In the alignment stage, the frame t_1 corresponding to the frame t_2 is sequentially searched according to the sign of $\angle p^{(c)}(t_1, t_2)$.

3.3 Calculation of phase-shift using EMD

Equation (5) is effective only if the phase is increases monotonically. Unfortunately, such requirement is not satisfied unless the original $f_i(t)$ has a zero-crossing point between local maxima [10]. For example, an analytic signal generated from $f_i(t)$ in Fig. 4 (a) has local
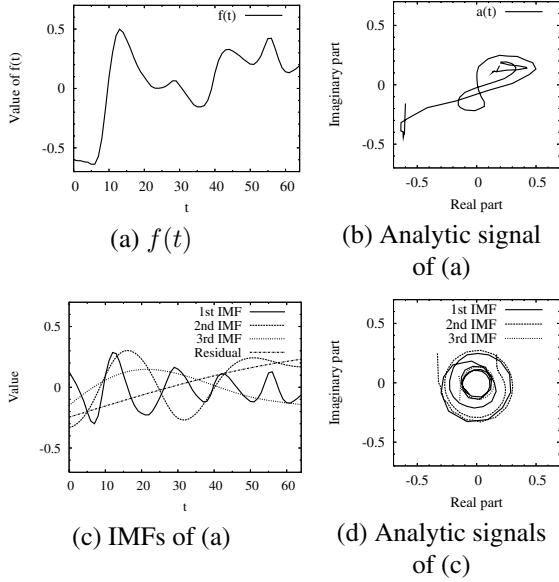
(a) $f(t)$

(b) Analytic signal of (a)

(c) IMFs of (a)

(d) Analytic signals of (c)

Figure 4. Examples of analytic signal.

loops in which the phase decreases (Fig. 4 (b)). In order to eliminate these loops, we apply the EMD [1] to decompose $f_i(t)$ to oscillation functions called "intrinsic mode functions (IMFs)" (Fig. 4 (c), (d)). Some of the IMFs should be excluded during the period where they are considered to make loops. Suppose that $b_i(t)$ is a sum of analytic signals of such IMFs and the residual, the following vector is subtracted from $\boldsymbol{\alpha}(t)$.

$$\boldsymbol{\beta}(t) = \begin{bmatrix} b_1(t) & \cdots & b_R(t) \end{bmatrix}^\top \quad (6)$$

Accordingly, the right side of Eq. (5) is modified as

$$\left[\boldsymbol{\alpha}^{(c)}(t_1) - \boldsymbol{\beta}^{(c)}(t_1)\right]^* \left[\boldsymbol{\alpha}^{in}(t_2) - \boldsymbol{\beta}^{(c)}(t_1)\right]. \quad (7)$$

3.4 Hilbert warping algorithm

The proposed algorithm for the alignment between a reference sequence $(1 \leq t_1 \leq T_1)$ and an input sequence $(1 \leq t_2 \leq T_2)$ is shown in Table 1. As illustrated in Fig. 5, this algorithm explores the time-warping path by tracing the node (t_1, t_2) where $\angle p^{(c)}(t_1, t_2) \approx 0$, and simultaneously computes the cumulative distance $D^{(c)}$. In this algorithm, the frame-to-frame distance $d^{(c)}(t_1, t_2)$ is defined as an Euclidean distance between ASVs by

$$d^{(c)}(t_1, t_2) = \sqrt{\left\| \boldsymbol{\alpha}^{(c)}(t_1) - \boldsymbol{\alpha}^{in}(t_2) \right\|^2}. \quad (8)$$

---

[1]The algorithm is described in [8]. We developed a library hht.h for using the EMD and Hilbert transform in MIST libraries [11].
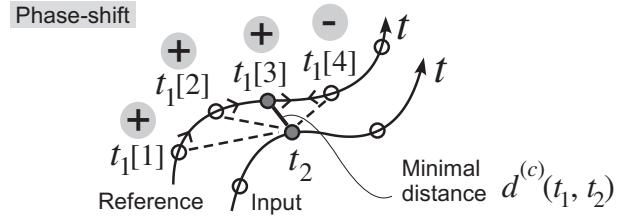
Figure 5. Phase-synchronization process for sequence alignment.

Table 1. Hilbert warping algorithm for calculating the cumulative distance $D^{(c)}$ to category c.

