People Recognition through Footstep Sound Using MFCC Extraction Method of Artificial Neural Network Back Propagation

Jan Everhard Riwurohi[1,2], Jazi Eko Istiyanto[1], Khabib Mustofa[1], Agfianto Eko Putra[1]
[1]Department of Computer Science and Electronic, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Department of Computer Systems, Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia

Summary
The sound of footsteps is a sound that is heard when people are walking, where the sound is the result of a jerk between the foot and the floor surface. This sound may vary from individual to individual due to some differences, including how to step, footwear used, the characteristics of the floor, height, and weight of the person. The sound of this footstep is a biometric feature and can be used as a person's identifier when the person is stepping. In some previous studies, the process of recording the footsteps on the data acquisition media is done on one trackside only, which is on the middle side of the track, as they used only one microphone. It means that the footsteps sound used to recognize a person was only the closest footsteps sound from the microphone. In this study, the sound of footsteps is recorded at the time when people are walking not only on one side of the course but three sides of different paths - the left side of the track, the middle side of the track and on the right side of the track. Thus, the footsteps sound used to recognize a person all the footsteps sounds recorded from all of those sides of the track. This way is aimed to the process of recognition of people based on the sound of footsteps on a data acquisition medium can run as natural as possible so that wherever the person stepped can be recognized well based on the sound of footsteps. This study created a media to support data acquisition process and to support the process of footstep recording and to put the microphone. The study used four microphones which are placed on the right and the left side of data acquisition track. The method of feature extraction used to extract the footstep sound is Mel Frequency Cepstrum Coefficient (MFCC), and the classification method for the recognition process of the person is using Back Propagation Neural Network. The accuracy of person recognition based on footsteps obtained in this study is 98.8% for the recognition of people walking on the left side of the track, 98.8% on the middle side of the track and 95% for the recognition of the person running on the right side of the track.

Keywords:
Footsteps sound, data acquisition media, microphone, MFCC, Artificial Neural Network Back Propagation.

1. Introduction

Biometrics is a process of learning about automatic methods to recognize humans based on one or more parts of the human body or human behavior itself, which has a certain uniqueness. In an application in the world of information and technology, biometrics is relevant to the technology which is used to analyze physical and human behavior in the authentication process [1]. There are many biometric properties such as iris [2], fingerprint [3], palm print [4] and sound [5] which have been used in the implementation of people recognition. The characteristic of biometrics can be divided into two categories; they are physiological characteristics (such as irises, fingerprints, and palms) and behavioral characteristics (such as sound, footsteps sound and gait).

The recognition of people based on the characteristics of footsteps, in general, can be divided into two parts. Those are such as the recognition of footsteps based on the large or small foot pressure on the floor and recognizing footsteps based on the size or magnitude of the sound heard as a result of foot beat [6]. In recognition of footsteps based on the size or magnitude of foot pressure on the floor, the sensor which is used should be placed on the floor to be passed by the person stepping so that the floor must undergo a physical change. In recognition of footsteps based on the large or small sounds that sounded when the person stepped then the sensor is only placed on the surface of the floor and the floor does not change physically.

The main benefit of the footstep is included in part of biometrics which is the footsteps can be recognized unrealizable by the person involved, and therefore the sensing system tends not change the behavior of the person concerned. The result gained in this study of footsteps determines and illustrates the potential footsteps as part of biometrics in the foreseeable future. [7]. Implementation of the people recognition based on the sound of this footstep can be applied in some ways, and one of them is applied to the home security system [8]. According to [8], the ability to identify occupants is one of the most important things regarding home security, which is part of the concept of intelligent home or implementation of the smart home.

The cycle of people walking is a cycle that describes the situation when footsteps on the ground and ends when the same leg steps on the ground again. This cycle is divided

Manuscript received April 5, 2018
Manuscript revised April 20, 2018

into two periods, namely stance and swing [9]. The cycle of people walking can be seen in Fig 1.

The stance period is used to describe the period when the foot touches the floor while the swinging period describes when the foot is in the air, in other words, there is no contact with the floor. The duration of the running cycle may vary depending on the speed of the person walking. The current cycle period comprises 60% for the stance period and 40% for the swing period [9]. In this period of support stance, there will appear the sound of a foot jerk on the floor. This sound will vary from one person to another. The difference is due to weight, the way the person is stepping and the characteristics of the existing floor. The sounds that appear and the duration of the sound will be a biometric feature that distinguishes between one people to another.

