Research Article
Gait Signal Analysis with Similarity Measure

Sanghyuk Lee[1] and Seungsoo Shin[2]

[1] *Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China*
[2] *Department of Information Security, Tongmyong University, Sinseonno, Nam-gu, Busan 608-711, Republic of Korea*

Correspondence should be addressed to Seungsoo Shin; shinss@tu.ac.kr

Human gait decision was carried out with the help of similarity measure design. Gait signal was selected through hardware implementation including all in one sensor, control unit, and notebook with connector. Each gait signal was considered as high dimensional data. Therefore, high dimensional data analysis was considered via heuristic technique such as the similarity measure. Each human pattern such as walking, sitting, standing, and stepping up was obtained through experiment. By the results of the analysis, we also identified the overlapped and nonoverlapped data relation, and similarity measure analysis was also illustrated, and comparison with conventional similarity measure was also carried out. Hence, nonoverlapped data similarity analysis provided the clue to solve the similarity of high dimensional data. Considered high dimensional data analysis was designed with consideration of neighborhood information. Proposed similarity measure was applied to identify the behavior patterns of different persons, and different behaviours of the same person. Obtained analysis can be extended to organize health monitoring system for specially elderly persons.

1. Introduction

Analysis on the human gait signal has been studied steadily by numerous researchers [1–3]. The research on the gait signal applies to the field of healthcare development, security system, and another related area. Research methodology is based on how to classify the signal and develop pattern recognition algorithm for the comparable data sets. This algorithm applies the processing of the different signals of the same person and same action signals from multiple persons. Developing methodology for gait discrimination is challenge. However human gait signal has high dimensional characteristics, hence analyzing and designing explicit classifying formula is needed [3].

Generally, human gait signals consist of walking, sitting, standing, stepping up and down, and other usual behavior. Such a usual behavior would be done in house life, then it is quite easy for us to identify when we watch them in real situation. In order to develop a more massive monitoring system and healthcare system to analyze and identify behavior signal of each person, decision and classifying system for

the gait signal is required. More specifically, decision whether he/she is doing in normal activities or not can be applied to the design of health care system. Therefore, obtained research output can be applied to the identification, healthcare, and other related fields. Additionally, more detail checking result even for the healthy people such as athletes can provide useful information whether he/she has suffered from other problems compared to the previous behavior.

To discriminate between different patterns, rational measure obtained from a statistical approach or heuristic approach is needed. By the point of statistical method, signal autocorrelation and cross correlation knowledge are useful because such formula provide us how much the signals are related with each other by the numeric value. Also, it is rather convenient to calculate due to the conventional software such as Matlab toolbox. However, it is not easy for high dimensional data to construct high dimensional correlation/covariance matrices. For heuristic approach, it needs preliminary processing for the signal, such as data redefinition and measure design based on the human thinking. Even the realization of measure based on heuristic

idea is considered, ordinary measure for discrimination has to be considered such as distance. Fortunately, similarity measure design for the vague data has become a more interesting research topic; hence, numerous researchers have been focused on the similarity measure, entropy design problem for fuzzy set, and intuitionistic fuzzy set [4–8].

Similarity measure [9–12] provides useful knowledge to the clustering, and pattern recognition for data sets [13]. However, most of the conventional results were not included in high dimensional data. Human gait signal represents high dimensional data characteristics. So similarity measure design problem for high dimensional data are also needed to deal with the human gait signal. Similarity measure research has rather long history; square error clustering algorithm has been used from the late 1960s [14]. And it was modified to create the cluster program [15]. Naturally, similarity measure topic has been moved to many areas such as statistics [16], machine learning [17, 18], pattern recognition [19], and image processing. Extended research on high-dimensional data can be applied to the security business including fingerprint and iris identification, image processing enhancement, and even big data application recently.

Then, distance between vectors can be organized by norms such as 1-norm, Euclidean-norm, and so forth. Similarity measure is also designed explicitly with the distance norm. Similarity measure design problem for high-dimension needs more considerate approach. Conventionally, the similarity measure has been designed based on the distance measure between two considered data, that is, distance measure was considered information distance between two membership functions. In the similarity measure design with distance measure, measure structure should be related to the same support of the universe of discourse [14, 15]. Additionally, similarity measure consideration on overlapped or nonoverlapped data is needed because many cases of high dimensional data are represented nonoverlapped data structure. With conventional similarity measure, nonoverlapped data analysis is not possible. Hence, similarity measure design for nonoverlapped data should be followed. In order to design similarity measure on nonoverlapped data, neighbor data information was considered. Data has to be affected from the adjacent information, so similarity measure on nonoverlapped data was designed. Inside of literature, artificial data was given to compare with conventional similarity measure; calculation result was also illustrated.

