American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-7, Issue-6, pp-223-233
www.ajer.org
Research Paper Open Access

Footstep Detection with HoloLens

Yingnan Ju
[1](ISE, School of Informatics, Computing, and Engineering, Indiana University Bloomington, US)
*Corresponding Author: Yingnan Ju

ABSTRACT: *This paper presents results on using AR/MR device (HoloLens) and machine learning approach for detecting whether a user wearing the device is touching the ground with the left or the right foot while walking at different speeds and in different directions. The accuracy of the prediction reached 88% even after the user's gait data had been collected for less than 1 minute. Applications include rendering footprints in games in real time, body movement recognition for personalization, recording and replaying guided tours by experts.*
KEYWORDS *-Augmented reality, mixed reality, HoloLens, gait identification, personalization*

Date of Sumisión: 06-06-2018 Date of aceptante: 21-06-2018

I. INTRODUCTION

Augmented Reality (AR) and Mixed Reality (MR) devices such as cell phones, tablets, or glasses support diverse user interactions with real world environments. However, current AR/MR devices detect rather limited human body movements with built-in sensors. For example, Microsoft HoloLens can only detect its own movements (including position, velocity and acceleration) with sensors and limited gestures (tap, bloom, and drag) with the built-in Kinect. Cell phones and tablets running AR SDK face similar problems. Basic human body movements, e.g., walking and running or nodding and shaking the head, cannot be discerned. Yet, basic body movements are important for enhancing the interaction between humans and machines as part of an immersive experience.

Fortunately, walking and running have unique patterns, that can be recognized using machine learning (ML) algorithms given sufficient data of body movements [1][2]. All needed data can be acquired from motion sensors readily available in cell phones, tablets, or glasses. The gait analysis method based on wearable sensors, which are inexpensive and can be applied outside the laboratory environment, was studied and has shown great prospects in the recent two decades [3]. This article presents first results from using basic motion sensors of wearable AR/MR devices to train a ML model for wearable AR/MR devices that is able to identify when a user is walking, when the foot is touching the floor, and which foot is stepping: left foot or right foot. The ultimate goal is to detect and act on identified user patterns, e.g., to give users audio feedback such as the sound of footsteps, and visual feedback such as footprints based on the detected paces of the user to offer a more immersive experience, as shown in Fig. 1.

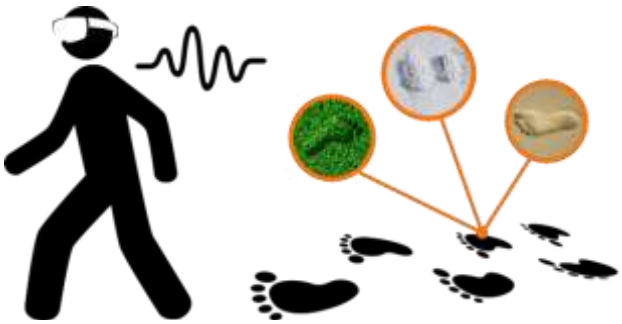


Fig. 1. Different feedbacks in different scenarios

The significance of our work is to improve the experience of mixed reality by introducing more reality (the real steps of the user) into the virtual world. This will bring a new kind of interaction between the user and the machine. The interaction is natural and will be processed unconsciously by the user while he/she is walking. Then the wearable AR/MR devices give corresponding feedbacks to the user and make the user feel the mixture of reality and the virtual world.

II. RELATED WORKS

There was no same kind of work before our research, but there were some similar researches in related fields, for example, vision-based gesture recognition or identification of user through the gait detected by cellphone. We will discuss and analyze these researches and tell the difference between them and our work.

A. Vision-based gesture recognition

Research of hand gesture recognition with focus on various recognition techniques has been focused on since 1999 [4]. There have been a large number of research works carried out during last twenty years [5]. Most AR/MR devices recognize hand gesture or body movements of a third party based on the pictures from the front/rear camera, with or without built-in depth sensors [6][7], but normally they are not able to recognize the wearer's body movements, especially for wearable AR/MR devices like HoloLens which has only a front camera.

The vision-based recognition could detect the walking through the camera and therefore, but it cannot recognize the gait of the user itself. Our research acquires motion data but not vision data, so it can recognize the footstep of the user.

B. Cellphone-based AR devices

Cellphone-based AR applications capture real-time pictures from the camera and display them on the screen. The user usually holds the cellphone within sight to experience the augmented reality. Although the cellphone has enough motion sensors to collect position, velocity and acceleration to realize a s similar result of our research, the disadvantage of cell phone-based AR is that there are too many disturbances to the position and rotation of the phone when it is hand-held. The position and rotation of a wearable AR/MR device like HoloLens (fixed on the head) is more stable when the user is walking. Therefore, wearable AR/MR devices like HoloLens are the best targets to implement the experiment, and they also benefit most from our research.