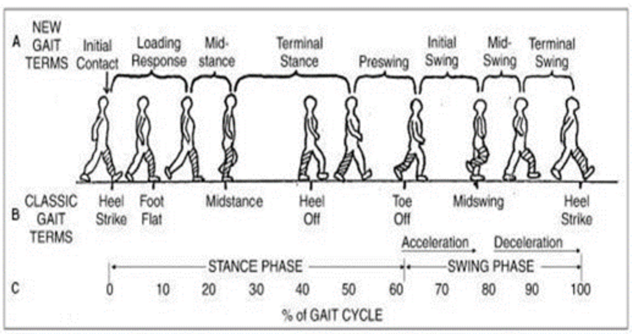

Fig. 1 The cycle of people walking [9]

In this research, the people recognition based on the sound of footsteps that sounded is only focused on the footsteps of one person walking. Footwear used by the person concerned is footwear which is made of wood or other of the same type so that when the person is walking will cause a sound that sounds like footsteps.

The sound occurs when stepping should be heard well, this is because the sensor which is used is a microphone. The room used in this study is a closed room so that other sounds that can interfere with the sound of footsteps can be eliminated as much as possible and to support the foot sound recording process, then this study made a media that serves as a data acquisition system. Data acquisition system here is a place for people stepping equipped with the layout of the arrangement of microphone sensors used.

2. Related Work

In contrast to the identification of other behavioral traits such as sound, iris and palm print, there are only a few published research papers on the identification of people based on footsteps that focus on the sound of footsteps. The study using a microphone to read footstep signal was first performed by Tanaka [10]. In this study, Tanaka used a microphone which is placed at the end of the walking path. The extraction method used in this research is FFT (Fast Fourier Transform), and the method for re-recognition used is Euclidean Distance. The number of people involved in this study was nine people, and the accuracy that is obtained from this study was 83%.

In 2004, Shoji [11] proposed a method of footstep identification by using Mel frequency analysis. The clustering approach k is used for classification. The research done by Shoji is a follow-up study that has been done before by Tanaka, Shoji also used one microphone that was placed at the end of the track position microphone. It is in the middle of the track width where the length of the track is 500 cm. The number of people involved in this study is five people, the footwear used is sabo, and each person will run at ten times. The extraction method which is used in this Shoji research is MLSA (Mel Log Spectrum Analysis) and the identification method used is K Means method. Accuracy result obtained in this Shoji research is 84%.

The research using a microphone as a sensor for reading other footsteps and it is done by She [12]. In this study, She focused on reading footsteps and other sounds in a house so that later can be distinguished between the sound of footsteps and other sounds. In this research, the extraction method that is used is EPD (End Point Detection), Footstep Impact Extraction and Footstep Rhythm Extraction. This study also reads footsteps for indoor and outdoor. The result of this research is 84,1%. Besides that, this research also arranges table containing other sounds that can disturb the footsteps, so that the result of this research can be used for the application to read the sound of footsteps between the sounds other that might happen.

The study of footsteps as a person identification was also done by Itai [13]. In his research, Itai used a data acquisition trajectory to record the sound of footsteps as used by Shoji [11]. In his research, Itai considered the sound of footsteps as a sound of a voice speaking while for the extraction process Itai used MFCC (Mel Frequency Cepstrum Coefficient and for the recognition process, and Itai also used DTW (Dynamic Time Wrapping). The weakness of this research is that the footstep sounds which are extracted only for the closest footstep from the microphone. The result of this footstep sound recognition in this research is 96%.

In 2017, Guo [14] proposed a footstep identification system for the smart home that used acoustic and acoustic psycho parameters as a feature of the nearest class footstep and for classification. The data acquisition system which was used is a development of Shoji [11], It used a microphone sensor with a track length of 5000 cm and a sensor that was placed at the center of the end of the track. The number of people involved in this study is ten people with different footwear. Everyone will step ten times,

where the scenario stepping from each person is at the end of the track towards the microphone. The result of accuracy obtained in this Guo research as a whole is 81%.
In this study, the microphones used to recgeord footsteps are not placed on the end of the track but the right and left side of data acquisition media. The way or scenario of the person stepping is not only on the middle side as found in the previous research but also on the left side of the track and the right side of the track so that everyone will step on three sides of the track. It is intended to be able to know exactly the sound patterns of each person's footsteps in each area along the path of the data acquisition system so that hopefully it can be obtained better accuracy than previous research.

3. Research Method

The research model which is proposed in this study can be seen in Fig 2.

