



Wearable Sensor-Based Biometric Gait Classification Algorithm Using WEKA

Ik-Hyun Youn^{1*}, Kwanghee Won¹, Jong-Hoon Youn¹, and Jeremy Scheffler², *Member, KIICE*

¹Department of Computer Science, University of Nebraska at Omaha, Omaha, NE 68106, USA

²Pius X High School, Lincoln, NE 68510, USA

Abstract

Gait-based classification has gained much interest as a possible authentication method because it incorporates an intrinsic personal signature that is difficult to mimic. The study investigates machine learning techniques to mitigate the natural variations in gait among different subjects. We incorporated several machine learning algorithms into this study using the data mining package called Waikato Environment for Knowledge Analysis (WEKA). WEKA's convenient interface enabled us to apply various sets of machine learning algorithms to understand whether each algorithm can capture certain distinctive gait features. First, we defined 24 gait features by analyzing three-axis acceleration data, and then selectively used them for distinguishing subjects 10 years of age or younger from those aged 20 to 40. We also applied a machine learning voting scheme to improve the accuracy of the classification. The classification accuracy of the proposed system was about 81% on average.

Index Terms: Classification, Gait analysis, Machine learning algorithm, WEKA

I. INTRODUCTION

Human gait is defined as a personal walking pattern using two limbs. Although the definition of human gait is simple, measurement of human gait patterns requires sophisticated techniques to capture the essence of human gait, including natural variations in gait [1]. Many researchers have recognized individual gait patterns as an authentication method. Unlike most of the previous gait classification approaches, this study uses an open data set with a large number of subjects for practical gait classification. Moreover, we chose machine learning algorithms to identify the effects of various factors influencing gait patterns. Particularly, we used a collection of machine learning algorithms in the Waikato Environment for Knowledge Analysis (WEKA)

open-source software package [2]. WEKA allowed us to apply various types of machine learning algorithms to find appropriate machine learning techniques for human gait classification. The goal of this gait classification study was to use machine learning algorithms to efficiently classify both mature and immature gait groups from a single sensor-based gait feature.

This study used an open gait database collected by an inertial sensor-based system [3]. In order to extract a vector of gait features, each gait needed to be accurately recognized. Temporal gait feature computation was applied to feature extraction because of its simple computation [4]. Statistical methods were also used to obtain more information about time series gait patterns [5]. For this study, we selected three machine learning algorithms from among those available in

Received 16 November 2015, Revised 26 November 2015, Accepted 17 December 2015

*Corresponding Author Ik-Hyun Youn (E-mail: iyoun@unomaha.edu, Tel: +1-402-250-8804)

Department of Computer Science, University of Nebraska at Omaha, Omaha, NE 68106, USA.

Open Access <http://dx.doi.org/10.6109/jicce.2016.14.1.045>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

WEKA based on relevant study [6]. The three algorithms resulted in an average of 81% accuracy in differentiating subjects who were below 10 years of age from the entire set of 350 participants. The proposed approach combines a majority voting technique to enhance classification accuracy.

The rest of the paper is organized as follows. Section II describes the characteristics of the open human gait data with an experimental environment, and elaborates on the proposed gait recognition and feature extraction techniques. In Section III, based on the features discussed in the previous section, we propose a new classification method and apply the gait recognition method using three machine algorithms. Finally, we present our conclusions in Section IV.

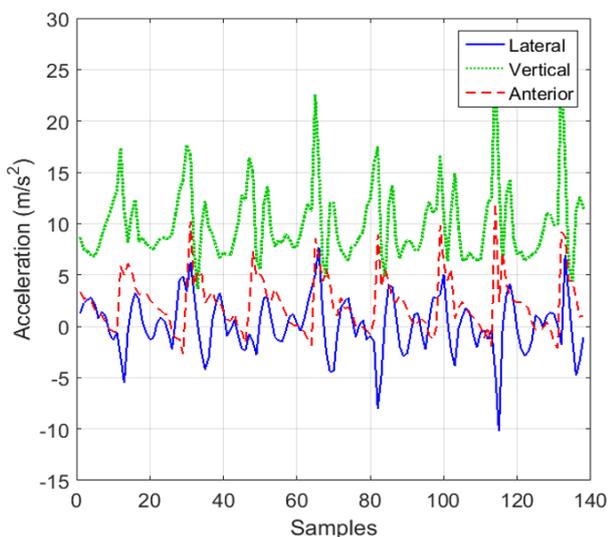


Fig. 1. Raw three-dimensional acceleration.