In the following section, preliminary results on the similarity measure on overlapped and nonoverlapped data were introduced. Proposed similarity measure was proved and applied to overlapped and nonoverlapped artificial data. In Section 3, gait signal acquisition system was considered with sensor and data acquisition Gait signal which was also illustrated with different behaviors. High dimensional similarity measure was proposed and proved in Section 4. Similarity measure design for high dimensional data was also discussed by way of norm structure. Similarity calculation results for different behavior and individuals were also shown. Finally, conclusions are followed in Section 5.

2. Similarity Measure Based Distance Property

In order to design similarity measure explicitly, usual measure such as Hamming distance was commonly used as distance measure between sets A and B as follows:

$$d(A, B) = \frac{1}{n}\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|, \tag{1}$$

where $X = \{x_1, x_2, \ldots, x_n\}$, $\mu_A(x_i)$ and $\mu_B(x_i)$ are fuzzy membership function of fuzzy sets A and B at x_i, and $|k|$ was the absolute value of k. The membership function of $A \in F(X)$ is represented by $A = \{(x, \mu_A(x)) \mid x \in X, 0 \leq \mu_A(x) \leq 1\}$, X is total set, and $F(X)$ is the class of all fuzzy sets of X. Similarity measure definition was defined with the help of distance measure [14]. There are numerous similarity measures satisfying the following definition.

Definition 1 (see [14]). A real function $s: F^2 \rightarrow R^+$ is called a similarity measure if s has the following properties:

(S1) $s(A, B) = s(B, A)$, $A, B \in F(X)$;

(S2) $s(D, D^C) = 0$, $D \in P(X)$;

(S3) $s(C, C) = \max_{A,B \in F} s(A, B)$, $C \in F(X)$;

(S4) $A, B, C \in F(X)$, if $A \subset B \subset C$, then $s(A, B) \geq s(A, C)$ and $s(B, C) \geq s(A, C)$;

where $R^+ = [0, \infty)$, $P(X)$ is the class of ordinary sets of X and D^C is the complement set of D. By this definition, numerous similarity measures could be derived. In the following theorem, similarity measures based on distance measure is illustrated.

Theorem 2. *For any set* $A, B \in F(X)$, *if* d *satisfies Hamming distance measure, then*

$$s(A, B) = d((A \cap B), [0]_X) + d((A \cup B), [1]_X) \tag{2}$$

is the similarity measure between sets A and B.

Proof. Commutativity of (S1) is clear from (9) itself; that is, $s(A, B) = s(B, A)$.

For (S2),

$$s(D, D^C) = d((D \cap D^C), [0]_X) + d((D \cup D^C), [1]_X)$$
$$= d([0]_X, [0]_X) + d([1]_X, [1]_X) = 0 \tag{3}$$

is obtained because of $(D \cap D^C) = [0]_X$ and $(D \cup D^C) = [1]_X$, where $[0]_X$ and $[1]_X$ denote zero and one over the whole universe of discourse of X. Hence, (S2) was satisfied.

(S3) is also easy to prove as follows:

$$s(C, C) = d((C \cap C), [0]_X) + d((C \cup C), [1]_X)$$
$$= d(C, [0]_X) + d(C, [1]_X) = 1. \tag{4}$$

It is natural that $s(C, C)$ satisfied maximal value. Finally,

$$d((A \cap B), [0]_X) = d((A \cap C), [0]_X),$$
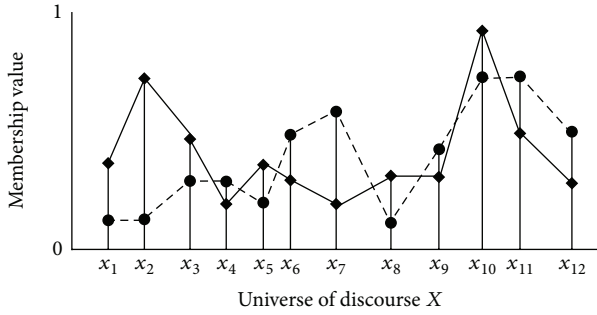$$d((A \cup C), [1]_X) < d((A \cup B), [1]_X) \tag{5}$$

Figure 1: Overlapped data distribution.

guarantees $s(A, C) < s(A, B)$, and

$$d\left((B \cap C), [0]_X\right) = d\left((A \cap C), [0]_X\right),$$
$$d\left((A \cup C), [1]_X\right) < d\left((B \cup C), [1]_X\right) \tag{6}$$

also guarantees $s(A, C) < s(A, B)$; therefore, triangular equality is obvious by the definition, and hence (S4) is also satisfied. □

Besides Theorem 2, numerous similarity measures are possible. Another similarity measure is illustrated in Theorem 3, and its proof is also found in the previous result [15, 16].