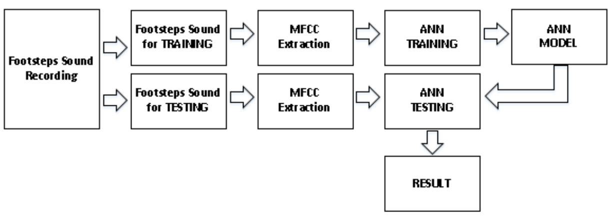

Fig.2 The proposed model research

3.1 Footstep Recording

Recording the footsteps of people which is stepping is done directly through the microphones. It is available on existing data acquisition media.

3.2 Data Separation

The result of footstep sound record will be separated into two; they are Training Data and Testing Data. This separation is done as part of the effort to form a Model of Artificial Neural Network (ANN) which will be used in the identification process. Formation of Neural Network Model is using Training Data, and Neural Network Testing is using Testing Data.

3.3 Extraction of MFCC (Mel Frequency Cepstrum Coefficient)

This MFCC extraction phase aims to change the footnote recording data from the time domain into the frequency domain to get the main characteristics of footsteps for each person. This characteristic is used as a feature that distinguishes between one individual to another. The MFCC extraction process for Training Data and Test Data is done separately.

3.4 Classification

Classification is intended to determine the suitability of the input value and output value. The input at the classification is in the form of the sound of footsteps of the person stepping while the output is the identity of the person concerned. The classification in this study is using ANN with Back Propagation method. The ANN used to obtain a high accuracy value during the identification process.
Characteristic of ANN used is designed by making a model of ANN. After ANN model is formed, then it is performed the training phase using Training Data. At the training phase, it can be seen whether the stability of ANN has been achieved or not. If ANN is not stable, the architecture of the ANN will be change by updating the structure and number of hidden layers. If the stability is achieved in the ANN model, then the next phase is testing phase using Data Test.

3.5 System Testing

Testing of the classification model of footstep sound feature performed in this study includes performance appraisal. Performance appraisal (performance assessment) is to test the ability of the model classification of characteristic sound patterns in classifying data that has errors in the classification. The formula that is used is usually the formula for calculating accuracy, recall, precision, and error rate based on Web formula [15].

4. Data Acquisition System

Media data acquisition system which is used to record the sound footstep in this study can be seen in Fig 3.

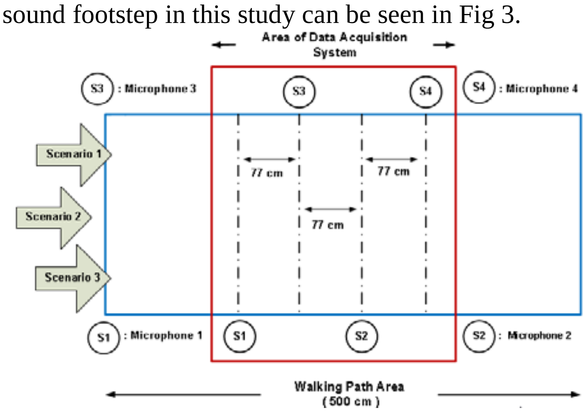

Fig. 3 Data Acquisition System for footstep sound recording.

In the research that was done by Tanaka [10], Shoji [11] and Itai [13] used only one microphone. Meanwhile, this study used four microphones. The use of more microphones aims to make the recorded footsteps more detailed because the sound of footsteps recorded not only comes from a microphone but also comes from 4 microphones.

In the research which was examined by Tanaka [10], Shoji [11] and Itai [13], the processed of footsteps were the sound of footsteps which were closest to the microphone, whereas in this study the footsteps read and processed were footsteps recorded along the path of the data acquisition system. Microphones are placed in crossed positions on the data acquisition system which is intended for the sound of pounding from the right foot, or the sound of pounding from the left foot can be recorded properly. It is considered that at the time people are walking there who tend to rest on the right foot but there is also a tendency to rest on the left foot [16]

The distance between the microphone sensors in this study was set to a value of 77 cm [17]. Nurcahyo stated that the step length (the distance between the right and left heel or vice versa on the move) for the Indonesians ranged from 76.5 cm up to 77 cm with the assumption that the height of observed Indonesian is ranged between 160 cm - 180 cm. The overall dimension used for the data acquisition system is the 500 cm walking run length, where this value refers to Tanaka's [10], Shoji [11] and Itai [13] studies with a track width of 150 cm.

5. The process of footstep sound recording

The process of recording the sound of footsteps on the research is done directly on the media data acquisition system Fig 3. This study involved ten subjects consisting of five men and five women. The recording process uses sampling frequency 44.100 Hz, 16-bit mono.