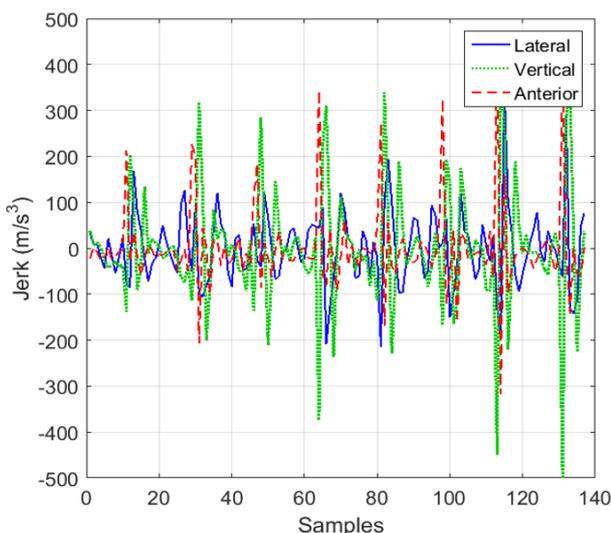


Fig. 2. Anterior direction jerk and identified step indices.

Table 1. Subject information

Age (yr)	No. of subjects	Maturity of gait
<10	157	Immature gait
20–40	193	Mature gait
Total	350	

II. GAIT FEATURE EXTRACTION

The data sets from the inertial sensor-based gait database [3] were analyzed through the gait recognition algorithm proposed in [7]. The first step of the proposed approach was the detection of each gait cycle. After the gait cycle detection, temporal and statistical gait features were computed. A total of 24 gait features were extracted for gait analysis in the next step.

A. Data Source

In this study, we used an open gait database collected by an inertial sensor-based system [3]. Although the data set includes sensing data collected from different types of sensors, we utilized only the data set collected from a single low-back trunk accelerometer. The experiment was conducted on 744 subjects with ages ranging from 2 to 78 years. The original research group evaluated the performance of the gait authentication scheme with various age groups. They observed that those below 10 years of age and those above 50 years of age showed a relatively low classification accuracy compared to other age groups. They concluded that classification accuracy depended on whether each age group had walking skill maturity or not.

For the experimental study, we selected two age groups: those with a mature gait and those with an immature gait. Table 1 shows the age range of each group.

B. Gait Recognition

In order to extract distinctive gait features, each gait needs to be accurately recognized. To identify each step motion, we captured heel-strike action by observing changes in acceleration at the heel strike moment.

While examining three-dimensional acceleration data, we noticed that a jerk, defined as a change in the rate of acceleration over time, showed more dramatic pattern changes than raw acceleration data at each heel strike. As depicted in Fig. 1, each directional acceleration showed a peak at the heel strike; however, the raw acceleration graph has several small peaks that can be considered a heel strike point for each step. Fig. 2 shows the jerk data obtained from the raw acceleration data. The jerk graph shows a clearly noticeable pattern for the heel strike action. In particular, the post-anterior jerk can be used to accurately identify each

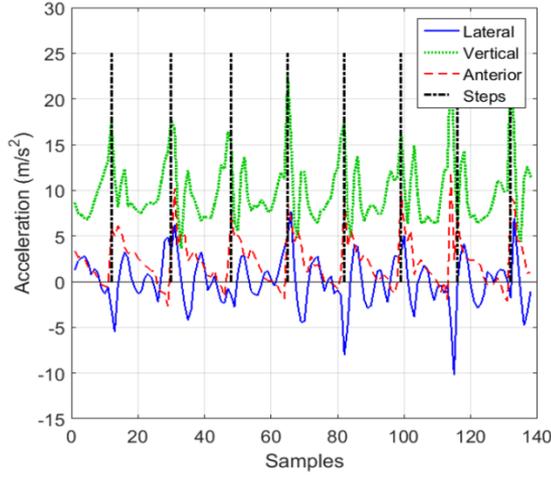


Fig. 3. Recognized gait using the threshold from the jerk and recognized steps.

heel strike.