Theorem 3. *For any set $A, B \in F(X)$, if d satisfies Hamming distance measure, then*

$$s(A, B) = 1 - d\left((A \cap B), (A \cup B)\right), \tag{7}$$

$$s(A, B) = 1 - d(A, A \cap B) - d(B, A \cap B), \tag{8}$$

$$s(A, B) = 2 - d\left((A \cap B), [1]_X\right) - d\left((A \cup B), [0]_X\right) \tag{9}$$

are the similarity measure between sets A and B.

Proof. Proofs are easy to be derived, and it was found in previous results [15, 16]. □

Besides similarity measures of (7) to (9), other similarity measures are also illustrated in previous results [15–17]. With similarity measure in Theorems 2 and 3, it is only possible to compute the similarity measure for overlapped data (Figure 1). Following data distributions of diamonds (♦) and circles (●) illustrates nonoverlapped data; it is appropriate to design similarity measure for data in Figure 2. Consider the nonoverlapped data distribution of diamonds (♦) and circles (●), the similarity measure of (7) to (9) cannot provide the appropriate solution for the nonoverlapped distribution. Two data pairs that constitute different distributions are considered in Figure 2. Twelve data with six diamonds (♦) and six circles (●) are illustrated with different combination in Figures 2(a) and 2(b). Similarity degree between circles and diamonds must be different between Figures 2(a) and 2(b) because of different distribution. For example, (7) represents

$$s(A, B) = 1 - d\left((A \cap B), (A \cup B)\right). \tag{10}$$

From (7), $d((A \cap B), (A \cup B))$ provides distance between $(A \cap B)$ and $(A \cup B)$. By the following definitions:

$$A \cap B = \min(A, B), \qquad A \cup B = \max(A, B). \tag{11}$$

Nonoverlapped data satisfies $A \cap B = \min(A, B) = 0$, and $A \cup B$ is defined as A or B. Hence, $s(A, B) = 1 - (1/N) \sum \max(A, B)$ shows similarity measure, where N is the total number of data sets A and B. From this property,

$$s(A, B) = 1 - \frac{1}{N} \sum \max(A, B)$$
$$= 1 - \frac{1}{12} \sum (0.5 + 0.8 + 0.6 + 0.4 + 0.5 + 0.6 + 0.7$$
$$+ 0.4 + 0.5 + 1 + 0.8 + 0.6) = 0.38, \tag{12}$$

and the same result is obtained for Figures 2(a) and 2(b). Hence, similarity measures (2) to (9) are not proper for the nonoverlapped data distribution. Two different data in Figure 2(a) are less discriminate than Figure 2(b). It means that similarity measure of Figure 2(a) has a higher value than Figure 2(b). Similar results are also obtained by the calculation of similarity measures (8) and (9).

Hence, it is required to design similarity measure for nonoverlapping data distribution. Consider the following similarity measure for nonoverlapped data such as Figures 2(a) and 2(b).

Theorem 4. *For singletons or discrete data $a, b \in P(X)$, if d satisfied Hamming distance measure, then*

$$s(a, b) = 1 - |s_a - s_b| \tag{13}$$

is a similarity measure between singletons a and b. In (13), s_a and s_b satisfy $d((a \cap \mathbf{R}), [1]_X)$ and $d((b \cap \mathbf{R}), [1]_X)$, respectively. Where $\mathbf{R}$ is whole data distribution including a and b.

Proof. (S1) and (S2) are clear. (S3) is also clear from definition as follows:

$$s(C, C) = 1 - |s_C - s_C|$$
$$= 1 - |d\left((C \cap \mathbf{R}), [1]_X\right) - d\left((C \cap \mathbf{R}), [1]_X\right)| = 1. \tag{14}$$

Finally, (S4) $A, B, C \in F(X)$, if $A < B < C$, then

$$s(A, B) = 1 - |d\left((A \cap \mathbf{R}), [1]_X\right) - d\left((B \cap \mathbf{R}), [1]_X\right)|$$
$$\geq 1 - |d\left((A \cap \mathbf{R}), [1]_X\right) - d\left((C \cap \mathbf{R}), [1]_X\right)| \tag{15}$$
$$= s(A, C),$$

because $d((B \cap \mathbf{R}), [1]_X) > d((C \cap \mathbf{R}), [1]_X)$ is satisfied. Similarly, $s(B, C) \geq s(A, C)$ is also satisfied. □

