2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

The footstep classification using continuous DP matching and threshold

Hiroshi Yasukawa† and Akitoshi Itai‡

†Aichi Prefectural University
1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, JAPAN
‡Chubu University
1200 Matsumoto-cho, Kasugai-shi, Aichi, 487-8501, Japan
Email:†yasukawa@ist.aichi-pu.ac.jp, ‡itai@cs.chubu.ac.jp

Abstract—Human footsteps include various characteristics determined by the gait, the footwear and the floor. In these years, the personal identification using footstep has been investigated. The footstep identification method based on continuous dynamic programing (DP) matching and short term Fourier transform (STFT) is proposed[1]. The conventional method shows the recognition rate of 95.1% for 10 subjects, however, all the footstep is classified as the registered footstep. This means that the non-registered subject, whose footsteps not used as the reference data, is recognized as one of the reference data.

In this paper, we propose the footstep classification using a continuous DP matching and threshold to reject a non-registered pedestrian.

1. Introduction

Characteristics of human footsteps are determined by the gait, the footwear and the floor environment. If personal characteristics and a walking state can be extracted from footsteps correctly, footsteps will lead to the creation of more powerful security systems such as enhanced video surveillance systems and biometric systems that will combine footstep analysis with fingerprinting and vein recognition. Moreover, by analyzing the difference of right-and-left footsteps and walking manner, it may be possible to create remote medical systems for continuous health checking.

In the field of indoor environmental sound analysis, some useful results are reported for a footstep based surveillance system. Tanaka[2] and Bin[3] presented footstep detection from various environmental sounds. However these researches did not focus on a personal identification. Shoji[4][5] proposed a feature extraction technique that uses mel-cepstrum analysis, the footstep intervals, and the degree of similarity of the spectrum envelope for footstep classification. Itai[6],[7] used psychoacoustics parameters and wavelet to represent a feature of footstep, respectively. These techniques requires the starting time of footstep to extract the feature parameter. DP matching[8][9] is basic technique in speech recognition to normalize the raw data in terms of time, moreover it can absorb fluctuation in data duration. Furthermore, continuous DP matching, which is the extended method of DP matching, searches a similar pattern to the reference pattern from the input pattern. Therefore, the footstep identification using continuous DP matching does not need the starting time of footstep. Nomura[1] proposed a footstep identification scheme based on DP matching and STFT. Miyoshi[10] reported the footstep classification with Gaussian mixture models. However, the conventional approach can not reject the others whose footsteps are not adopted as reference data. In this paper, we propose the footstep classification using the DP matching and thresholding.

Table 1: Recording environment

Location	A passage in Aichi Prefectural University
Subject	15 students
Footwear	A pair of slippers
Number of trials	Ten times per subject
Sampling frequency	44100Hz

2. Recording Condition

The measurement conditions used throughout this paper are described here. We measure footstep amplitude values at the passage in an indoor environment. The recording conditions for the 15 subjects are shown in Table 1. Examinees walk along the indoor passage in Aichi prefectural university for ten times wearing the specified slippers. Footsteps are captured by a microphone (ONO SOKKI LA-5120, frequency range: 20-12500Hz) and collected by a computer for sampling at 44.1 kHz. The recording environment is shown in Fig. 1. Subjects walk along a 5m track towards the microphone; this distance is chosen to yield natural footsteps and to avoid tiring the subject. In this paper, 15 subjects perform only 10 measurement sessions. Examples of footstep waveform thus recorded over 3 seconds are shown in Fig. 2. The vertical axis shows an amplitude, the horizontal axis indicates a time (seconds). The footstep includes large spikes at 0.5 second interval. Those spikes indicate an impact sound of the footwear against a