The recording mechanism is everyone should step in one direction according to the direction of the arrows in Fig 3. The direction of stepping in the recording process of each subject follows three ways or scenarios. The first scenario is everyone will step on the left side of the data acquisition media path; the second is everyone will step on the middle side of the data acquisition media path, and the third is everyone will step on the right side of the data acquisition media path.

The aim in applying those ways scenario in a variety of stepping is to get various variations of footsteps on existing data acquisition media. Each person will step 24 times for each way or step scenario. The number of footsteps for each person in 3 ways or a stepped scenario is 72 footsteps. The total number of footsteps for ten people is 720 footsteps. In the recording process, the sound of footsteps recorded is the sound of footsteps walking alone.

The result of the footstep recording will be classified into two groups: Training Data and Testing Data. The composition used to separate the training data and the test data is 70:30 [18], based on the composition the amount of sound recording data used for the training process is 480 foot sounds recording while for the test process is 240 sound recordings footsteps. The result of footsteps sound recording can be seen in Fig 4.

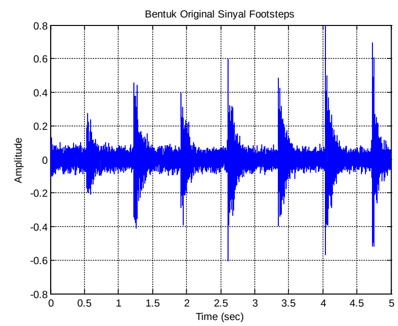

Fig. 4 The result of footsteps sound recording

6. Mel Frequency Cepstrum Coefficient Method (MFCC)

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions which cannot perceive frequencies over 1Khz. In other words, in MFCC is based on the known variation of the human ear's critical bandwidth with frequency [19]. MFCC has two types of the filter which are spaced linearly at a low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture the important characteristic of phonetic in speech. The MFCC phase used in this study can be seen in Fig 5.

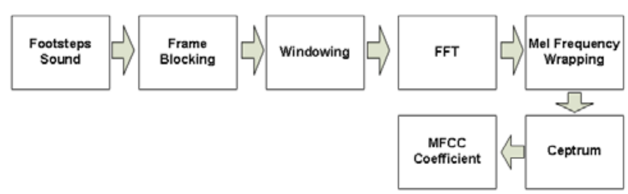

Fig. 5 The phases of MFCC used [19]

6.1 Frame Blocking

At this phase, the sound signal consisting of sample S (X (S)) is divided into several frames containing N samples, each frame is separated by M (M <N).

6.2 Windowing

After the Frame Blocking process has been done, then the next step is windowing each frame to minimize signal discontinuity at the beginning and end of each frame. If the window is defined as w(n), 0 ≤ n ≤ (N-1), which N is the number of samples in each frame, then the result of this process is signal [20], corresponding to Eq. (1):

$$w(n) = 0.54 + 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad (1)$$

in which: 0 ≤ n ≤ (N-1)

6.3 Fast Fourier Transform (FFT)

The next process is the Fast Fourier Transform, which converts each frame containing N samples from the time domain to the frequency domain. FFT is a fast algorithm for the implementation of Discrete Fourier Transform (DFT) which is operated on a discrete time signal consisting of N samples, which can be seen in Eq. (2) [20]:

$$f(n) = \sum_{k=0}^{N-1} y_k \, e^{-2\pi kn/N} \qquad (2)$$

in which: n = 0,1,2, … ,(N-1)

6.4 Mel Frequency Wrapping

The frequent formula which is used to convert from frequency to Mel Scale is according to Eq. (3) [20]:

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \qquad (3)$$

where: f is the frequency in Hertz

6.5 Cepstrum

The cepstrum is the reverse term for the spectrum. In this last step, Mel log spectrum is converted to cepstrum using Discrete Cosine Transform (DCT).
The result of this process is called Mel Frequency Cepstrum Coefficients (MFCC). This MFCC is the result of a cosine transformation of the short-term power spectrum logarithm that is expressed in the Mel-frequency scale. If Mel Power Spectrum Coefficients is denoted as Sk, k = 1, 2, …, K in Eq. (4), Minh N.Do [21] defines the coefficient of MFCC (ĉn) as:

$$\widehat{c} = \sum_{k=1}^{K} (\log S_k) \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right] \qquad (4)$$

in which: k=1,2,…,K

An example of a cepstrum resulting from MFCC process from the footstep sound on Fig 4 can be seen in Fig 6.