Similarity measure (13) is also designed with the distance measure such as Hamming distance. As noted before, conventional measures were not proper for nonoverlapping
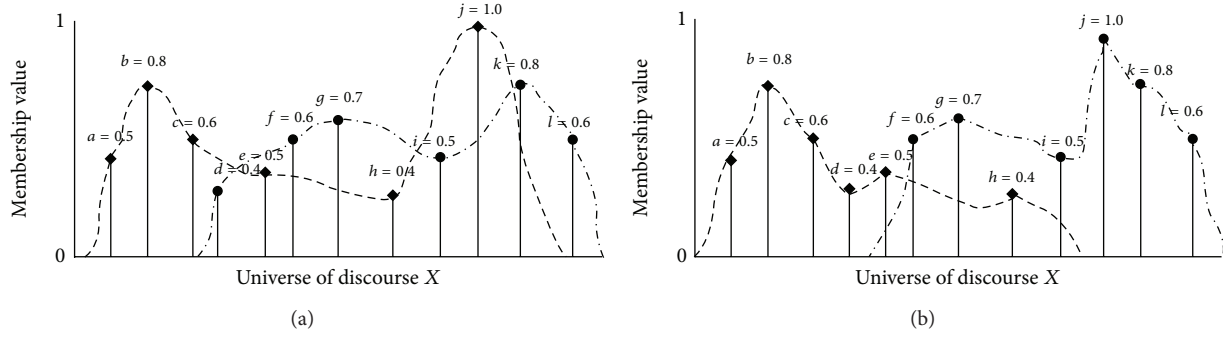
(a)

(b)

Figure 2: (a) Data distribution between circle and diamond. (b) Data distribution between circle and diamond.

continuous data distributions; this property is verified by the similarity measure calculation of Figures 2(a) and 2(b).

Next, calculate the similarity measure between circle and diamond with (13).

For Figure 2(a),

$$
\begin{aligned}
s\left(\blacklozenge,\bullet\right) &= 1 - \left|d\left(\left(\blacklozenge\cap\mathbf{R}\right),[1]_{X}\right) - d\left(\left(\bullet\cap\mathbf{R}\right),[1]_{X}\right)\right| \\
&= 1 - 1\times\left(6\left|\left((1-0.5)+(1-0.8)+(1-0.6)\right.\right.\right. \\
&\qquad\qquad\left.+(1-0.5)+(1-0.4)+(1-1)\right) \\
&\qquad\quad -\left((1-0.4)+(1-0.6)+(1-0.7)\right. \\
&\qquad\qquad\left.\left.\left.+(1-0.5)+(1-0.8)+(1-0.6)\right)\right|\right)^{-1} \\
&= 1 - \frac{1}{6\left|2.3-2.4\right|} = 0.983
\end{aligned}
$$

(16)

is satisfied.

For the calculation of Figure 2(b),

$$
\begin{aligned}
s\left(\blacklozenge,\bullet\right) &= 1 - \left|d\left(\left(\blacklozenge\cap\mathbf{R}\right),[1]_{X}\right) - d\left(\left(\bullet\cap\mathbf{R}\right),[1]_{X}\right)\right| \\
&= 1 - 1\times\left(6\left|\left((1-0.5)+(1-0.8)+(1-0.6)\right.\right.\right. \\
&\qquad\qquad\left.+(1-0.4)+(1-0.5)+(1-0.4)\right) \\
&\qquad\quad -\left((1-0.6)+(1-0.7)+(1-0.5)\right. \\
&\qquad\qquad\left.\left.\left.+(1-1)+(1-0.8)+(1-0.6)\right)\right|\right)^{-1} \\
&= 1 - \frac{1}{6\left|2.8-1.8\right|} = 0.833.
\end{aligned}
$$

(17)

Calculation result shows that the proposed similarity measure is possible to evaluate the degree of similarity for nonoverlapped distributions. By comparison with Figure 2, distribution between diamond and circle in Figure 2(a) shows more similar.

## 3. Human Behavior Signal Analysis and Experiments

Gait signals are collected with experiment unit; acquisition system (Figure 3) system is composed with all in one sensor in


Figure 3: Data acquisition experiment.

which accelerator, magnetic, and Gyro sensor, mobile station, and connector are integrated. Signal acquisition experiment was done as shown in the following figures.