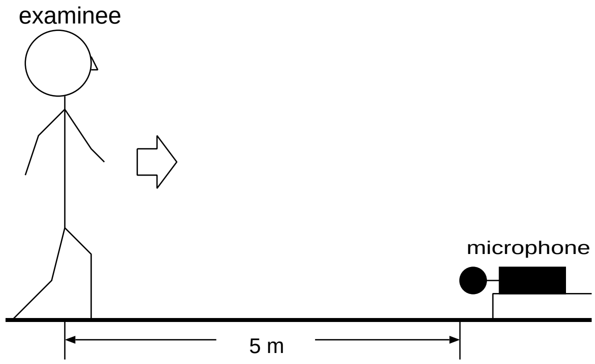

Figure 1: Recording environment

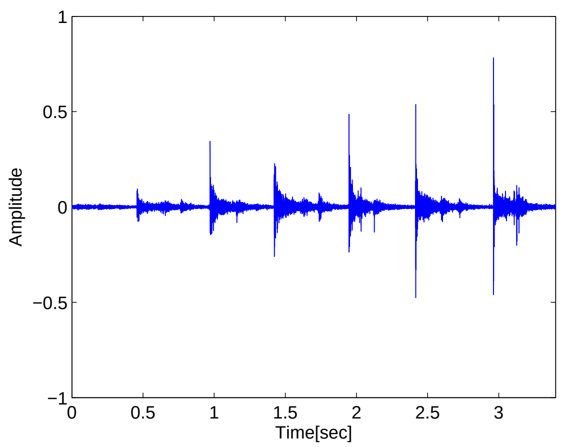

Figure 2: Example of footstep waveform

floor. From the conventional research, it is known that the fricative sound between a floor and footwear is included after the impact sound. In addition, some footsteps form a non-stationary waveform with a spike and high frequency signals.

3. Recognition Method

3.1. Method for Speech Recognition

In a speech recognition, if two people speak the same word, the duration of their phonemes will differ. DP matching can expand or contract the time axis (dynamic time warping), and absorb the temporal fluctuations in the raw data. A simple recognition technique based on DP matching calculates the correlation of spectrum patterns of the input and the reference pattern. This method focuses on the time change in the frequency components in the speech signal. We note that the impact and fricative sounds in footsteps exhibit nonlinear changes in time domain. In this paper, we employ a cosine distance and a continuous DP matching to calculate the similarity of footstep spectrograms.

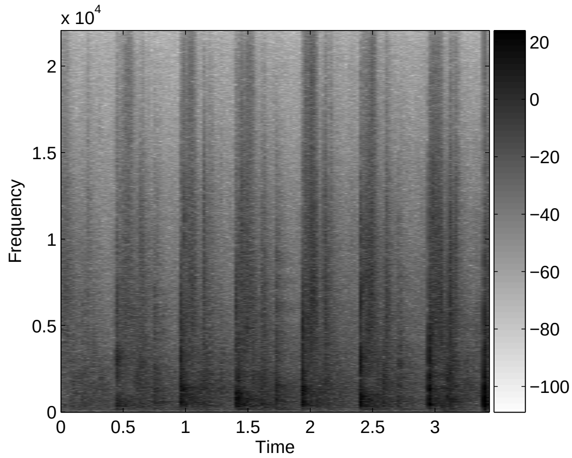

Figure 3: Example of footstep spectrogram

3.2. Similarity of Spectrograms

The spectrogram of the footstep data is generated by STFT. As shown in (1), the spectrogram $\boldsymbol{A}$ is represented as a two dimensional matrix.

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,i} & \cdots & a_{1,I} \\ \vdots & \ddots & \vdots & & \vdots \\ a_{f,1} & & a_{f,i} & & a_{f,I} \\ \vdots & & \vdots & \ddots & \vdots \\ a_{F,1} & \cdots & a_{F,i} & \cdots & a_{F,I} \end{bmatrix}. \tag{1}$$

The columns indicate the time index i, while the rows represent the frequency index f. An example of footstep spectrogram is shown in Fig.3.

In Fig.3, the vertical axis represents a frequency, the horizontal axis is a time (seconds). A spectrum similarity of each frame is calculated using the cosine distance. i-th frame spectrum of input data $\boldsymbol{A_i}$ and j-th frame spectrum of reference data $\boldsymbol{B_j}$ are given by:

$$\boldsymbol{A}_i = \begin{bmatrix} a_1 & a_2 & \cdots & a_f & \cdots & a_F \end{bmatrix}, \tag{2}$$

$$\boldsymbol{B}_j = \begin{bmatrix} b_1 & b_2 & \cdots & b_f & \cdots & b_F \end{bmatrix}. \tag{3}$$

$$(1 \leq i \leq I, 1 \leq j \leq J)$$

The cosine distance between frequency components $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ is given as follows:

$$\boldsymbol{D}(i,j) = \frac{|\boldsymbol{A}_i| \cdot |\boldsymbol{B}_j|}{\sqrt{\sum_{f=1}^{F} a_f^2}\sqrt{\sum_{f=1}^{F} b_f^2}} \tag{4}$$

where $|\boldsymbol{A}_i|$ is an array such that each element is the absolute value of the corresponding element of $\boldsymbol{A}_i$; $\cdot$ indicates inner product. If (4) is calculated for all combinations of (i, j), $\boldsymbol{D}$ is expressed by a two-dimensional array which is the similarities between the frequency components of (i, j). The cosine distance represents correlation using a scalar value