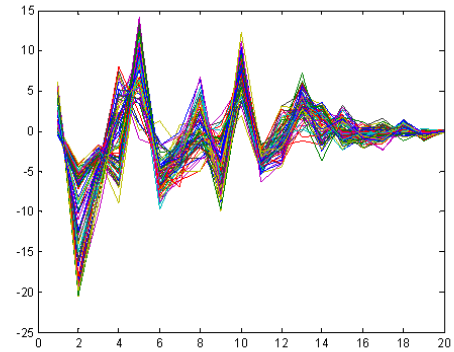

Fig. 6 Cepstrum result of MFCC process

The result of MFCC process from the footsteps sound that has been recorded is in the form of cepstrum as can be seen in Fig 6. Cepstrum in this study is a matrix of 78 x 20. That Cepstrum still not describe the specific characteristics of footsteps for everyone. Therefore, the main features of the cepstrum that can represent the cepstrum of each person.
In several studies related to the footsteps identification of people walking, to obtain the main characteristic of an existing feature matrix from the extraction results, some previous studies used the minimum, maximum value and average value approach of a given set of matrix values. In the research which was done by Orr [22], Orr changed the matrix value of foot pressure profile to the minimum and maximum values for each value in the foot profile to matrix profile. Suutala [23] changed the value of the feature foot matrix to the minimum, maximum and average values for each value of the matrix feature. Qian [24] changed the value in the data frame matrix of the footstep to a minimum, maximum and averages for each value that contained the footnote data frame.
In this study, the matrix value on cepstrum will be changed to the minimum, maximum and average value of each cepstrum value and to get the correct data centering measure will also add the standard deviation and median values. The minimum, maximum, mean, standard deviation and median values will become the input value of ANN.

7. Artificial Neural Network

Artificial Neural Network (ANN) is a neural network model that mimics the working principle of the neuron of the human brain (biological neurons). ANN first appeared after a simple model of artificial neurons introduced by McCulloch and Pitts in 1940 [25]. The simple model is based on the function of the biological neuron which is the basis of the signaling unit of the nervous system.

The working principle of ANN is based on the working mechanism of information distribution of human neural network system. However, because of the limitations of the ANN structure, only a small part of the human imaging system can be replicated [25].

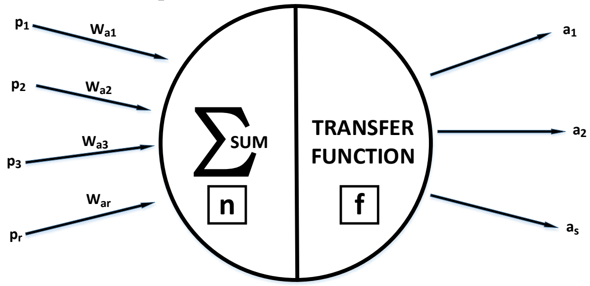

Fig. 7 Illustrated Model of JST [25]

The learning or training process of ANN is a process of altering or adjusting the strength of the relationships between the connected nodes [25]. Weight values express the strength of the relationships between nodes. The reliability of an ANN depends on its success in finding the weighted factor.

Back Propagation is one of the developments of Single Layer Neural Network architecture. This architecture consists of input layer, hidden layer, and output layer [25]. Each layer consists of one or more artificial neurons. Reverse propagation algorithm is one of the supervised learning techniques used in this research. Therefore, it is necessary to understand some important elements in reverse propagation method.

When the output does not result as expected, the output is propagated backward to the hidden layer and from the hidden layer to the input layer. The backpropagation network architecture as shown in Fig. 8, x1 to xn is the input layer, z1 to zp is the hidden layer, and y1 up to ym is the output layer.

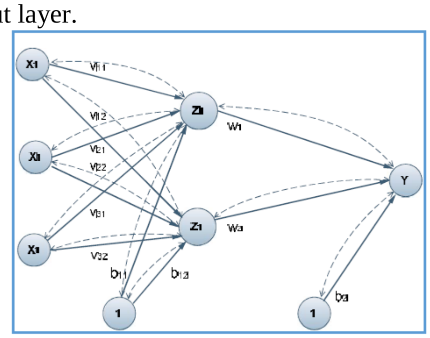

Fig. 8 The architecture of Back Propagation Network [26]

ANN architecture which is used in this study consisted of five inputs representing the minimum, maximum, mean, standard deviation and median values which are the specific characteristics of cepstrum. The number of hidden layers used is two hidden layers. There are ten output layers which are a representation of the number of people involved in this study. ANN architecture which is used in this study can be seen in Fig 9.

