

Article

Estimating Spatiotemporal Information from Behavioral Sensing Data of Wheelchair Users by Machine Learning Technologies

Ikuko Eguchi Yairi ^{1,*}, Hiroki Takahashi ¹, Takumi Watanabe ¹, Kouya Nagamine ¹, Yusuke Fukushima ¹, Yutaka Matsuo ² and Yusuke Iwasawa ²

- ¹ Department of Science and Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554, Japan; hirotaka@yairilab.net (H.T.); watanabe@yairilab.net (T.W.); nagamine@yairilab.net (K.N.); fukushima@yairilab.net (Y.F.)
- ² Department of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; matsuo@weblab.t.u-tokyo.ac.jp (Y.M.); iwasawa@weblab.t.u-tokyo.ac.jp (Y.I.)
- * Correspondence: i.e.yairi@sophia.ac.jp; Tel.: +81-3-3238-3300

Received: 30 January 2019; Accepted: 4 March 2019; Published: 15 March 2019



Abstract: Recent expansion of intelligent gadgets, such as smartphones and smart watches, familiarizes humans with sensing their activities. We have been developing a road accessibility evaluation system inspired by human sensing technologies. This paper introduces our methodology to estimate road accessibility from the three-axis acceleration data obtained by a smart phone attached on a wheelchair seat, such as environmental factors, e.g., curbs and gaps, which directly influence wheelchair bodies, and human factors, e.g., wheelchair users' feelings of tiredness and strain. Our goal is to realize a system that provides the road accessibility visualization services to users by online/offline pattern matching using impersonal models, while gradually learning to improve service accuracy using new data provided by users. As the first step, this paper evaluates features acquired by the DCNN (deep convolutional neural network), which learns the state of the road surface from the data in supervised machine learning techniques. The evaluated results show that the features can capture the difference of the road surface condition in more detail than the label attached by us and are effective as the means for quantitatively expressing the road surface condition. This paper developed and evaluated a prototype system that estimated types of ground surfaces focusing on knowledge extraction and visualization.

Keywords: human sensing; wheelchair; road accessibility; feature extraction; deep learning

1. Introduction

Providing accessibility information of the road for people with difficulties in moving, such as elderly people, mobility impaired people, and visually impaired people, is one of the important social issues. One method of solving these issues using information communication technology is to develop an accessibility map as a large geographic information system to provide the accessibility information [1–3]. The conventional method for gathering accessibility information on a large scale is as follows: a method for experts to evaluate sidewalks and their images for each case [4]; crowdsourcing methods to recruit information from volunteers [5,6]; and so on. In all these methods, human labor is indispensable.

Recent expansion of intelligent gadgets, such as smartphones and smart watches, familiarizes humans with sensing their activities. Focusing on the fact that the observed values of acceleration sensors installed in wheelchairs were influenced by the condition of the road surface, we have been proposing a system which evaluates road surface condition by machine learning from acceleration



sensor data. Research on human action recognition by applying machine learning to raw human behavior data measured by an accelerometer attached to the body is as follows: examining the possibility of various machine learning methods for human action recognition using data sets, such as everyday life and assembly work at a factory [7–9]; aiming at improving learning efficiency by imaging time series data of human actions [10]; distinguishing involuntary body vibrations due to illness from voluntary exercise [11,12]; and so on. This paper estimates the road surface condition using the deep convolutional neural network (hereinafter referred to as DCNN) that is one of the most famous expression learning techniques for developing impersonal models in voice recognition, and analyzes the learned DCNN. Our goal is to realize a system that provides the road accessibility visualization services to users by pattern matching using impersonal models, while gradually learning to improve service accuracy using new data provided by users. As the first step, this paper evaluates features acquired by the DCNN, which learned the state of the road surface from the data in supervised machine learning techniques. There are no other studies investigating and verifying that feature quantities extracted from the learned DCNN capture a more detailed road surface condition than the given road surface condition label.

In the rest of this paper, the outline of the proposed system is shown in Section 2. In Section 3, there is a preliminary analysis of the wheelchair sensing data, before the machine learning is performed, to clarify the relationship between the barrier of the sidewalk and the vibration value, and the relationship between the physical burden of the wheelchair user and the vibration value. Section 4 describes collection of wheelchair sensing data, assignment of learning labels, and data classification learning by the DCNN. In Section 5, analysis of the relationship between the data and the reaction pattern of the learned DCNN is conducted, and the analysis result is reported. Section 6 discusses future tasks, and Section 7 concludes this paper.