Gait patterns are composed of walking, step up and step down for 20 persons. For each behavior, signals are measured with all in one sensor which integrated with three sensors (accelerator, magnetic, and Gyro sensors); each sensor represents three dimension direct signals. Four sensors are attached to waist, two legs, and head. Example of obtained gait signals is illustrated in the following Figure 4. Among numerous cases, walking and stair up signals are illustrated with acceleration sensor in Figures 4(a) and 4(b), magnetic sensor in Figures 4(c) and 4(d), and Gyro sensor in Figures 4(e) and 4(f), respectively. Full signal was illustrated in Figure 4(f) for stair up with Gyro sensor; we can notice 12 signals for x-y-z direction, and it shows almost the same pattern for similar gait. Hence, x-z direction signals are considered in each figure. Due to the fact that signal patterns are almost the same and numerous quantity, we collect two directional signals. From the top, the first two signals represent z-x signals at head, and next ones are waist and left and right leg signals, respectively. We also carried out preprocessing to make synchronize signals and obtained gait signals that are illustrated in Figure 4.
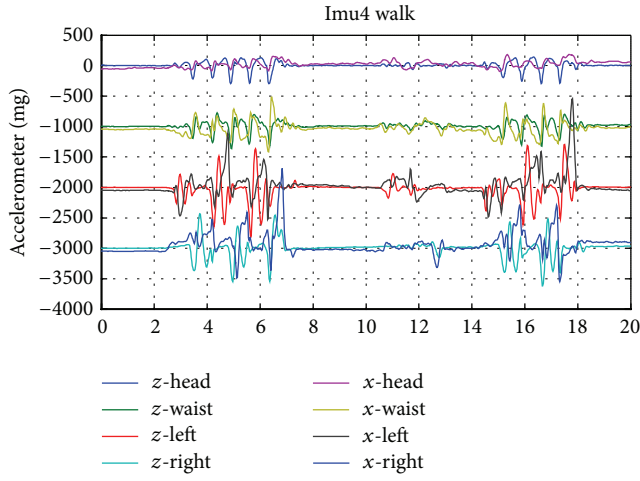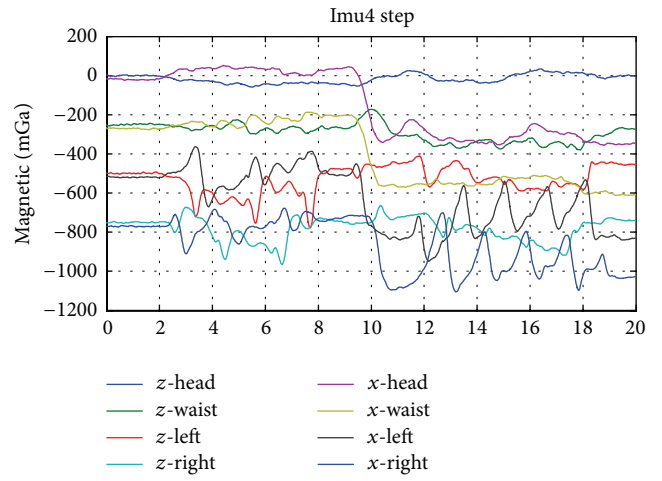
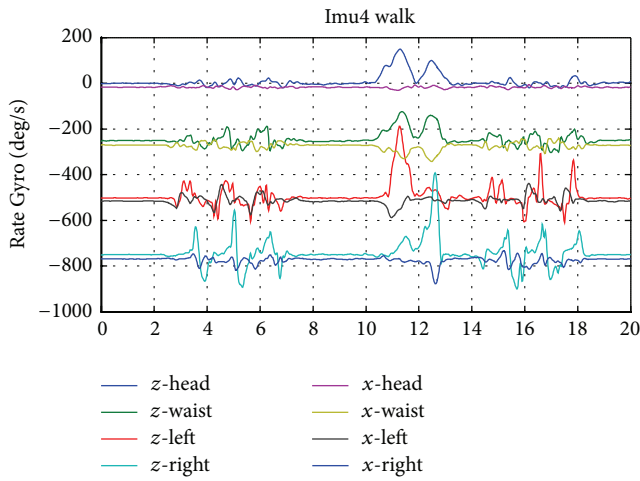(a) Walking with acceleration sensor
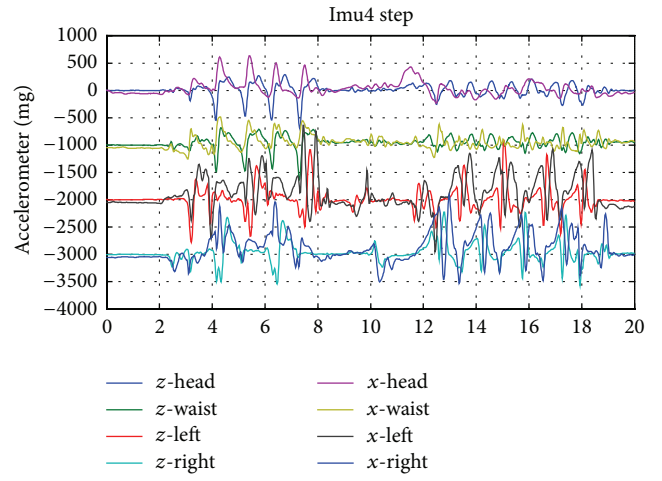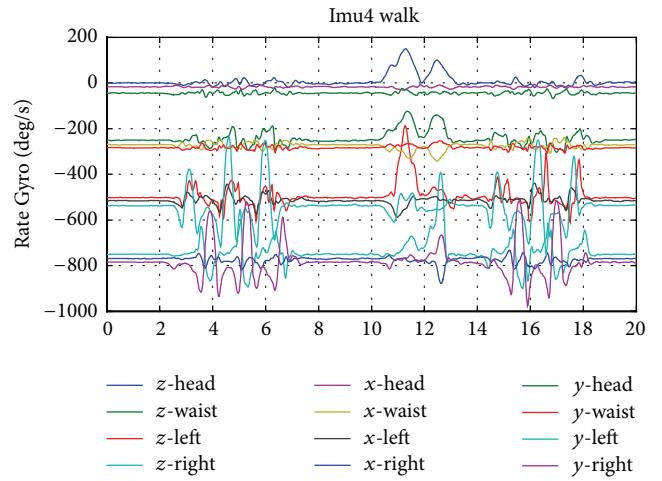
(b) Stair up with acceleration sensor
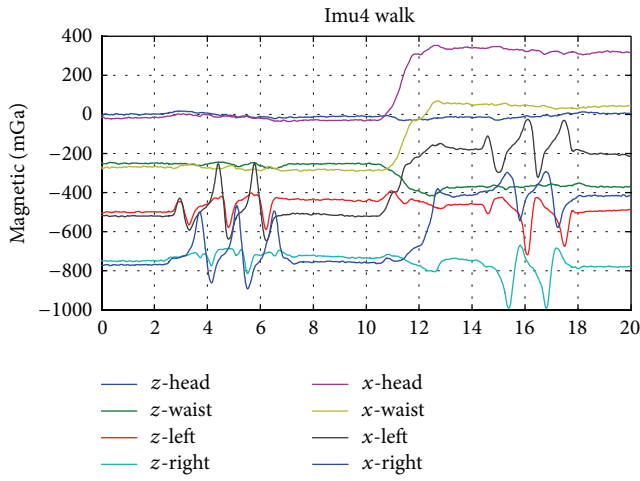
(c) Walking with magnetic sensor