In Fig. 3, the black dotted lines delimit each step. The black vertical lines are matched to original raw acceleration data to identify each step. The raw acceleration data was then used to extract gait features in the next step.

C. Feature Extraction

There are three major approaches to extracting gait features: temporal feature-based, frequency analysis-based, and statistical analysis methods [3]. In this study, we applied both temporal feature analysis and statistical approaches. Since the temporal gait feature computation is relatively simple, it can be applied to online gait classification systems [4]. Typically, statistical methods require more computations than the methods using temporal gait feature extraction; however, statistical methods can provide more information about time series gait patterns [5].

Table 2 categorizes the 12 selected fundamental gait features into either temporal or statistical gait features.

We applied several well-known statistical techniques to capture the characteristics of natural gait variation. Initially, a total of 12 gait features were extracted from the accelerations of each gait cycle, and then the averages of each feature were used for classification. The root mean square (RMS) of vertical acceleration and the signal magnitude area (SMA) of 3-axis signals were computed using Eqs. (1) and (2), respectively.

$$X_{RMS} = \sqrt{\frac{1}{N}(x_1^2 + x_2^2 + \dots + x_n^2)}, \quad (1)$$

$$SMA = \frac{1}{t} \times \left(\int |x(t)| dt + \int |y(t)| dt + \int |z(t)| dt \right). \quad (2)$$

Table 2. Fundamental gait features

Type	Name	No. of features
Temporal	Step duration	1
	Standard deviation	3
	Mean	3
Statistical	Root mean square	3
	Signal magnitude area	1
	Vector magnitude	1
Total basic features		12

Table 3. Symmetry gait features

Type	Name	No. of features
Temporal symmetry	Symmetry of step duration	1
	Symmetry of standard deviation	3
Statistical symmetry	Symmetry of mean	3
	Symmetry of root mean square	3
	Symmetry of signal magnitude area	1
	Symmetry of vector magnitude	1
Total symmetry features		12

The 12 gait features listed in Table 2 were used to determine gait symmetry features. Gait symmetry is defined as a perfect agreement between the actions of the lower limb [8]. Two consecutive step features were used to calculate the symmetry between the movement of the left and right limbs. A list of symmetry gait features is shown in Table 3.

III. CLASSIFICATION

To efficiently categorize the subjects into two different groups, we have tested various classification algorithms in WEKA to choose higher accuracy algorithms which require inexpensive computation cost in testing phase. We chose three well-known machine-learning approaches: support vector machine (SVM) [9], random forest [10], and logistic regression [11]. The SVM classifier trains a hyperplane that maximizes the margin between two different clusters. In our experiments, linear kernel SVM was used. The resultant support vectors and weights can be represented simply by a linear combination of them. Random forest is an ensemble learning approach which makes use of randomized decision trees. It is easy to implement and has shown reasonable performance in many applications. On the other hand, logistic regression makes use of a logistic function. The logistic function takes the inner product of coefficients and a feature vector, and maps the feature vector to a specific class (0 or 1). The classifier can be trained by finding the best coefficients for all given feature vectors and the class labels of the training data set.

We also used the three trained classifiers to build a

combined classifier that takes the label of majority votes among the results of the three algorithms. Three different classifiers can generate different decision boundaries. Thus, the combination of these decision boundaries can represent more complex shapes, and it can correctly classify some difficult examples. Moreover, once the classifiers are trained, the computational cost of the testing phase of the algorithms is relatively inexpensive. For example, the linear kernel SVM and logistic regression predict the label of a given feature vector in two steps. First, they take the inner product of a feature vector and the coefficients. Then, the values are used to evaluate a step function.

As we mentioned earlier, the data set contains 3-axial acceleration values collected from walking subjects. We chose two age groups from the open gait database [3]. One was an age group younger than 10 years old, and the other was a group of subjects in their 20s or 30s. The numbers of subjects for the age group 0–10 years and for the age group 20–40 years were 157 and 193, respectively.