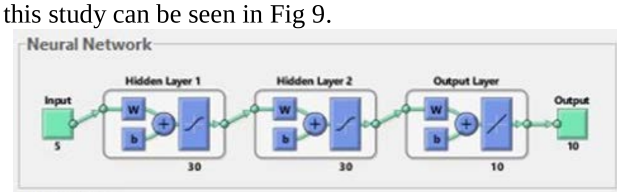

Fig. 9 The ANN architecture used

8. Result and Discussion

The result of the accuracy of person recognition based on footstep sounds obtained from this research can be seen in confusion matrix Fig 10, Fig 11 and Fig 12.

Target Class	Output Class 1	2	3	4	5	6	7	8	9	10	Total Target	Precision	Recall
1	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
2	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
3	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
4	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
5	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
6	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
7	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	7 9%	0 0%	0 0%	1 1%	8	88% 12%	100% 0%
8	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	8	100% 0%	100% 0%
9	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	8	100% 0%	100% 0%
10	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	8	100% 0%	89% 11%
Total Output	8	8	8	8	8	8	7	8	8	9	Total Data	80	
											Testing Accuration	98.75%	
											Average of Testing Precission	98.75%	
											Average of Testing Recall	98.89%	
											Error Rate	1.25%	

Fig. 10 Confusion Matrix footsteps in the first way or scenario

Fig. 10 shows that there exist misclassified class on class 7 which is classified as class 10. Hence, the precision of class 7 is 88%, and the recall of class 10 is 89%.

Target Class	Output Class 1	2	3	4	5	6	7	8	9	10	Total Target	Precision	Recall
1	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
2	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
3	0 0%	0 0%	7 9%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	1 1%	8	88% 0%	100% 0%
4	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
5	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
6	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
7	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	8	100% 12%	100% 0%
8	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	8	100% 0%	100% 0%
9	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	8	100% 0%	100% 0%
10	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	8	100% 0%	89% 11%
Total Keluaran	8	8	7	8	8	8	8	8	8	9	Total Data	80	
											Testing Accuration	98.75%	
											Average of Testing Precision	98.75%	
											Average of Testing Recall	98.89%	
											Error Rate	1.25%	

Fig. 11 Confusion Matrix footsteps in a second way or scenario

It can be seen from Fig. 11 that there exist misclassified class on class 3 which is classified as class 10. Hence, the precision of class 3 is 88%, and the recall of class 10 is 89%.

Target Class \ Output Class	1	2	3	4	5	6	7	8	9	10	Total Target	Precision	Recall
1	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
2	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
3	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	89% 0%
4	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	89% 0%
5	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
6	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	0 0%	0 0%	8	100% 0%	100% 0%
7	0 0%	0 0%	0 0%	1 1%	0 0%	0 0%	6 8%	0 0%	0 0%	1 1%	8	75% 12%	100% 0%
8	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	0 0%	0 0%	8	100% 0%	100% 0%
9	0 0%	0 0%	1 1%	0 0%	0 0%	0 0%	0 0%	0 0%	6 8%	1 1%	8	75% 0%	100% 0%
10	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	8 10%	8	100% 0%	80% 11%
Total Output	8	8	9	9	8	8	6	8	6	10	Total Data	80	
											Testing Accuration	95.00%	
											Averag of Testing Precision	95.00%	
											Average of Testing Recall	95.78%	
											Error Rate	5.00%	

Fig. 12 Confusion Matrix footsteps sound in a third way or scenario

There exist misclassified class on class 9 which is classified as class 3 and class 10 (see Fig. 12). The misclassification happens in the class as well. Class 7 is classified as class 4 and class 10. Hence, the precision of class 7 and class 9 is 75%, the recall of class 3 and 4 become 89%, and the recall of class 10 becomes 80%.

The decline of precision and recall on those class explained above, can cause the decline the accuracy of the system for every way or scenario of walking. The decline of accuracy leads to people's footsteps misidentification.

The variance of the footsteps ways or scenarios that have been prepared before lead to the misclassification.

The difference of those footsteps variance will cause the change of the distance between the sound source (footsteps sound) and the microphone. The difference leads to the intensity of the recorded sound.

A summary of the three confusion matrixes in Fig 10, Fig 11 and Fig 12 can be seen in Table 3.