2. System for Estimation of Road Accessibilities

This section introduces the authors' proposed system for providing road accessibility information which are helpful for all walkers, especially for wheelchair users. Figure 1 shows the outline of the proposed system. In the proposed system, vibration waveforms during movement are collected by an acceleration sensor installed in a wheelchair. After extracting road surface information from vibration waveforms using machine learning, the extracted road surface information is accumulated and visualized on a map.



Figure 1. Outline drawing of road surface feature automatic evaluation system by wheelchair sensing and machine learning.

The simplest type of accessibility visualization utilizing human sensing is simple wheelchair trails [13]. Such trails provide practical information for wheelchair users regarding wheelchair accessible roads and facilities. The information is useful but not sufficient. The trail approach provides the fact that someone could pass there, but not all wheelchair users can pass in the same way. The physical abilities of wheelchair users are more diverse than generally imagined; there are users like Paralympic athletes, and some users may damage their bodies with only a few vibrations. Important information for wheelchair users is the physical state of the road surface, such as the angle

of the slope of the sidewalk, height of curb, roughness of the road surface, and so on. This information about the physical state of the road surface is not only helpful for wheelchair users to make decisions of access/avoidance of the roads according to their physical conditions and abilities, but also all people with difficulties in moving in the same way. The information about the physical state of the road is the foundation of the road accessibilities. This paper hereinafter calls this information the road accessibility information. The purpose of this paper is to propose a system for providing this road accessibility information by human sensing and machine learning techniques.

Vibration waveforms from acceleration sensors are reasonable in detecting the road accessibility information because they are influenced by the state of the road surface. Extracting useful information from noisy vibration waveforms of accelerometers mounted on various wheelchairs is not impossible because of the recent success of developing impersonal models in pattern matching tasks by the DCNN. Our final goal is to realize the system that provides road accessibility visualization services to users by online/offline pattern matching using impersonal models, while gradually learning to improve service accuracy using new data provided by users. As the wheelchair traveling data of more places is gathered by more users after the service is launched, the model for pattern matching is incrementally strengthened. Along with maturity of the model, it will also be possible to extract the road accessibility information from the running data of baby strollers and bicycles in the same way as wheelchairs. This paper aims to establish a fundamental method of information extraction from wheelchair running data using machine learning.

The mobility support information system for people with difficulties in moving has been proposed as the following examples.

- Creating walking space network data composed of links with information such as width, step, crossing gradient of walking route, etc., and nodes connecting links with latitude/longitude information [14–16];
- Collecting walking space network data by on-site survey by community [17];
- Providing information on the road surface condition to users' mobile terminals [18,19].

All of these have the problem that it is difficult to collect wide-area road surface information because it costs a huge amount of human power. On the other hand, the following has been tried as a method for evaluating the road surface condition by automatic processing.

- Using high-resolution satellite images to classify land cover according to the physical condition of the ground surface, such as agricultural land, grazing land, barren areas [20–22];
- Detecting depression of a road surface using a method based on machine learning which is based on acceleration sensor data and GPS data installed in cars [23,24];
- Detecting recesses and abnormal traffic conditions on the road surface by acceleration sensor data, GPS data, voice data, etc. of smartphones for cities [25,26].

In any of these, it is difficult to obtain a detailed road surface condition required for mobility impaired people using wheelchairs. Therefore, we have aimed at developing a method of automatic extracting of detailed information on the road surface by machine learning from acceleration sensor data installed in a wheelchair. It is not easy to extract the influence of the condition of the road surface from the raw data of the acceleration sensor [27,28]. It is important to convert from the observed acceleration sensor data to indexes representing the state of the road surface by expression learning techniques.

3. Preliminary Analysis

3.1. Road Barrier and Vibration

To analyze the relationship between physical vibration and subjective feeling of the road barrier, we employed a vibration acceleration level (VAL) with decibel units (dB) that was defined as

 $20\log_{10}(a/a0)$, where a and a0 indicate the root-mean-square of three-axis acceleration values and a reference acceleration, respectively. a0 is usually set to 10^{-5} m/s² by the Japanese Industrial Standards (JIS) C1510-1995.