C. Cell phone-based gait recognition

There are several papers about approaches for gait identification based on biometric gait using accelerometer [8][9]. Some papers focus on the identification of a specific person based on the pattern of motion data collected, e.g. the owner of the cellphone [9][10], and some papers focus on the identification of some specific disease which can cause unique patterns of gait [11][12]. Our research focuses more on recognizing detailed phase of walking.

In summary, our research is a small-scale exploration in AR/MR field and we are trying to detect one of the basic body movements of the user via simple motion data. There was no same research and experiment, so there was no baseline of the result of the experiment and we will use random baseline in the evaluation section.

III. METHODS

A. Gait cycle analysis

The typical walking pattern of human is shown in Fig. 2. In a left-handed coordinate system, the positive x, y and z axes (x+, y+, and z+) point right, up and forward, respectively. If a human is walking forward along z-axis (z+), the body will slightly lean to the left/right side (x-/x+) as the left/right foot is going to touch the ground. When a foot is starting to touch the ground, the body and the head of the human is at a lower position (y-) because the user is swinging one leg and there is an angle between legs and the ground. Therefore, when the user is leaning to the left and the head or body is at a lower position (x- and y-), we will know that one foot of the user starts to touch the ground and that foot is the left foot. The similar situation happens to other side (right: x+ and y-).

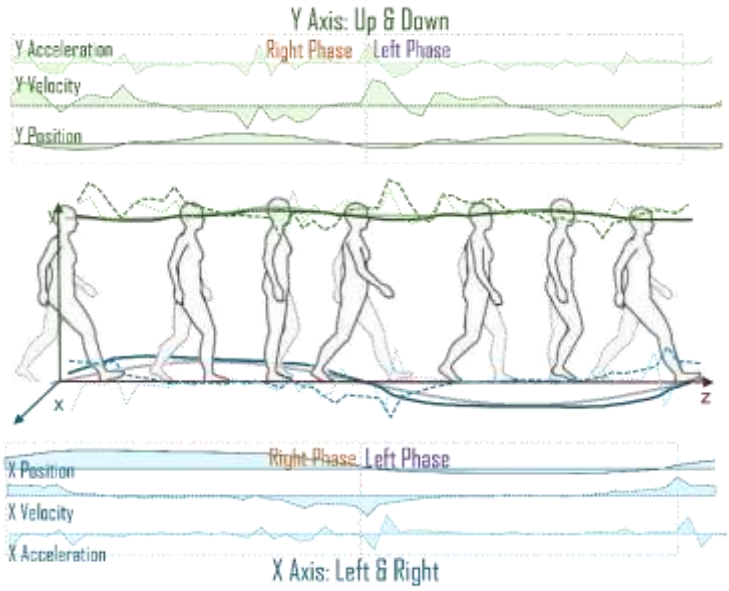


Fig. 2. Human gait cycle analysis

B. Data acquisition

We chose Microsoft HoloLens to be the wearable AR/MR device, which is a mixture of glasses and helmet and can be fixed steadily on the head of the user, as shown in Fig. 3. HoloLens can store position and rotation data as the user walks in natural status and the precision of the position data can reach 0.1mm.

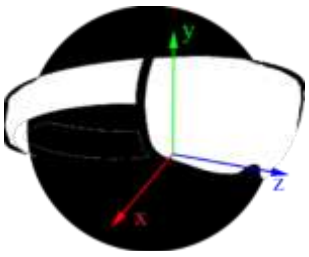


Fig. 3. HoloLens and the coordinates system

When an application starts in HoloLens, it will set up a left-handed coordinate system (the positive x, y and z axes point right, up and forward, respectively) and the position of HoloLens at that exact time is the origin point. Then we can get the position data relative to the origin point from HoloLens as the components of the position on three axes were stored in the HoloLens:

$$Position = \{P_x, P_y, P_z\}$$

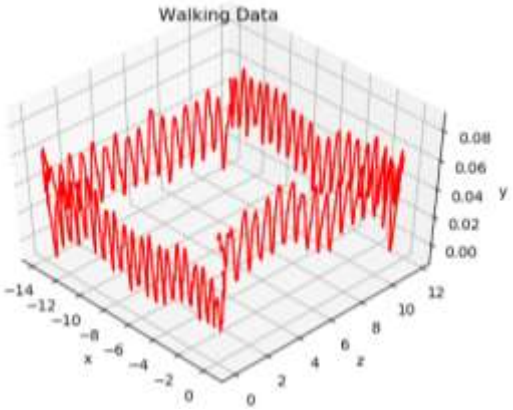


Fig. 4. Walking trace data in space

The data was logged in the "*Update*" function that was called every frame, it is called 60 times in one

second if the program suffered no lag or no frame drop. However, in reality, the frames are not always stable and the interval between frames is not consistent. Therefore, we logged data in the "*FixedUpdate*" function, which is called every fixed framerate frame. We set framerate to 50 to make sure there would be consistent 50 updates each second and the interval was 20 milliseconds constantly.