Table 3: Summary of Confusion Matrix

Scenario	The result of Classification Testing			
	Accuracy (%)	Precision (%)	Recall (%)	Error Rate (%)
First Way or Scenario	98.75%	98.75%	98.89%	1.25%
Second Way or Scenario	98.75%	98.75%	98.89%	1.25%
Third Way or Scenario	95.00%	95.00%	95.78%	5%

Based on the results of this research, it can be seen from the value of accuracy obtained for each way or scenario stepping that the distance between the microphones with the source of the footsteps sound is very influential.

The position of microphone placement on the data acquisition system affects the number of footage variations that can be recorded. This position is quite important because on any side the person stepping can still be recognized well.

The number of microphones which is used to record footstep sound also provides an important contribution to the accuracy of people's recognition based on footsteps. The number of microphones used will affect the quality of the sound footsteps recorded.

The study, based on Table 3, has the better accuracy compared to the previous studies. Itai and Yasukawa (2006) used only one microphone which placed on the middle side of the track, resulted in 96% of accuracy. Meanwhile, this study used four microphones and placed in the left, the middle and the right side of the track; obtained 98.75% of the accuracy. It can be summarized that the number of the sensors and changing the position of the microphones increased 2.75% of accuracy.

9. Conclusion

The results of this study prove that the sound of footsteps can be in the form of a person's biometric traits, which can be used as identifiers of the person when walking. It can be seen from the accuracy obtained in Table 3.

The accuracy of person recognition based on footsteps sound depends on the position of the microphone and the number of microphones used in the data acquisition system. This is because the microphone position and the number of microphones used, affect the accuracy of foot recognition for various variations of the person's steps on the existing data acquisition. The various variations of the person's steps can be recognized because the microphones are placed at the right and the left side of data acquisition track. Hence, the distance between the source of the sound (footsteps sound) and the microphones is relatively the same in every way or scenario.

In the previous studies, identification of footsteps of person was based on one way of stepping, which was on the middle side. While in this study used three ways or scenarios with the average accuracy is 97.5%.

References

[1] Babich A, K, Nagwanshi and S. Dubey, 2012, Biometric authentication using human footprint," Int. J. Appl. Inform. Syst., vol. 3, no. 7, pp.1-6.

[2] Abhyankar, A., Schuckers, S,2010, A novel biorthogonal wavelet network system for off-angle iris recognition', Pattern Recognit., 43, pp. 987–1007

[3] Pankanti, S, Prabhakar, S, Jain, A.K, 2002, On the individuality of fingerprints, IEEE Trans. Pattern Anal. Mach. Intell., Vol 24, pp. 1010–1025

[4] Kong, A., Zhang, D., Kamel, M, 2009, A survey of palmprint recognition, Pattern Recognit., Vol 42, pp. 1408–1418

[5] Vajaria, H., Islam, T., Mohanty, P., 2007, Evaluation and analysis of a face and voice outdoor multi-biometric system, Pattern Recognit. Lett., Vol 28, pp. 1572–1580

[6] Middleton L, Buss, A.A., Bazin, A.I., and Nixon, M.S., 2005, A floor sensor system for gait recognition, Proc. IEEE Workshop Automatic Identification Advanced Technologies, pp 44 – 50

[7] Rodriguez, R.V., Lewis, R.P., Mason, J.S.D., and Evans, N.W.D., 2008. Footstep Recognition for a Smart Home Environment. Int'l Journal Smart Home, Vol. 2, pp 95-110.

[8] Gafurov, D., 2007, A Survey of Biometric Gait Recognition: Approaches, Security, and Challenges. NIK Conference, pp 53 – 60.

[9] Boulgouris, N.V., Hatzinakos, D., Plataniotis, K.N., November 2005, Gait Recognition: A challenging signal processing technology for biometric identification, IEEE Signal Processing Magazine, pp 78 – 90.

[10] Tanaka, M., and Inoue, H., 1999. A Study on Walk-Recognition by Frequency Analysis of Footsteps. IEEE Japan, Nonlinear Signal, and Image Processing.

[11] Shoji, Y., Takasuka, T., and Yasukawa, H., 2004. Personal Identification Using Footstep Detection. Proc IEEE Intelligent Signal Processing and Communication Systems (ISPACS).

[12] She, B., 2004, Framework of footstep detection in indoor environment, International Congress On Acoustics; 18th.

[13] Itai, A., and Yasukawa, H., 2006, Footstep Recognition with Psyco acoustics Parameter, Circuits, and Systems. APCCAS 2006, IEEE Asia Pacific Conference on Digital Object Identifier, pp 98 – 102.