Nine participants with mobility impairments conducted an experiment of wheelchairs moving around Akihabara Station in Tokyo. In order to obtain pure movement data, each participant was asked to move their own wheelchair which is being used in their everyday life, and the vibration from the wheelchair moving was obtained from a Sun SPOT three-axis accelerometer of Sun Microsystems attached on the axle of the rear wheels. The moving vibration data was sent to a base station from the accelerometer and the location data obtained from a GPS receiver were automatically recorded into the laptop contained in the backpack attached to the back of each wheelchair. The experiment was recorded by video taking, and each subject was asked his/her personal feelings about road barriers through an interviewer.

To find a higher VAL on the route intuitively, we separated the moving route represented on the Google map by every 0.02 s, and each part of the route was painted with one of 13 colors according to the averaged VAL value, i.e., red color for a VAL value over 116 dB. Notice that the location data obtained from the GPS receiver was revised by hand carefully to remove the multipath effect.

Figure 2 shows the result of a 50 year old male participant who had about 40 years' experience in using a wheelchair. He has been using it both inside and outside all the time. In contrast, Figure 3 shows a whole picture of the visualized VAL of a 30 year old male participant who had about 10 years' experience in using a wheelchair. Since he usually sails a kayak, a human-powered light narrow boat that has one or more covered cockpits, he enjoyed the wheelchair moving even when the body vibration turned out to be so high. From those results, it can be observed that the road barrier from a wheelchair moving does not only come from the vibration level like VAL, but also from vibration pattern changes.



Figure 2. A 50 years old male participant with a rheumatoid and paralyzed arm moves his manual wheelchair. Although he tried to control his wheelchair carefully throughout this experiment, higher VAL values were detected at uneven road surfaces.



Figure 3. A 30 years old male participant moved his wheelchair so fast, and higher VAL values were often detected. Regardless, he enjoyed his wheelchair moving.

3.2. Physical Burden and Vibration

It is known that the road with low accessibility increases the physical burden of the manual wheelchair. In this section, we analyzed the relationship between the physical burden and vibration of the manual wheelchair user. The authors have proposed a method to analyze and quantify the relationship between the acceleration change caused by thrust when pushing the wheelchair by hand and the physical burden during manual wheelchair traveling from the acceleration sensor mounted on the smartphone [29]. Evaluation of the burden has been studied mainly in the field of sports and healthcare, and it has been utilized [30–32]. NASA-TLX (NASA Task Load Index) [33] and the SWAT (Subjective Workload Assessment Technique) [34] are the evaluation indexes using a psychological scale. Representative examples of evaluation methods using biometrics include the evaluation of muscle fatigue using electromyograms [35], and the evaluation of exercise intensity and psychological stress by heart rate [36,37].

The physical burden of manual wheelchair driving is influenced by the magnitude of thrust which is the load of the rowing action. Since the force and the acceleration are proportional to each other from Newton's equation of motion, the acceleration waveform of the wheelchair changes depending on the degree of burden during wheelchair traveling. We hereinafter refer to the action while the hands are in contact with the handrim of the manual wheelchair as "padding". Focusing on the peak to peak value (hereinafter P-P value) of the waveform at the time of padding, correlation with heart rate was confirmed through experiments.

A total of eleven persons, including five manual wheelchair users and six wheelchair inexperienced users, wore an iPhone and a heartbeat sensor and ran a preselected route. As a result, wheelchair running data consisting of 14,189 rowing actions for a total of 4.5 h was obtained. Routes were selected from four types near Ichigaya station in Tokyo (route A is about 2200 m, route B, C and D is about 1350 ± 50 m). The iPhone was installed under the seat of the wheelchair and the acceleration data on the direction of traveling axis and the GPS data were acquired at a 50 Hz sampling rate. Since acceleration is susceptible to noise, a value smoothed by a simple moving average of 10 points before and after was used. A Poral H7 heartbeat monitor was used for heartbeat sensing, and data was acquired in RRI (the R wave-to-R wave (RR) interval) format at each beat. The heartbeat data was converted to bpm format after artifacts and noise were removed.

Two types of sample extraction methods with rowing and padding, and five kinds of feature amounts of P-P value, maximum value, minimum value, average value, and time width were combined to create 10 pattern data sets. As a result of comparing correlation coefficients between these data sets and heart rate data, the maximum value was 0.78 (P < 0.001) of the padding and P-P values. Figure 4 shows the smoothed row vibration P-P value and heart rate of the participant having the highest

correlation value of 0.92. The two data were normalized with minimum value 0, maximum value 1. It was confirmed that the P-P value of the acceleration data of the manual wheelchair was highly correlated with the burden of padding, that is, the road surface barrier.