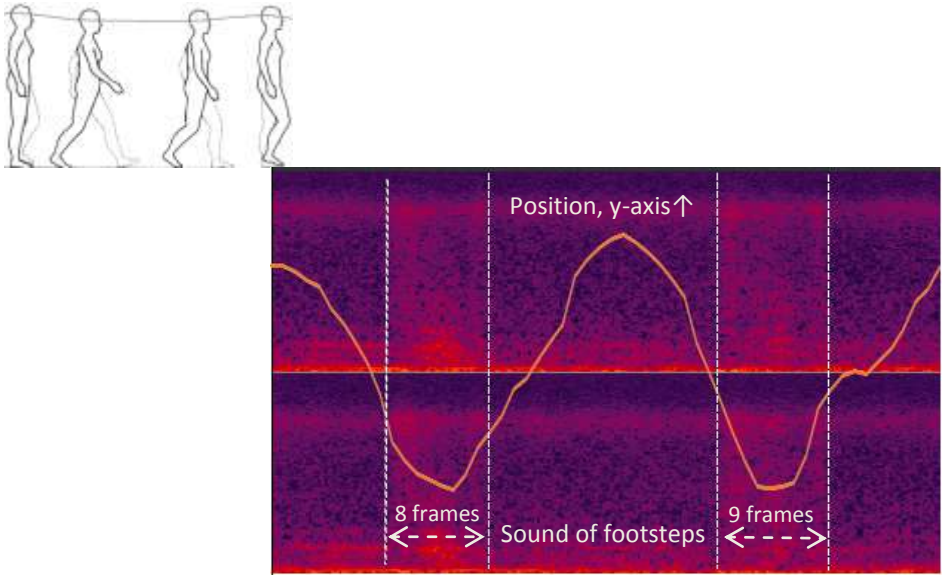


Fig. 5. Soundwave of footstep in one gait cycle

To help labeling the spatial data that we get from HoloLens for machine learning, we also recorded the sound of footsteps in real time. With the help of the visualized analysis of the soundwave, we could label the data points directly based on the timing of the footsteps. In Fig. 5 we could see the timing of the footsteps of the user and the corresponding position of the user synchronized at the same time point. If we put the visualization of the soundwave and the spatial waveform on y-axis together, as in Fig. 5, we can see that the y-axis values of the position (p_y) from data points collected were at the bottom of the wave when the user's foot was touching the ground, which means the user's head was at a lower position. That feature ($p_y < 0$) could be used to label the data points as "a foot starting to touch the ground". As mentioned in section A, to identify which foot, we also needed the x-axis value (p_x). When the user's body and head was leaning to the left, x-axis value is negative in left-handed coordinate system ($p_x < 0$), and vice versa for the right side ($p_x > 0$). These features could be used to label the spatial data points as "L" (left foot starting to touch the ground, $p_y < 0\ and p_x < 0$), "R" (right foot starting to touch the ground, $p_y < 0\ and p_x > 0$) and "N" (intermediate states between L and R, $p_y > 0$) approximately as a comparison with the data labeled directly based on the visualized analysis of the sound recorded (also labeled as "L", "R" and "N").

C. Data process

From the position data that we got from HoloLens, we calculate a corresponding velocity via taking derivatives of the position data and calculate the acceleration via taking a second derivative:

$$Position = \{p_x, p_y, p_z\}$$

$$Velocity = \{v_x, v_y, v_z\} = \frac{dPosition}{dt}$$

$$Acceleration = \{a_x, a_y, a_z\} = \frac{d^2Position}{dt^2}$$

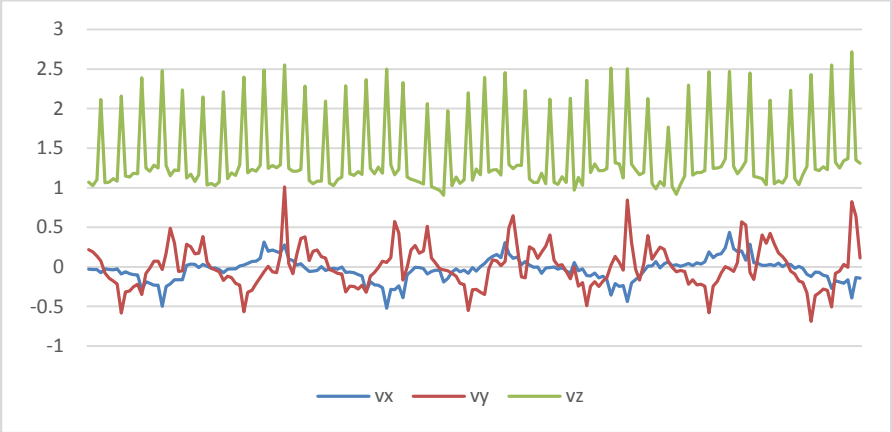


Fig. 6. Typical periodical walking velocity data in several cycles

Since it was difficult for a user to walk precisely along the z-axis (pointing forward) in the coordinate system defined in HoloLens (see Fig. 7), there was always an inevitable angle between the actual walking direction of the user and the z-axis when the user was walking in only one direction. This would result in a cumulative difference between the actual spatial data collected by HoloLens and the ideal spatial data exactly along the z-axis. Therefore, we needed to process the raw spatial data collected from HoloLens to facilitate data labeling based on the features of spatial data for the comparison.

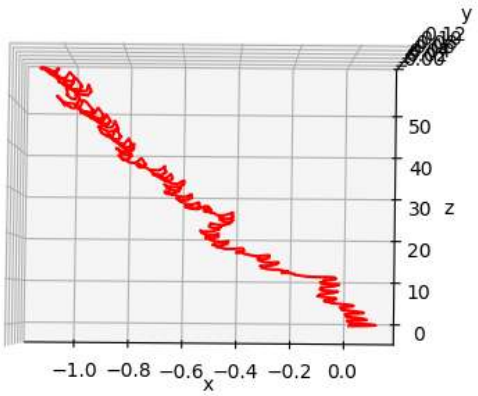


Fig. 7. Actual walking path of the user

To process the spatial data from HoloLens when the user was walking in only one direction (along z-axis, pointing forward), we subtracted the average of x, y, and z axes values of position of 2n+1 spatial data points around each data point (n points before and after each point) to eliminate the accumulation difference from the ideal situation, so as to obtain a stable periodic data in x, y and z axes:

$$p_x{}'_i = p_{x_i} - \frac{\sum_{j=i-n}^{i+n} p_{x_j}}{2n+1}$$

$$p_y{}'_i = p_{y_i} - \frac{\sum_{j=i-n}^{i+n} p_{y_j}}{2n+1}$$

$$p_z{}'_i = p_{z_i} - \frac{\sum_{j=i-n}^{i+n} p_{z_j}}{2n+1}$$

The value of the n affected the accuracy of the generated position p_x', p_y', and p_z'. Also, since these 2n+1 points involve n data points after each data point, it would inevitably bring lags to the newly generated spatial data if it was applied in real scenario. The time of lag was:

$$\text{time}_{lag} = n * \frac{1}{framerate}$$

n	Legend: px average px px′

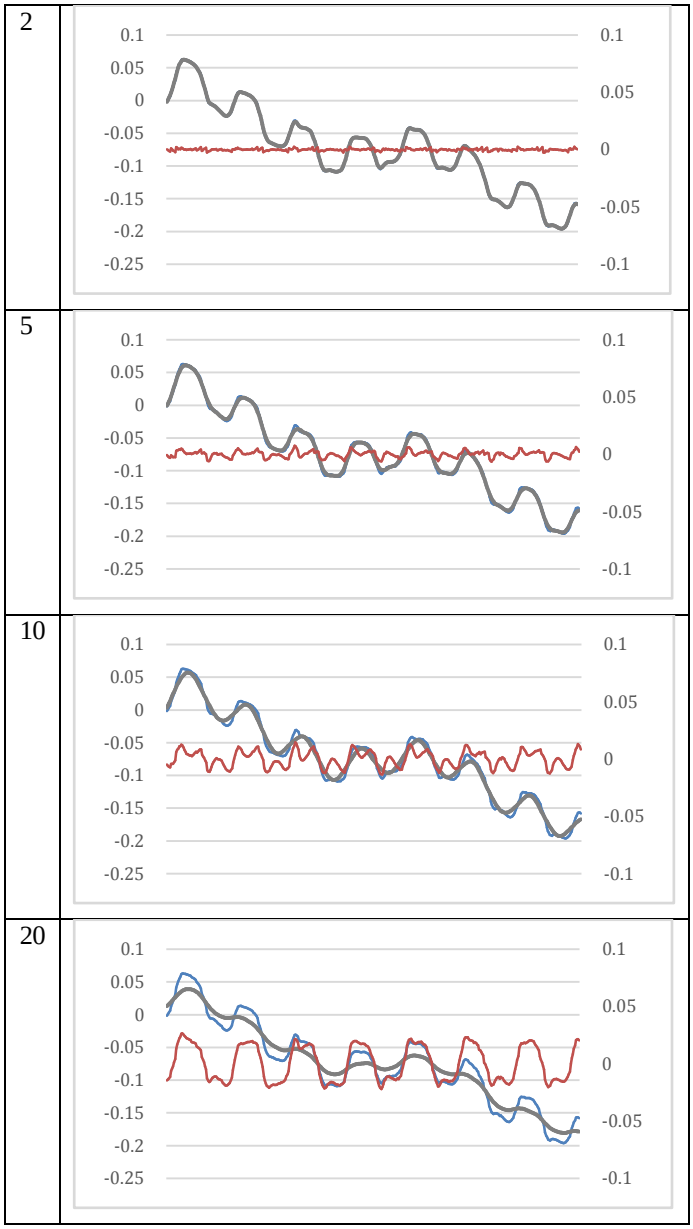


Fig. 8. Comparison between p_x and p_x' with different values of n

Fig. 8 shows how p_x'changed with different values of n. The blue curves and grey dotted curve represented p_x and average of 2n+1 $p_x$respectively; the orange curve, p_x', was the difference between p_x (blue curve) and average of 2n+1 p_x (grey dotted curve). It is clear from Fig. 8 that as n increased, the newly generated p_x' had better similarity with the original p_x. Therefore, we generated p_x', p_y', and p_z' with n=20.

As mentioned before, the frame rate of data acquisition was 50, so when n was 2, 5, 10, and 20, the corresponding lag was 40ms, 100ms, 200ms and 400ms respectively. According to the spatial data collected and the sound recorded, we could know that the period of each walking cycle was about 800ms – 1200ms, and the duration of sound of each footstep was about 8 – 12 frames (160 – 240ms). Therefore, the lag (400ms) was unacceptable in real application and that made this approach (labeling data based on spatial data collected) only applicable to the analysis and comparison in our experiment.

When the user was walking in all directions, the spatial patterns of walking were scattered to both x and z axes, which made it unable to label data with the value of p_x and p_y and therefore impossible for machines to learn the pattern of the walking from spatial information. Therefore, we need to further process the data and normalize the velocity of the user on both x-axis and z-axis. We could get the user's current walking direction $(V_x + V_z)$ based on the average value of m historical velocity vector before each point and set this direction as the new virtual z'-axis:

$$V_{x,i} = \frac{\sum_{j=i-n}^{i-1} v_{x,j}}{n}$$
$$V_{z,i} = \frac{\sum_{j=i-n}^{i-1} v_{z,j}}{n}$$

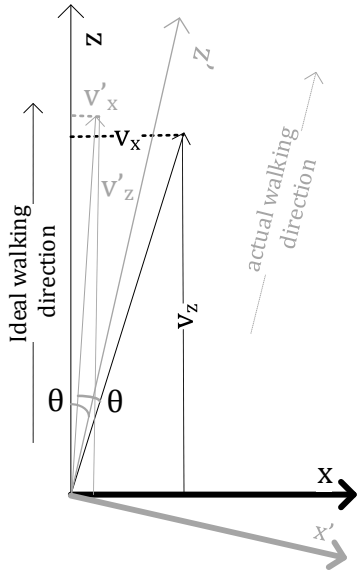


Fig. 9. Normalization of the vector of the velocity

As is shown in Fig. 9, the angle between this new z'-axis and the original z-axis of the coordinate system was θ degrees. Then we rotated the user's current velocity vector $(v_x + v_z)$ $-\theta$ degrees to get a new velocity vector $(v_x^{'} + v_z^{'})$ and the components of the new velocity vector on x-axis $(v_x^{'})$ and z-axis $(v_z^{'})$ could be calculated:

$$v^{'}_{x,i} = cos\theta V_{x,i} - sin\theta V_{z,i}$$
$$v^{'}_{z,i} = sin\theta V_{x,i} + cos\theta V_{z,i}$$