[14] Guo, F, and Wang, X, 2017, Robust footstep identification system based on acoustic local features, IET Biom., Vol. 6 Iss. 6, pp. 387-392

[15] Webb, A. R., and Copsey, K. D.,2011, Statistical Pattern Recognition. Third. A John Wiley & Sons, Ltd, Publication

[16] Nhut DT, Rosa, P.F.F., Lima, S.S., Botelho, W.T., Nascimento, A.F., and Alaluna, M.S., 2012, A pervasive identification and adaptation system for the smart house, AIAI, pp 141–154.

[17] Nurcahyo, Putra, D., 2015, Sistem Biometrika: Konsep Dasar, Teknik Analisa Citra dan Tahapan Membangun Aplikasi Sistem Biometrika. Andi Offset, Yogyakarta.

[18] Tahir,A., 2015, Implementasi Jaringan Syaraf Tiruan dengan Algoritma Propagasi Balik untuk Pengenalan Pola Angka, JTRISTE, Vol.2, No.1, Maret 2015, pp. 1-12, ISSN: 2355-3677.

[19] Ganchev, T., Fakotakis, N., and Kokkinakis, G., 2005. Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task, Proc. of the SPECOM, pp 23 – 35.

[20] Setiawan A, Hidayanto, Isnanto RR, 2011, Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel Frequency Cepstrum Coefficient (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer,Jurnal TRANSMISI Vol 1(2), pp 82 – 86

[21] Lyons,J, 2013 , Mel Frequency Cepstral Coefficient (MFCC) tutorial, accessed on 24 March 2018.

[22] Orr, R.J., and Abowd, G.D., 2000, The Smart Floor: A Mechanism for Natural User Identification and Tracking. Proc. CHI ’00 Extended Abstracts on Human Factors in Computing Systems, pp 152 – 160.

[23] Suutala, J., and Roning, J., 2005. Combining Classifiers with Different Footstep Feature Sets and Multiple Samples for Person Identification. Proc. Int’l Conf. Acoustics, Speech, and Signal Processing, Vol. 5.

[24] Qian, G., Zhang, J., and Kidane, A., 2010. People Identification Using Floor Pressure Sensing and Analysis. IEEE Sensors J., vol. 10, Vol. 9, pp. 1447-1460.

[25] Puspitanigrum D. 2006, Pengatar Jaringan Syaraf Tiruan. Yogyakarta: Andi Yogyakarta.

[26] Kusumadewi S. 2004, Membangun Jaringan Syaraf Tiruan Menggunakan MATLAB & EXCEL LINK. Yogyakarta. Graha Ilmu.

Jan Everhard Riwurohi. He received his Undergraduate degree (SKom) in Computer Engineering from Universitas Budi Luhur, Indonesia; Master degree in Industrial Control (MT.) from Universitas Indonesia. He is a lecturer at the Department of Computer Engineering, Faculty of Information Technology, Universitas Budi Luhur. His research interests are automation control and robotics

Jazi Eko Istiyanto. A Full Professor (awarded 2010) of Electronics and Instrumentation at Universitas Gadjah Mada, a Ph.D. (completed 1995) in Electronic Systems Engineering and M.Sc (1988) in Computer Science, and Post Graduate Diploma (1987) in Computer Programming and Microprocessor Application, from University of Essex, UK. He currently also serves as the Chairman of Indonesia Nuclear Energy Regulatory Agency. He also holds a B.Sc (1986) in Physics from Universitas Gadjah Mada. His research interests cover Embedded Systems, Digital Systems Optimization, and Information Security. Prof. Dr. Istiyanto is with the Department of Computer Science and Electronics, Universitas Gadjah Mada.

Khabib Mustofa. He received his Undergraduate degree (S.Si.) in Computer Science from Universitas Gadjah Mada, Indonesia; Master degree in Computer Sciences (M.Kom.) from Universitas Gadjah Mada, and Dr. techn from Institute for Software Engineering and Interactive Systems, Technische Universit¨at Wien (Vienna University of Technology), Austria. He is a lecturer at the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. His research interests cover Web technology (Semantic web, Web Services), mobile application and Information Management.

Agfianto Eko Putra. A Ph.D. (graduated in 2010) and M.Si (graduated in 1998) in Physics from Universitas Gadjah Mada. He also holds S.Si (graduated in 1992) in Physics from Universitas Gadjah Mada. He is a lecturer at Electronics & Instrumentation, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. His research interests cover Digital Signal Processing, Embedded Electronics, and Satellite System.