To evaluate the performance of the three existing algorithms and one combined method, we prepared five sets of experiments. For each set, we randomly assigned each feature vector to one of 10 subsets (of equal size) to perform a 10-fold cross-validation. The 10-fold cross-validation divides the data set into 10 subsets; each trial alternately chooses one subset for the testing phase and the remaining nine subsets for the training phase. We measured the accuracy of each algorithm. Where TP is a measure of true positive, TN is a measure of true negative, and N is the number of total feature vectors, accuracy is measured by Eq. (3):

$$Accuracy = \frac{(TP + TN)}{N}. \tag{3}$$

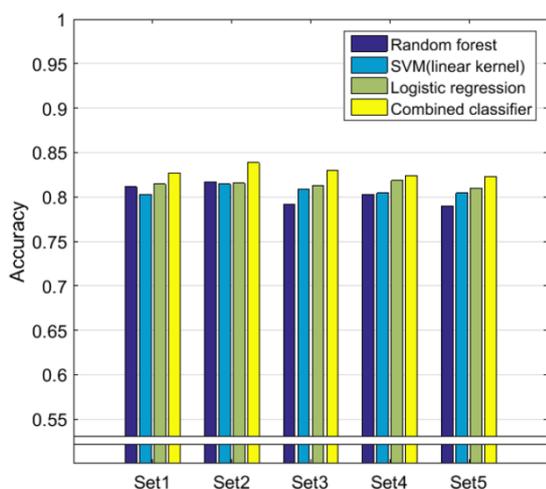


Fig. 4. Classification accuracy of machine learning algorithms for gait classification between two age groups.

IV. RESULTS

Table 4 and Fig. 4 show the accuracy of classification algorithms for each experimental set. The logistic regression classifier shows the highest average accuracy among three machine-learning algorithms. The combined classifier slightly increases the classification accuracy (by about 2%) for each experiment. We selected 10 features using a parameter selection method in WEKA. Table 5 shows the top six features that were identified. Then, we measured the accuracy of each classifier after adding the remaining features one at a time until the 10th feature.

The machine learning algorithms with the features proposed in this paper achieved about 81% accuracy on average. Although we defined 24 gait features for the classification in Section II, they are not equally important for the classifiers. Based on the results of our experiments, the classifiers with all 24 features achieved the best prediction accuracy in general. However, in order to minimize the computational complexity of the classifier, we may not want to use all features. Thus, the goal of this experiment was to find a minimum set of gait features without compromising the accuracy of the classifiers.

First, we evaluated the importance of the gait features using the support vector machine attribute selection algorithm available in WEKA. Table 5 lists the top six gait features chosen by the attribute selection algorithm.

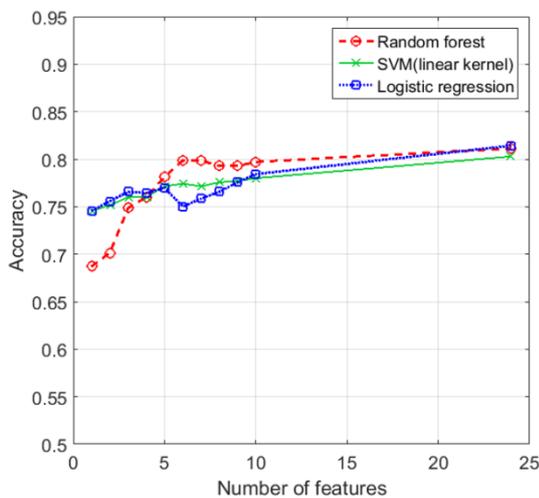
Fig. 5 shows the prediction results of the top 10 gait features as well as all 24 gait features discussed in Section II. Based on the results of our experiments, the step time, a key temporal gait feature, was identified as the most important gait feature. As shown in Fig. 5, for both SVM and logical regression, the prediction accuracy with only the step time feature was about 75%. The standard deviation of the lateral and vertical acceleration dispersion features improved the accuracy of the classifier when it was combined with the step time parameter. Finally, overall step acceleration features such as the signal magnitude of the area and the vector magnitude of each step also helped improve the performance of the classifiers.