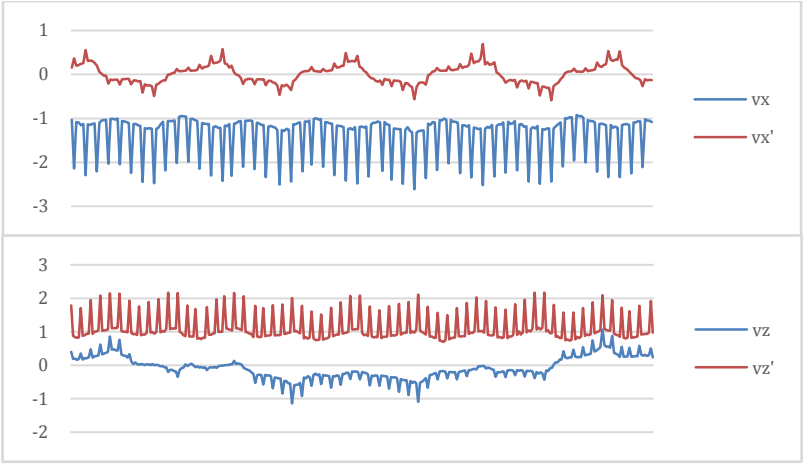


Fig. 10. Comparison between v_x, v_z and normailized $v_x^{'}, v_z^{'}$

Fig. 10 shows the comparison between the original velocity vectors v_x, v_z and the normalized velocity vectors $v_x^{'}, v_z^{'}$. The value of m was 40 and it covered more than the period of one step if the user walked about two steps per second (25 frames for each step). It can be seen from the Fig. 10 that the normalized $v_x^{'}$ and $v_z^{'}$ had the similar waveforms as the typical velocity waveforms shown in Fig. 6.

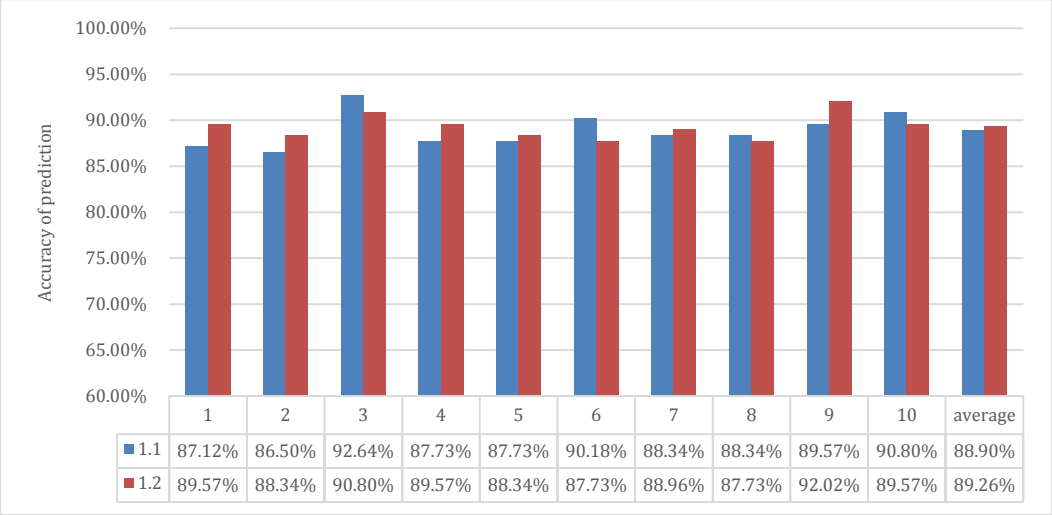
IV. EVALUATION

We collected data from HoloLens fixed on the head of the user when the user walked naturally in only one direction and in different directions with different speeds.

A. User walking in only one direction

In this part, we used part of data labeled based on the visualized analysis of the sound recorded as the training set (training set 1.1) and other part as the test set (test set 1). The size of the training set was 1696 velocity vectors (64 steps) and the size of each test set is 163, total 10 groups of test set. At the same time, as the comparison, we also labeled the data based on the features of spatial information as another training set (training set 1.2). We used support vector machine (SVM) as the supervised machine learning model with the linear kernel and the result is shown below in Table 1.

TABLE I. ACCURACY OF PREDICTION OF GROUP 1: USER WALKING IN ONLY ONE DIRECTION



	1	2	3	4	5	6	7	8	9	10	average
1.1	87.12%	86.50%	92.64%	87.73%	87.73%	90.18%	88.34%	88.34%	89.57%	90.80%	88.90%
1.2	89.57%	88.34%	90.80%	89.57%	88.34%	87.73%	88.96%	87.73%	92.02%	89.57%	89.26%

Training Set 1.1: data labeled based on the sound
Training Set 1.2: data labeled based on the features of spatial information
Test set 1: Data labeled based on the sound

It can be seen from Table 1 that both training sets (1.1 & 1.2) had good accuracy (> 87% in each group and > 88% overall) of prediction when the user was walking in only one direction, compared with the random baseline (33% for three labels: “L”, “R”, and “N”). However, the data labeled based on the features of spatial information (training set 1.2) was not available when the user was walking in all directions and it would bring lags (400ms when n=20). Therefore, only data labeled based on the sound (training set 1.1) was suitable for applications in real scenarios.