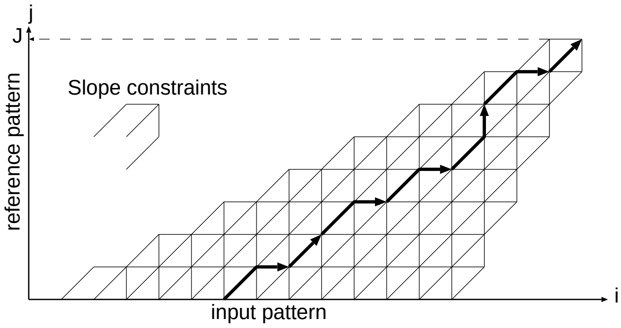

Figure 4: Continuous DP matching

from -1 to 1. When the cosine distance is close to 1, there is a strong correlation between the two frequency components.

3.3. Continuous DP Matching

In the conventional method, the starting time of footstep is required to calculate the feature parameters. However, a continuous DP matching does not use the starting time in the recognition process. This paper employs a continuous DP matching to calculate the similarity between two walking footsteps. The cumulative distance between input and reference data is calculated to evaluate the similarity.

Consider the 2D plane shown in Fig. 4. The discrete points in Fig. 4 are indexed by the time series. All grid points represent a distance between $\mathbf{A}(i)$ and $\mathbf{B}(j)$. The basis problem of DP matching is to find the shortest path from node (1, 1) to (I, J). In continuous DP matching, the partial interval of the input pattern is matched to the reference pattern with shifting one frame. Assuming that the distance between ith frame in the input pattern and jth frame in the reference pattern is $d(i, j)$, the cumulative distance $g(i, j)$ is defined as following equation:

$$g(i, j) = \min \begin{bmatrix} g(i-2, j-1) + 2d(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-1, j-2) + 2d(i, j-1) + d(i, j) \end{bmatrix}. \quad (5)$$

The cosine distance $\mathbf{D}$ is calculated by using a continuous DP matching. However, $1 - \mathbf{D}$ is applied to each grid in order to calculate the minimum cumulative distance using DP matching since DP matching employs the minimization problem. The MATLAB source developed by Dan Ellis(Columbia University)[8] is used to calculate cosine distance and realize a continuous DP matching.

4. Proposed Method

4.1. Footstep Classification

The personal footstep identification using continuous DP matching and threshold is introduced here.

Table 2: Classification parameters

Window length of STFT	46.4 msec
Shift width of window	23.2 msec
Reference data	1 of 10 walking footstep for 10 subjects
Length of reference data	1.6 sec

The similarity of input and reference spectrograms is calculated by the above method. The Table 2 lists the parameters for feature extraction and classification. The length of clipped reference data is 1.6 seconds, the length of STFT window is 46.4 msec (2048 samples), the shift width of STFT window is a half of STFT window. These parameters yield the better detection ratio in [1]. The reference data clipped from recorded footstep data includes the footstep of three steps. On the other hand, input data is the raw recorded data during 5m track. To evaluate the rejection performance, footsteps of 10 subjects are adopted as reference data (registered subjects). Other 5 subjects are the non-registered subjects. The one of 10 footstep data for registered subjects is clipped and used as reference data. This means that the footstep of 5 non-registered subjects are not classified to any reference data. The recognition process consists of two steps:

The candidate detection

The cumulative distances between one input data and all reference data are calculated using continuous DP matching. The input data is classified as the reference data which yields the minimum cumulative distance. This classification result is used as the candidate.

The rejection using threshold

If the the cumulative distance of the candidate is greater than threshold T, the input footstep is rejected. This means the input data is recognized as the non-registered subject. In the case of the cumulative distance is less than T, the input data is classified as the subject of candidate footstep. In this approach, the small T yields the excessive rejection for registered footsteps.