(d) Stair up with magnetic sensor

(e) Walking with Gyro sensor

(f) Stair up with Gyro sensor

FIGURE 4: Gait signal with all in one sensor.

We get the signals from the control unit, and the signal is processed in a note book. Signal characteristics were considered peak value and magnitude distance between each gait signal. Next, by the application of the similarity measure, we get the calculation of each action such as walking, step up, and so on.

4. Numerical Decision Calculation

4.1. High Dimensional Analysis. Research on big data analysis has been emphasized by research outcomes recently [7, 8, 13, 19]. Big data examples are illustrated as follows.

   (i) Biomedical data such as DNA sequence or Electroencephalography (EEG) data. It contains not only high dimension but also large number of channel data.

   (ii) Recommendation systems and target marketing are important applications in the e-commerce area. Sets of customers/clients' information analysis help to predict their action to purchase based on customers' interest. It also includes a huge amount of data and high dimensional structure.

   (iii) Industry application such as EV station scheduling problem needs geometrical information, city size, population, traffic flow, and others. Hence, number of station and station size constitute huge data and high dimension.

Data might be expressed as high dimensional structure such as

$$D_i = \left(d_{i1}, d_{i2}, d_{i3}, \ldots, d_{ij} \right),$$
$$\text{where } i = 1, \ldots, n, \quad j = 1, \ldots, m, \tag{18}$$

where n and m denote the number of data and dimension, respectively.