Hilbert warping algorithm
/* Initialization */
1   $D^{(c)} \leftarrow 0, \quad t_1[1] \leftarrow 1, \quad t_2 \leftarrow 1, \quad i \leftarrow 1$
2   **do**
3   **do**
/* Search by the sign of the phase-shift */
4      $t_1[i+1] \leftarrow t_1[i] + \text{sgn}\angle p^{(c)}(t_1[i], t_2)$
5      $i \leftarrow i + 1$
6   **until** sign of $\angle p^{(c)}(t_1[i], t_2)$ changes
/* Distance $d^{(c)}(t_1, t_2)$ is calculated */
7     $D^{(c)} \leftarrow D^{(c)} + \min_i d^{(c)}(t_1[i], t_2)$
8     $t_1[1] \leftarrow \arg\min_{t_1[i]} d^{(c)}(t_1[i], t_2)$
9     $t_2 \leftarrow t_2 + 1, \quad i \leftarrow 1$
10   **until** t_2 reaches the last frame
11   **return** $D^{(c)}$

Finally, the input sequence is classified to

$$\hat{c} = \arg\min_c \Big(D^{(c)} \;+\; \sum_{t_1=1}^{t_1'-1} d^{(c)}(t_1, 1) \\ + \; \sum_{t_1=t_1''+1}^{T_1} d^{(c)}(t_1, T_2) \Big) \quad (9)$$

where t_1' and t_1'' are the frame numbers which are aligned to $t_2 = 1$ and $t_2 = T_2$, respectively.

This method avoids the over-fitting to incorrect categories because the searched path $(\angle p^{(c)}(t_1, t_2) \approx 0)$ does not coincide with the path giving the minimal $D^{(c)}$ if the two sequences cannot be aligned consistently.
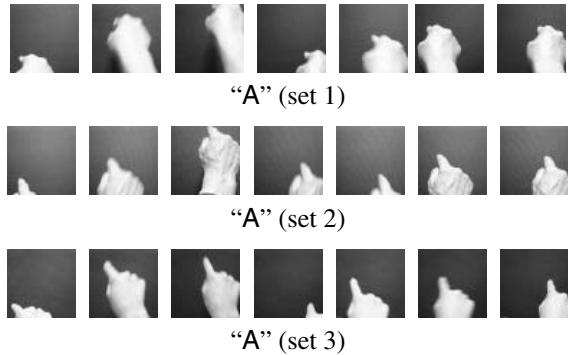
Figure 6. Example of images in datasets.

Figure 7. Recognition rates of finger-writing characters.

Table 2. Confusion matrices for categories H **and** M $(R = 5)$**.**

(a) DTW

Input	Result	
	H	M
H	8/10	1/10
M	2/10	8/10

(b) HW + EMD

Input	Result	
	H	M
H	9/10	0/10
M	0/10	10/10

4. Experimental result

An experiment was conducted using finger-writing character datasets [2] (Fig. 6) which consisted of 10 datasets written by 10 persons individually. Each dataset contained 26 image sequences of finger-writing letters (uppercase A–Z). Recognition rates were evaluated by leave-one-out cross-validation; all the sequences except for an input dataset were used as references. The classification was based on the nearest neighbor rule (1-NN).

The performance of the proposed method (HW+EMD) was compared with the DTW. The cumulative distance $D^{(c)}(T_1, T_2)$ of the DTW was calculated by

$$D^{(c)}(0,0) = 0 \qquad (10)$$
$$D^{(c)}(t_1, t_2) = \min_k \left\{ D^{(c)}(t_1 - k, t_2 - 1) \right\}$$
$$+ \ d^{(c)}(t_1, t_2), \quad (0 \le k \le 2), (11)$$

where $d^{(c)}(t_1, t_2)$ here is an Euclidean distance in the eigenspace. The proposed method was compared also to the simple Hilbert warping method without EMD (simple HW). This simple HW used Eq. (5) instead of Eq. (7).

4.1 Recognition accuracy

Figure 7 shows the recognition rates. The horizontal axis of the graph represents the dimension R of the eigenspace. According to the results, the proposed method outperformed the DTW. For example, categories H and M were distinguished properly (Table 2). Unlike the DTW, the proposed method

---

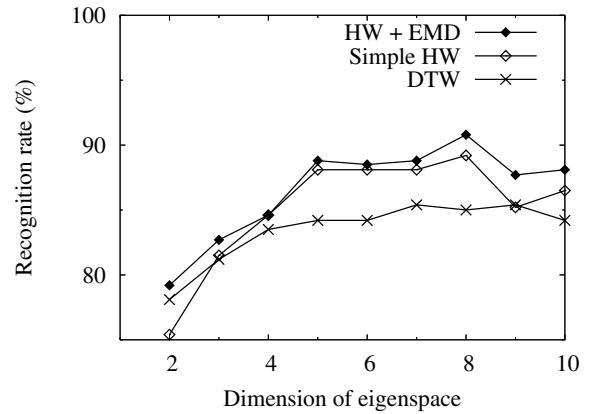[2]The datasets can be downloaded for evaluation freely from *http//www.murase.m.is.nagoya-u.ac.jp/˜hishi/finger-writing.html*.

avoided the over-fitting to category M. Distance matrices $d^{(c)}(t_1, t_2)$ for recognizing category H in dataset 1 are presented in Fig. 8. From the lower-right sub-figure of Fig. 8, we can see that the different category was successfully rejected. As described in 3.4, the search path is composed of ASV pairs with the same instantaneous phase. Accordingly, the phase-synchronization gave the proper time-warping path for the classification.