Figure 4. Comparison between normalized P-P values and heart rates at padding of the participant, having the highest correlation value of 0.92.

The difference when comparing the heartbeat and the vibration value by the method as shown in Figure 4 is hereinafter referred to as an error value. As the result of plotting the error value on the map, it was found that a difference of 0.33 or more (heart rate > vibration value) occurred at the intersection, the point where the inclination at right angles to the traveling direction was large, and the place where the road width narrowed. As confirmed by video, these were all cases when the user operated the manual wheelchair with one hand. The vibration pattern of one hand operation is different from the case of two handed operation. By using the pattern classification approach of the vibration waveform instead of the P-P value of the vibration value, it could be detected that the state of the road surface was different from the usual state.

4. Road Accessibility Estimation

In this section, we describe the method and results for evaluating the road surface condition by the DCNN through supervised learning. The data acquisition method, the labeled data set creation method, the DCNN structure, and the classification accuracy of the road surface condition are explained step by step.

4.1. Road Sensing and Labeling for Supervised Learning

A total of nine wheelchair users, including six manual wheelchair users and three electric wheelchair users, participated in the experiment. Their actions when traveling about 1.4 km of specified route around Yotsuya station in Tokyo were measured by the acceleration sensor (iPod touch) installed in the lower part of the wheelchair seat, and positioning data of the Quasi-Zenith Satellite System (QZSS) was added. Acceleration values in the *x*, *y*, and *z* axes of the acceleration sensor were sampled at 50 Hz, and a total of 1,425,798 samples (about 8 h) were obtained. In order to confirm the situation where the acceleration sample was acquired, the video image of the participant's wheelchair running and of the running route, there are many flat sidewalks, which are not necessarily smooth. If the user did not have a problem in raising and lowering wheelchair ramp slopes, it was thought that excessive burden on the body and risk of accident had less chance of occurring on the route.

The data set which consists of the acceleration values and four types of labels—slope/curb/tactile block/other—was created by checking the state of the road surface and participants from all of the captured video images during experiments. Each of these four labels represents an important road surface feature: in order, a continuous gradient, an abrupt step, a continuous unevenness. In order to obtain sufficient classification learning accuracy, at least the data ratio of any label needed to be 1% or more of the total number of data. These labeled data were the number of points and the ratio of data sufficient for classification learning. The three-axis acceleration data was sliced into 7016 pieces by

a sliding window method with a window width of 400 samples (about 8 s) and an overlapping rate of 0.5 and labeled. This window width and overlap rate were determined by searching for suitable values by research using the same data [38].

4.2. Machine Learning and Estimation Results

The structure of the DCNN used in this paper is shown in Figure 5. This network was composed of six layers of input layer, three convolution layers, a fully connected layer, and an output layer. By using the hierarchical structured network to learn functions from input to output, it was possible to proceed simultaneous learning of the feature extractor h and the classifier f that are effective for data classification. In the convolution processing part, following the representative research [11], two processes of convolution and max pooling were used. Here, w in the figure was a weight to be convoluted with respect to the input data, which was a parameter learned to successfully map the input to the output. Feature Map (N) in the figure means that different kinds of different weights were learned to obtain N types of different outputs, respectively. As another setting, Adam was used for learning of the neural network, and the learning rate was set to 0.0001. Rectified linear units were used for the activation function, except for the softmax function, which was used in the output layer.



Figure 5. The structure of the deep convolutional neural network (DCNN) used in this paper.

Table 1 compares the existing method with the DCNN in Figure 4 for the accuracy of the road surface classification task based on the four kinds of labels. Here, each table's value is the average value of nine trials obtained by the leave-one-subject-out method [39]. The leave-one-subject-out method is an evaluation method that can verify the recognition accuracy for unknown users by evaluating with user data not included in the learning data. In this paper, we repeatedly evaluated the model trained with the data set of wheelchair users for eight people with the data set of the remaining one person. Mean F-score and accuracy of each class were used as the evaluation index. This classification problem is class imbalance. A high average F value means that it is well recognizable for barriers with few spots appearing in the data set. Details of the method used for comparison are as follows.