Direct data comparison is applicable to overlapped data with norm definition including Euclidean norm such as

$$\mathbf{d}_1 \left(D_i, D_j \right) = \sum_{k=1}^{m} \left| d_{ik} - d_{jk} \right|, \quad \mathscr{L}_1 \text{ norm,}$$

$$\mathbf{d}_2 \left(D_i, D_j \right) = \sqrt{\sum_{k=1}^{m} \left(d_{ik} - d_{jk} \right)^2}, \quad \text{Euclidean-norm,} \tag{19}$$

$$\mathbf{d}_p \left(D_i, D_j \right) = \left(\sum_{k=1}^{m} \left| d_{ik} - d_{jk} \right|^p \right)^{1/p}, \quad \mathscr{L}_p \text{ norm}.$$

Information distributions show various configurations, and hence it needs consider various types of distance measure to complete discriminative measure. Furthermore analysis of similarity and relation between different information should be considered carefully when it represents high-dimensional data. Specifically, m dimension represents the independent number of characteristics or attributes.

TABLE 1: Similarity measure comparison between patterns.

Similarity measure	Values
$s\left(W_i, SU_i\right)$	0.344
$s\left(W_i, SD_i\right)$	0.278
$s\left(SU_i, SD_i\right)$	0.550

TABLE 2: Average similarity measure between individuals.

Similarity measure	Values
$s\left(PW_i, PW_j\right)$	0.624
$s\left(PSD_i, PSD_j\right)$	0.643
$s\left(PSU_i, PSU_j\right)$	0.712

4.2. Illustrative Example. Hence, analysis and comparison with each attribute provide explicit importance of each data. Similarity measure provides analysis between patterns, such as

$$s\left(D_i, D_j\right), \quad \forall i, j = 1, \ldots, n. \tag{20}$$

And comparison with different patterns for the same person is carrying out between walking, step up, or step down. Gait signal related with his/her movement was gathered to analyze their different patterns and different persons. Each signal constitutes walking, stair up, and stair down with 20 persons' gaits were gathered. Hence, personal information can be represented by multidimensional information such as

$$P_i = \left(W_i, SU_i, SD_i\right), \quad \text{where } i = 1, \ldots, 20. \tag{21}$$

And among 20 personal data, 3 different behaviors were also expressed. Considering the data, it is obvious that data is overlapped. Hence, it is clear that similarity measures of (2) and (7)–(9) provide similarity calculation for overlapped data.

Similarity measure between person to person is expressed as

$$s\left(P_i, P_j\right) = 1 - d\left(\left(P_i \cap P_j\right), \left(P_i \cup P_j\right)\right). \tag{22}$$

Also different action from the same person as follows:

$$s\left(W_i, SU_i\right) = 1 - d\left(\left(W_i \cap SU_i\right), \left(W_i \cup SU_i\right)\right). \tag{23}$$

Normalized similarity calculation results are illustrated in Table 1.

Results illustrate that the stair up and down shows higher similarity than the others. However, even similarity calculation result is higher than others; it is not much close to one, it just satisfies 0.55. Due to different directions stair up/down should have basic limit to close maximum similarity.

Table 2 shows the average similarity between different individuals. Results show that the stair up is the closest even with a different gait. Naturally, walking pattern represents the least similar. In Table 3, walking similarity between different

TABLE 3: Similarity measure comparison between individuals (walking).

Similarity measure	1	2	3	4	5	6	7		⋯		18	19	20
1	1	0.511	0.621	0.420	0.462	0.581	0.617				0.581	0.470	0.623
2	0.511	1	0.592	0.590	0.490	0.632	0.543				0.603	0.619	0.577
3	0.621	0.592	1	0.510	0.611	0.640	0.599				0.710	0.566	0.582
4	0.420	0.590	0.510	1	0.701	0.653	0.589		⋯		0.612	0.629	0.585
5	0.462	0.490	0.611	0.701	1	0.599	0.603				0.489	0.581	0.620
6	0.581	0.632	0.640	0.653	0.599	1	0.576				0.499	0.545	0.629
7	0.617	0.543	0.599	0.589	0.603	0.576	1				0.710	0.627	0.634
											0.590	0.611	0.589
⋮			⋮						⋮		0.429	0.570	0.630
											0.559	0.628	0.549
18	0.581	0.603	0.710	0.612	0.489	0.499	0.710	0.590	0.429	0.559	1	0.345	0.626
19	0.470	0.619	0.566	0.629	0.581	0.545	0.627	0.611	0.570	0.628	0.345	1	0.556
20	0.623	0.577	0.582	0.585	0.620	0.629	0.634	0.589	0.630	0.549	0.626	0.556	1

individuals is illustrated; symmetric results of 20 persons are listed in Table 3.