B. User walking in all directions

As mentioned before, when the user was walking in different directions, the spatial patterns of walking were scattered to both x and z axes and therefore we needed to normalize the velocity of the user on both x-axis and z-axis in order to obtain periodic spatial information and therefore only normalized data labeled based on the visualized analysis of the sound could be used as the test set in machine learning. One training set was data labeled based on sound of the user walking in all directions (training set 2). The size of the training set 2.1 was 1532 velocity vectors (55 steps). We also used the two training sets (training set 1.1 & 1.2) in section 1 for comparison. There were four test sets: three sets were normalized data of the user walking in all directions in different speeds (test set 2.1: normal speed, 2 steps per second; test set 2.2: approximately 1.25x normal speed; test set 2.2: approximately 1.5x normal speed); and the fourth test set was the same test set in section 1 (test set 1) for comparison. We used support vector machine (SVM) as the supervised machine learning model with the linear kernel and the result is shown below in Table II.

TABLE II. ACCURACY OF PREDICTION OF GROUP 2: USER WALKING IN ALL DIRECTIONS

Test Sets / Training sets	2.1	2.2	2.3	1
2	89.03%	87.33%	88.02%	83.99%
1.1	82.10%	80.05%	81.94%	88.90%
1.2	75.11%	72.35%	73.93%	89.26%

Training set 2: Normalized data labeled based on the sound, user walking in all directions
Training Set 1.1: data labeled based on the sound, user walking in one direction
Training Set 1.2: data labeled based on the features of spatial information, user walking in one direction
Test set 2.1: Normalized data labeled based on the sound, user walking in all directions (Normal walking speed)
Test set 2.2: Normalized data labeled based on the sound, user walking in all directions (approximately 1.25x normal walking speed)
Test set 2.3: Normalized data labeled based on the sound, user walking in all directions (approximately 1.5x normal walking speed)
Test set 1: Data labeled based on the sound, user walking in one direction

It can be seen from the Table II that the accuracies of training set 2 for test sets 2.1, 2.2 and 2.3(user walking in all directions in different speed) were all more than 87%. The accuracies of training set 1.1 for test sets 2.1, 2.2 and 2.3 were still more than 80%, which showed the good consistency of the data characteristics of the normalized data.

C. Further analysis of the result

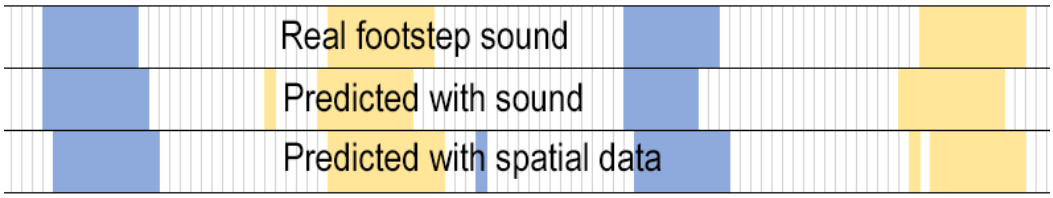


Fig. 11. Comparison between the actual result and expected result of the prediction

We compared the expected output (based on real footstep sound) and actual output (predicted with data labeled based on sound and data labeled based on spatial information) of timing of footsteps in section 1 and 2. A small sample of the result of section 1 is shown in Fig. 11. Each small block represented a frame (20ms) and yellow and blue blocks represented the timing of left foot and right foot touching the ground respectively. The first line was based on the timing of real sound of footsteps recorded; the second and third lines were based on the result of training sets 1.1 and 1.2 in section 1 respectively. We found out that most mismatches between the real timing and the predicted timing were fragmentary with length less or equal than 2 frames, as is shown in Table 3 and Fig. 12. These fragmentary mismatches could trigger false positive. Therefore, in real application, there could be one more step to filter these mismatches to improve the accuracy. If the application thinks it is a new step only when it detects two or three consecutive states, most potential false alarms could be filtered. If the threshold of the filter is set higher, there will be less mistakes, but the lag will also be higher, and a balance should be considered in real application.