The results indicate also that the EMD is necessary especially when the dimension of the feature vector is small.
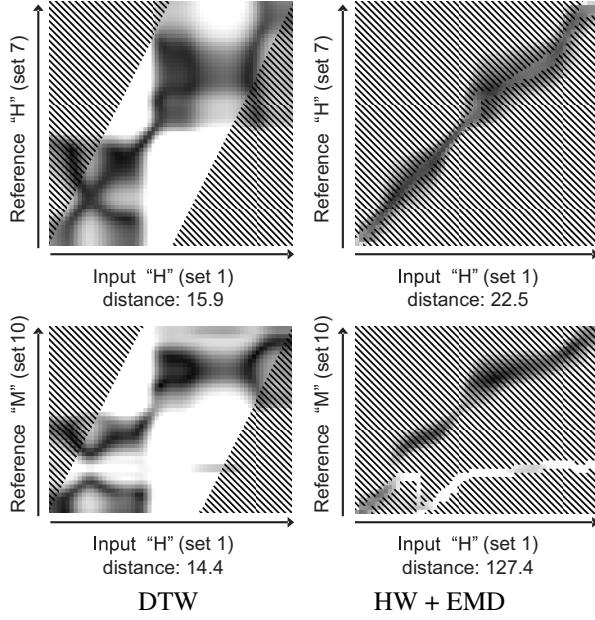
4.2 Computational cost

The computation time for recognizing one sequence is shown in Table 3, where the results of the Hilbert warping methods include also the time required for the Hilbert transform.

The proposed method was approximately three times faster than the conventional DTW, since the calculation of $d^{(c)}(t_1, t_2)$ was drastically reduced as shown in Fig. 8. The EMD was useful also in terms of speed because the monotonicity of the phase contributes to the

Figure 8. Example of distance matrices. Values of $d^{(c)}(t_1, t_2)$ are shown by the intensity (0: black). Nodes filled with oblique lines were not searched.

Table 3. Average computation time for recognizing one sequence. The number of eigenvectors was 5. The experiment was performed on a Pentium IV 3 GHz PC.

	DTW	Simple HW	HW + EMD
Time [ms]	159.2	58.5	53.5

efficient alignment of sequences.

5. Conclusion

In this paper, a Hilbert warping algorithm for sequence classification is proposed. The sequence alignment process is based on the phase-synchronization of analytic signals, which is suitable for classification. The experimental result showed the high classification performance of the proposed method for finger-writing character recognition.

Acknowledgement

References

[1] D. Gavrila, "The visual analysis of human movement: A survey," Computer Vision and Image Understanding, vol.73, no.1, pp.82–98, January 1999.

[2] L. Jin, D. Yang, L. Zhen, and J. Huang, "A novel vision based finger-writing character recognition system," Proc. 18th Int. Conf. on Pattern Recognition vol.1, pp.1104–1107, Hong Kong, China, August 2006.

[3] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.677–695, July 1997.

[4] H. Sakoe and S. Chiba, "A dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoustics, Speech and Signal Processing, vol.26, no.1, pp.43–49, February 1978.

[5] S. Hahn, "Hilbert transforms in signal processing," Artech House, Norwood, Maryland, 1996.

[6] H. Murase and S. Nayar, "Visual learning and recognition of 3-d objects from appearance," Int. Journal of Computer Vision, vol.14, no.1, pp.5–24, January 1995.

[7] T. Watanabe and M. Yachida "Real-time gesture recognition using eigenspace from multi-input image sequences," Proc. 3rd Int. Conf. on Automatic Face and Gesture Recognition, pp.428–433, Nara, Japan, April 1998.

[8] N. Huang and S. Shen, "Hilbert-Huang transform and its applications," World Scientific, Interdisciplinary Mathematical Sciences, vol.5, FarrerRoad, Singapore, 2005.

[9] A. Maheswaran and B. Davis, "Analytical signal processing for pattern recognition," IEEE Trans. Acoustics, Speech and Signal Processing, vol.38, no.9, pp.1645–1649, September 1990.

[10] T. Zagajewski, "Criticism of the definition of instantaneous frequency," Bull. of the Polish Academy of Sciences, vol.37, no.7–12, pp.571–580, November 1989.

[11] MIST project, *http://mist.suenaga.m.is.nagoya-u.ac.jp/trac-en/*.