Detection of finger gesture using singular spectrum transformation

Y. Tamura\textsuperscript{1}, T. Umetani\textsuperscript{1} and H. Nakamura\textsuperscript{2}
\textsuperscript{1} Department of Intelligence and Informatics, Konan University, Okamoto, Higashinada, Kobe, 658-8501, Japan
\textsuperscript{2} Department of Helical Plasma Research, National Institute for Fusion Science, Oroshi, Toki, 509-5292, Japan
E-mail: tamura@konan-u.ac.jp

Abstract. The purpose of this study is to detect finger movement using a singular spectrum transformation method. Human gesture recognition is essential for realizing natural user interfaces. However, constructing a robust, easily installable interface is extremely difficult. Our proposed method uses singular spectrum transformation to classify finger movements. This method robustly classifies gestures and behavior.

1. Introduction
Gesture recognition of finger movement is among the most important issues for realizing natural user interfaces. Touch devices, such as those installed on smart phones, have become widely used. These devices are controlled by fingers. However, a natural interface requires a special device and is applied to a two-dimensional (2D) display. On the other hand, devices such as Microsoft Kinect and Leap motion sensors are operable in 3D space and have wider usability than touch devices. Unfortunately, the user’s indication is less easily detected by 3D interfaces since its accuracy is lower and the degree of freedom is higher than that of the touch devices. Moreover, the raw data generated by 3D sensors is largely obscured by noise.

This study proposes a finger movement detection and classification method based on singular spectrum transformation (SST) \cite{1}. The SST method has been applied in diverse fields such as the earth sciences, physics, and financial research. In human interface research, SST has been applied to the segmentation of human motions \cite{2}. The present study applies SST to finger gesture recognition in noisy environments.

2. System configuration
Our system consists of a data glove and a PC. We used the data glove because it collects data at higher frequency than a Kinect or Leap motion sensor. The 5DT data glove is shown in Fig. 1. This data glove is installed with a single flexure sensor in each finger. The flexure of each finger is measured at 60 Hz, higher than the measuring frequency of the Kinect sensor.

3. Method
Figure 2 presents a flowchart of our proposed system.
In this study, we detect finger movement and the change-point of finger motions by an SST-based method. Since SST is a non-parametric decomposition method, it is suitable for analyzing unsteady datasets. The SST operates as follows: Consider a one-dimensional time series dataset \( x(1), x(2), x(3), \ldots, x(t) \). A \( w \)-component vector \( \mathbf{j} \) is defined as

\[
\mathbf{j}(t-1) = [x(t-w) \ x(t-w+1) \cdots x(t-1)]^T.
\]  

(1)

In addition, the \( n \)-column Hankel matrix (i.e., the trajectory matrix) \( \mathbf{H}(t) \) is given by

\[
\mathbf{H}(t) = \begin{bmatrix} \mathbf{j}(t-n) & \mathbf{j}(t-n+1) & \cdots & \mathbf{j}(t-1) \end{bmatrix}.
\]  

(2)

The Hankel matrix can be decomposed by singular value decomposition as

\[
\mathbf{H}(t) = \mathbf{U}(t)\mathbf{W}(t)\mathbf{V}(t)^T.
\]  

(3)

By applying this method to the time series of finger flexure data in the past and future datasets, the change-points of the motion are extracted from the dataset. The past dataset comprises data from time \( (t-n-w-1) \) to \( (t-1) \), whereas the future dataset contains data from \( (t+s) \) to \( (t+s+w+n) \), where \( s \) is the time shift from the past dataset, and \( w \) and \( n \) in Eqs. 1 and 2, respectively, are window range parameters.