TABLE III. DISTRIBUTION OF MISMATCHES WITH DIFFERENT LENGTHS

length of mismatch	1	2	3	4
	34	8	2	0

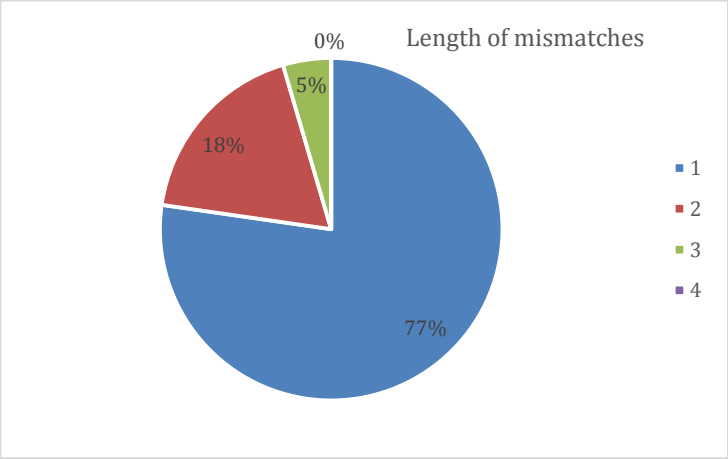


Fig. 12. Propotion of mismatches with different lengths

V. CONCLUSION AND FUTURE WORK

In this paper we analyzed the HoloLens position data to detect when the user was using the left or right foot to touch the ground while he was walking with a wearable AR/MR device. The approach was based on machine learning with the data from the motion sensors of the wearable device and the visualized analysis n of the sound of footsteps recorded. The results of the evaluation section showed that if we used the training set with the normalized data of the user walking in all directions and the data was labeled based on the visualized analysis of the sound recorded, the accuracy of prediction of the left/right foot could be more than 88% for different walking speeds. If we used the training set with data of the user walking in only one direction and the data was also labeled in the same method, the accuracy of prediction of left/right foot could also be more than 80%.

In our experiment, we collected only position data from HoloLens. In the future, we will consider adding rotation data into the training sets. More data will be collected from more people of different demographic groups and more methods of machine learning will be implemented to improve the accuracy and reduce the potential lag.

ACKNOWLEDGEMENTS

The research was financially supported by the Cyberinfrastructure for Network Science Center(CNS) of Indiana University Bloomington. Thank my advisor Professor Katy Börner who gave me a lot of suggestions. Thank Ms. Nitocris Perez who supported me the HoloLens for the data collection. Thank Yue Chen for data collecting.

REFERENCES

Journal Papers:

[1]. Yam, C., Nixon, M. S., & Carter, J. N. (2004). Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, *37*(5), 1057-1072.
[2]. Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, *10*(2), 1154-1175.
[3]. Tao, W., Liu, T., Zheng, R., & Feng, H. (2012). Gait analysis using wearable sensors. *Sensors*, *12*(2), 2255-2283.
[4]. Wu, Y., & Huang, T. S. (1999, March). Vision-based gesture recognition: A review. In *Gesture Workshop* (Vol. 1739, pp. 103-115).
[5]. Sarkar, A. R., Sanyal, G., & Majumder, S. (2013). Hand gesture recognition systems: a survey. *International Journal of Computer Applications*, *71*(15).
[6]. Mohatta, S., Perla, R., Gupta, G., Hassan, E., & Hebbalaguppe, R. (2017, March). Robust Hand Gestural Interaction for Smartphone Based AR/VR Applications. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on* (pp. 330-335). IEEE.
[7]. Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: A survey. Artificial Intelligence Review, 43(1), 1-54.
[8]. Iso, T., & Yamazaki, K. (2006, September). Gait analyzer based on a cell phone with a single three-axis accelerometer. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 141-144). ACM.
[9]. Thang, H. M., Viet, V. Q., Thuc, N. D., & Choi, D. (2012, November). Gait identification using accelerometer on mobile phone. In *Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on* (pp. 344-348). IEEE.
[10]. Juefei-Xu, F., Bhagavatula, C., Jaech, A., Prasad, U., & Savvides, M. (2012, September). Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on* (pp. 8-15). IEEE.
[11]. Klucken, J., Barth, J., Kugler, P., Schlachetzki, J., Henze, T., Marxreiter, F., ... & Winkler, J. (2013). Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease. *PloS one*, *8*(2), e56956.

[12]. Moore, S. T., MacDougall, H. G., Gracies, J. M., Cohen, H. S., & Ondo, W. G. (2007). Long-term monitoring of gait in Parkinson's disease. *Gait & posture*, *26*(2), 200-207.

Yingnan Ju."Footstep Detection with HoloLens. "American Journal Of Engineering Research (AJER), Vol. 7, No. 6, 2018, PP.223-233.