Similar results for stair up and down are obtained.

5. Conclusions

Gait signal identification was carried out through similarity measure design. Gait signal was obtained via data acquisition system including mobile station, all in one sensor attached to the head, waist, and two legs. In order to discriminate the gait signal with respect to different behaviors and individuals, similarity measure design was considered. Similarity measure was considered with the distance measure. For data distribution, overlapped and nonoverlapped distribution were considered, and similarity measure was applied to calculate the similarity. However, the conventional similarity measure was shown that it was not available to calculate the similarity on non-overlapped data. To overcome such a drawback, the similarity measure was considered with data information of neighbor. Closeness between neighbor data provides a measure of similarity among data sets; hence, the similarity measure was calculated. Calculation proposed two different artificial data, and the proposed similarity measure was useful to identify nonoverlapped data distribution. It is meaningful that similarity measure design can be extended to high dimensional data processing because gait signal was considered as a high dimensional data. With data acquisition system, 20 person gait signals were collected through experiments. Different gaits, walking, stair up, and stair down signals were obtained, and similarity measure was applied. By calculation, similarity between stair up and stair down showed higher similarity than others. Individual similarity for a different gait signal was also obtained.

Gait signal analysis can be used for behavior decision system development; it is also naturally extended to health care system, especially to elderly people. Additionally, it is also useful for athlete to provide useful information if he/she is suffering from different actions compared to previous behavior.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] D. H. Fisher, "Knowledge acquisition via incremental conceptual clustering," *Machine Learning*, vol. 2, no. 2, pp. 139–172, 1987.

[2] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

[3] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *Computer Journal*, vol. 26, no. 4, pp. 354–359, 1983.

[4] R. S. Michalski and R. E. Stepp, "Learning from observation: conceptual clustering," in *Machine Learning: An Artificial Intelligence Approach*, pp. 331–363, Springer, Berlin, Germany, 1983.

[5] H. P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *Journal of the American Statistical Association*, vol. 62, no. 320, pp. 1159–1178, 1967.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, NY, USA, 1990.

[7] "Advancing Discovery in Science and Engineering. Computing Community Consortium," Spring 2011.

[8] Advancing Personalized Education, *Computing Community Consortium*, 2011.

[9] S. Lee and T. O. Ting, "The evaluation of data uncertainty and entropy analysis for multiple events," in *Advances in Swarm Intelligence*, pp. 175–182, Springer, New York, NY, USA, 2012.

[10] S. Park, S. Lee, S. Lee, and T. O. Ting, "Design similarity measure and application to fault detection of lateral directional mode flight system," in *Advances in Swarm Intelligence*, vol. 7332 of *Lecture Notes in Computer Science*, pp. 183–191, Springer, New York, NY, USA, 2012.

[11] S. Lee and T. O. Ting, "Uncertainty evaluation via fuzzy entropy for multiple facts," *International Journal of Electronic Commerce*, vol. 4, no. 2, pp. 345–354, 2013.

[12] S. Lee, W. He, and T. O. Ting, "Study on similarity measure for overlapped and non-overlapped data," in *Proceedings of the International Conference on Information Science and Technology (ICIST '13)*, pp. 48–53, March 2013.

[13] "Smart Health and Wellbeing," Computing Community Consortium, 2011.

[14] X. C. Liu, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy Sets and Systems*, vol. 52, no. 3, pp. 305–318, 1992.

[15] S. Lee, W. Pedrycz, and G. Sohn, "Design of similarity and dissimilarity measures for fuzzy sets on the basis of distance measure," *International Journal of Fuzzy Systems*, vol. 11, no. 2, pp. 67–72, 2009.

[16] S. Lee, K. H. Ryu, and G. Sohn, "Study on entropy and similarity measure for fuzzy set," *IEICE Transactions on Information and Systems*, vol. 92, no. 9, pp. 1783–1786, 2009.

[17] S. H. Lee, S. J. Kim, and N. Y. Jang, "Design of fuzzy entropy for non convex membership function," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Contemporary Intelligent Computing Techniques*, vol. 15 of *Communications in Computer and Information Science*, pp. 55–60, 2008.

[18] Y. Cheng and G. Church, "Biclustering of expression data," in *Proceedings of the 8th International Conference on Intelligent System for Molecular Biology*, 2000.

[19] D. M. West, *Big Data for Education: Data Mining, Data Analytics, and Web Dashboards*, Governance Studies at Brookings, Washington, DC, USA, 2012.