Motion change is clearly detected by a “degree of change” factor, defined as follows. We first construct a hyperplane \( \mathbf{A}_p(t) \) using the \( n \) largest left-singular vectors \( \mathbf{u}_{p1}, \mathbf{u}_{p2}, \cdots, \mathbf{u}_{pn} \) in the past matrix \( \mathbf{U}_p(t) \) of Eq. 3. We also define the largest left-singular vector \( \mathbf{u}_f(t) \) in the future matrix \( \mathbf{U}_f(t) \). If the past and the future datasets are similar, the inner product between vector \( \mathbf{u}_f(t) \) and the vector formed by projecting \( \mathbf{u}_f(t) \) onto the hyperplane \( \mathbf{A}_p(t) \) is approximately 0. In the opposite scenario, this inner product approximates to 1. Thus, we define the factor \( S(t) \) as follows:

\[
S(t) = 1 - \mathbf{u}_f(t)^T \mathbf{P}(t)
\]  

(4)

\[
\mathbf{P}(t) = \mathbf{A}_p(t)^T \mathbf{A}_p(t)^T
\]  

(5)
where all components of matrix $A_p$ are orthonormal bases.

4. Results
In this section, we show the result of the extraction using the SST method. The user wears a data glove on the right hand and sits on a chair next to a desk. The user taps the desk, and the changing finger movements are detected by the flexure sensor.

The original SST is a non-parametric method; however, it is necessary to define the parameters $w, n$, and $s$ in Eqs. 1 and 2. Here we define $w = n$ and $s = -(w + n)/2$. In this definition of $s$, the central datum of the future data becomes $t$ and the window range becomes $2w$. Moreover, in this experiment, the hyperplane $A_p$ was obtained only from $u_{p1}, u_{p2}$.

4.1. Parameter dependency
The parameter $w$ must be set sufficiently long so that all gesture signals are included. On the other hand, if $w$ is too short, large amounts of noise are processed in the output data. Therefore, we tested the dependency of output of $w$. To this end, we set $w = 30, 60, 90$. Since the refresh rate of the data glove is 60Hz, the longest time range is 1.5 s.

The user was then requested to tap the desk 10 times. The results are shown in Fig. 3. Note that the 10th tapping signal cannot be analyzed in $w = 60$ and $w = 90$ since the window range is wider than the time range of this signal. Therefore, if nine peaks are detected, we infer that the method has detected all the gestures.

The analysis reveals that $w = 60$ is a suitable window size under this condition, $w = 30$ is slightly oversensitive whereas $w = 90$ misses some change-point signals. Therefore, we set $w = 60$ in the following experiment.

![Figure 3](image)

**Figure 3.** Dependency of output on the parameter $w$, which determines the window size of signal processing. Left image: Raw data of finger flexure. The rate of finger flexure is normalized from 0.0 to 1.0. Right image: The result of SST calculation. Circles and squares indicate the change-points for $w = 60$ and $w = 90$, respectively.

4.2. Influences of noise
In the previous section, the signal was relatively clear and most of the change-points in the data were distinguishable by human eyes. However, visual detection is much more difficult in a noisy environment. Therefore, we added white noise to the data and reapplied the SST method. Figure 4 shows the result of this experiment. The original signal is normalized from 0.0 to 1.0 and superimposed with white noise. The noise ratio is 0.1, 0.3 or 0.5. This result shows that noise in the input data obscures the change-points to some extent; however, the essential features of the original signal are preserved, especially at noise level 0.5.
Figure 4. Experiment of signal detection in a noisy environment. The upper panels show the signal superimposed with white noise at a ratio of 0.1 (left), 0.3 (center), and 0.5 (right). The lower panels show the corresponding SST results.

5. Conclusion
We detected the change-points of finger motions using a data glove and an SST data processing method. We showed that this method detects finger movements in a noisy environment. Generally, useful data are difficult to extract from raw data and some parameters must be defined by a machine learning method. However, our results demonstrate the ability of the SST method to remove continuous noisy data and extract objective data with only a few specified parameters. Although all the change-points were not stably detected in a noisy environment, the salient features of the original signal were largely preserved. This implies that the SST method is applicable to finger detection in a low-accuracy system. However, a suitable algorithm is required that automatically extracts the change-points from the SST results. Moreover, the signals in this experiment were very peaky since the tapping was performed over short time periods. Therefore, the method requires further testing on long-term gestures.

Acknowledgement
This study is supported by KAKENHI (26330237).

References
[1] J. B. Elsner and A. A. Tsonis 1996 Plenum Press
[2] H. Nakanishi, S. Kanata, H. Hattori, T. Sawaragi and Y. Horiguchi 2011 Journal of Advanced Computational Intelligence and Intelligent Informatics 15 8