Table 4. Classification accuracy of machine learning algorithms for gait classification between two age groups

Trial	Random forest	SVM linear kernel	Logistic regression	Combined classifier
Set1	0.8114	0.8029	0.8143	0.8271
Set2	0.8171	0.8143	0.8157	0.8386
Set3	0.7914	0.8086	0.8129	0.8300
Set4	0.8029	0.8043	0.8186	0.8229
Set5	0.7900	0.8043	0.8100	0.8229
Average	0.8026	0.8069	0.8143	0.8286

Table 5. Top six gait features to predict two age groups with highest accuracy

No.	Feature	Description
1	Step time	Average step time
2	SD of vertical accuracy	Standard deviation of vertical acceleration for each step
3	RMS of lateral accuracy	Root mean square of lateral acceleration for each step
4	SD of vertical accuracy	Standard deviation of lateral acceleration for each step
5	Signal magnitude area	Signal magnitude area of each step
6	Vector magnitude	Vector magnitude of each step

**Fig. 5.** Classification accuracy of machine learning algorithms according to the number of features.

Based on the results of this experimental study, we recommend the top six gait features shown in Table 5 as a minimum set of gait features to be used for the classifier while maintaining prediction accuracy.

V. DISCUSSION AND CONCLUSIONS

This study examined the classification accuracy of subjects from two different age groups. We treated a younger age group as an immature gait group and an adult group as a mature gait group. We chose machine learning algorithms to handle multidimensional gait features. We applied various classifiers conveniently to compare compatible machine learning algorithms for a vector of gait features using WEKA. As a result of the experiments, we achieved about 80% classification accuracy in distinguishing between the immature gait group and the mature gait group.

REFERENCES

- [1] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, pp. 148-155, 2002.
- [2] G. Holmes, A. Donkin, and H. Witten, "WEKA: a machine learning workbench," in *Proceedings of the 1994 2nd Australian and New Zealand Conference on Intelligent Information System*, Brisbane, Australia, pp. 357-361, 1994.
- [3] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognition*, vol. 47, no. 1, pp. 228-237, 2014.
- [4] D. Gafurov, E. Snekkenes, and P. Bours, "Improved gait recognition performance using cycle matching," in *Proceedings of 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Perth, WA, pp. 836-841, 2010.
- [5] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," in *Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE2007)*, Wuhan, China, pp. 543-546, 2007.
- [6] H. Chan, M. Yang, H. Wang, H. Zheng, S. McClean, R. Sterritt, and R. E. Mayagoitia, "Assessing gait patterns of healthy adults climbing stairs employing machine learning techniques," *International Journal of Intelligent Systems*, vol. 28, no. 3, pp. 257-270, 2013.
- [7] I. H. Youn, S. Choi, R. Le May, D. Bertelsen, and J. H. Youn, "New gait metrics for biometric authentication using a 3-axis acceleration," in *Proceedings of 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, pp. 596-601, 2014.
- [8] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and limb dominance in able-bodied gait: a review," *Gait & Posture*, vol. 12, no. 1, pp. 34-45, 2000.
- [9] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293-300, 1999.
- [10] A. Liaw and M. Wiener, "Classification and regression by RandomForest," *R News*, vol. 2, no. 3, pp. 18-22, 2002.
- [11] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons, 2005.



Ik-Hyun Youn

is currently a Ph.D. student in the College of Information Science and Technology at the University of Nebraska at Omaha. He received his M.E. from Korea Maritime and Ocean University in 2011 and B.E. from Mokpo National Maritime University in 2004. His current research interests are focused on mobility monitoring using wireless sensors and IT solutions to improve maritime transportation safety.



Kwanghee Won

received his Ph.D. degree in Computer Science and Engineering from Kyungpook National University, Korea. He received B.S. and M.S. degrees in Computer Engineering from the same university. He is currently a Research Scholar in the College of Information Science & Technology, University of Nebraska Omaha. His research interests include 3D computer vision and mobile sensors.



Jong-Hoon Youn

received the M.S. and Ph.D. degrees in Computer Science from Oregon State University in 1999 and in 2002, respectively. He is currently an Associate Professor in the Department of Computer Science at University of Nebraska, Omaha (UNO). His current research interests are focused on wireless sensor networks, development of mobile applications, design and analysis of low-power communication protocols, and mobility monitoring using wireless sensors.



Jeremy Scheffler

currently teaches physics courses at Pius X High School in Lincoln, Nebraska. He received his M.A.T. in 2006 and B.S. in 2001 from the University of Nebraska-Lincoln. His contributions to this research resulted from his participation in the Infusing Mobile Platform Applied Research into Teaching (IMPART) Research Experience for Teachers (RET) supported by the National Science Foundation under Grant No. CNS-1201136.