4.2. Recognition Result

The performance of proposed method is evaluated by using the recognition and rejection ratio. The recognition ratio is defined as the ratio between "the number of correctly classified footsteps for registered subjects" and "the number of footsteps for registered subjects". The rejection ratio is the ratio between "the number of correctly rejected footsteps for non-registered subjects" and "the number of footsteps for non-registered subjects". Note that the recogni-

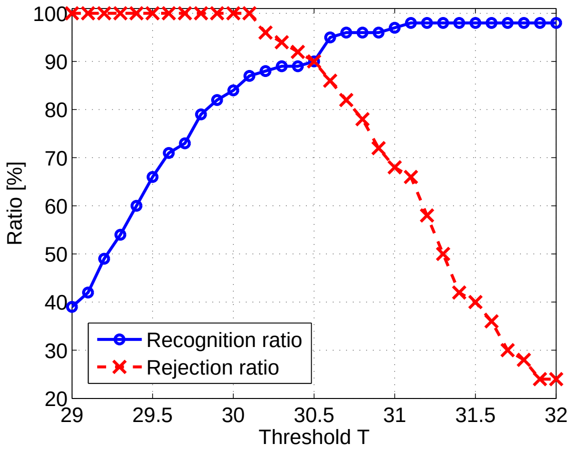

Figure 5: Recognition and rejection ratio

tion and rejection ratio indicate the detection and rejection accuracy for registered and non-registered subjects, respectively. These ratios are calculated using the leave-one-out approach. Recognition results are shown in Fig.5. The vertical axis represents the ratio, the horizontal axis indicates the threshold T for cumulative distance. The solid and broken line show the recognition and rejection ratio, respectively. T is a scalar value from 29 to 32 at 0.1 interval.

From Fig.5, the recognition rate increases with threshold. On the other hand, the rejection ratio is decreased with increasing the threshold. When we focus on the ratios in $T = 30.5$, the recognition and rejection ratio is 90%. In addition, $T = 30.1$ performed the perfectly rejection while the recognition ratio is 87%. These results indicate that the threshold approach successfully rejects the non-registered subject.

Nomura shows that a continuous DP matching and a short term Fourier transform yields the accuracy of 95.1% by using the footstep of 1.6 second[1]. In the case of $T = 30.1$, the recognition ratio of our result is 8% less than the conventional research. However, the proposed method achieved the rejection of non-registered subjects whose footsteps are not prepared as reference data.

5. Conclusion

This paper proposed the footstep classification scheme that uses continuous DP matching and thresholding. The similarity of input and reference footstep is calculated by using a cosine distance of footstep spectra. Experimental results show that the recognition and rejection ratio for 15 subjects exceeds 90% when the threshold T is 30.5. The completely rejection of non-registered subjects is achieved by $T = 30.1$. This result shows that the continuous DP matching and thresholding approach is effective to the footstep classification. In addition, non-registered subjects whose footstep is not used for reference data is correctly rejected by using threshold. The proposed method also indicates the possibility that the environmental abrupt sound, i.e. claps and knocking, can be rejected without the preprocessing and classification. Future tasks include examining various walking speeds and realizing recognition system.

References

[1] Y.Nomura, T.Isomura, and H.Yasukawa, “Personal Footstep Identification Using DP Matching,” Proc. of NCSP2008 , pp.295-298, 2008.

[2] M.Tanaka, and H.Inoue, “A Study on Walk Recognition by Frequency Analysis of Footsteps,” (in Japanese), Trans. IEE of Japan, Vol.119-C, No.6, pp.762-763, 1999.

[3] B.She, “Framework of footstep detection in indoor environment,” Proc. of ICA2004, Mo.P2, Engineering acoustics (I)(EAP01), pp.715-718, 2004.

[4] Y.Shoji, T.Takasuka and H.Yasukawa, “Personal Identification Using Footstep Detection,” Proc. of ISPACS2004, pp.43-47, 2004.

[5] Y.Shoji, A.Itai and H.Yasukawa, “Personal Identification Using Footstep Detection in Indoor Environment,” IEICE Trans. Fund., Vol.E89-A, NO.8, pp.2072-2077, 2005.

[6] A.Itai and H.Yasukawa, “Footstep Recognition with Psycho-Acoustics Parameter,” Proc. of 2006 IEEE Asia Pacific Conference on Circuits and Systems, pp.994-997, 2006.

[7] A.Itai and H.Yasukawa, “Personal Identification Using Footstep Based on Wavelets,” Proc. of ISPACS2006, pp383-386, 2006.

[8] D.Ellis, “Dynamic Time Warp in Matlab”, http://labrosa.ee.columbia.edu/matlab/dtw/

[9] J.Griffin, “Discrete-Time Processing Of Speech Signals,” Wiley-IEEE Press, 1993.

[10] M. Miyoshi, K.Mori, Y,Kashihara, M.Nakao, S.Tsuge, M.Fukumi, “Personal identification method using footsteps,” Proc. of SICE Annual Conference, pp.1615-1620, 2011.