- **Raw:** The data set for nine people was classified by the K-nearest neighbor method (KNN). Input was a one-dimensional vector obtained by simply combining three-axis acceleration values for 400 windows. 1, 5, 10, 15, 20, 25, and 30 were used for the value of K, but Table 1 shows the result of K = 1, which was the most accurate.
- **MV:** As the same as Raw, KNN was used and Table 1 is the result of K = 1. The difference was that the average value and standard deviation of each axis of *x*, *y*, *z* of the three-axis acceleration for 400 windows was applied as six meta value inputs.
- **SVM:** The data set was classified by a support vector machine (SVM). The same input as MV was used. A radial bias function was used as a kernel. A grid search was performed for γ (10⁻¹²-10²) and C (10⁻⁷-10⁷), and the value with the highest accuracy in the verification data was taken as the hyper parameter.
- **Heuri:** The same classifier as the SVM was used. In addition to the six inputs used in SVM, the maximum, minimum, zero crossing, average crossing, average of difference, standard deviation of difference, maximum of difference, minimum of difference, FFT frequency

component, the intensity of the 0 spectrum of the FFT frequency component, energy, and entropy was also used for input. In addition, correlation for each axis pair of x, y, z and maximum value of cross-correlation (correlation coefficient in time shift with the highest correlation coefficient taking into consideration the time shift) were also calculated and used.

• ECDF: We used the support vector machine as a classifier, and the empirical cumulative distribution function (ECDF) as an input. The ECDF has advantages compared to other feature extraction methods, such as FFT, PCA, statistical quantity, etc., in multiple tasks [40]. For the number of interpolation points, which was a hyper parameter, it has been reported that the sensitivity is low, so it was fixed to 10 in the experiment.

Table 1. Performance comparison between conventional methods and the DCNN in supervised learning. MV: statistical meta value of raw data; SVM: support vector machine; ECDF: empirical cumulative distribution function.

Method	Raw	MV	SVM	Heuri	ECDF	DCNN
Mean F-score	0.22	0.58	0.56	0.63	0.65	0.71
Accuracy	0.36	0.79	0.74	0.83	0.83	0.88

As Table 1 shows, the method using the DCNN had higher accuracy than the conventional method. As the result of the classification accuracy of the DCNN was high in the road condition classification task using wheelchair traveling data, our proposed method is reasonable and practical.

5. Analysis of the Learned DCNN

5.1. Analysis Overview

In this section, it is investigated what type of road surface features the DCNN has learned. The DCNN seems to capture more detailed characteristics of the road surface as the output pattern of all the fully connected layers of 400 units was divided into a plurality of clusters. The analysis method is explained in order of Step 1 to Step 5 below.

Step 1: Acquire the DCNN output pattern

As shown in the previous section, 400 units of output pattern of the fully connected layer was acquired by creating a DCNN model trained with eight data sets and inputting the remaining one data set as test data. "Feature amount" is hereinafter used as the term referring to the output pattern of these 400 units.

Step 2: Clustering of feature amounts

Clustering of feature amounts for each data set of nine participants was done. K-means was used as a clustering method.

Step 3: Visualize on the map

The clustering results were analyzed by determining the color of data for each cluster and visualizing the traveling points where individual data were obtained on the map.

Step 4: Determining the optimal number of clusters

It was confirmed that all participants had a case where feature amounts for uplink slope and down slope data were classified into different clusters. In such a case, it was considered that there was a high possibility that the features of the detailed road surface were accurately captured even in other clusters. In this paper, the number of clusters at that time was taken as the optimum number of clusters. Table 2 shows the optimum number of clusters for all participants (E1–3: electric wheelchair user, M1–6: manual wheelchair user).

Step 5: Investigation of correspondence between feature amounts and road condition

Moving pictures taken during wheelchair running were analyzed and the relation between the feature amounts of clusters and the corresponding road surface condition was examined.

Table 2. The optimum number of clusters for all participants (E1–3: electric wheelchair user, M1–6: manual wheelchair user).

Participants	E1	E2	E3	M1	M2	M3	M4	M5	M6
Optimum number of clusters	14	13	9	8	11	11	9	11	10

5.2. Results

Figure 6 shows the feature amount clustering result example when E1, having the highest classification accuracy of the DCNN, was used as test data. When visualizing on the map, the position information corrected based on the moving image was used. As shown in Figure 6, each user was traveling three times on the route, and on the second lap, the route was traveled in the opposite direction between the first and third laps.



Figure 6. The clustering result of the DCNN feature output of E1's acceleration data in 14 clusters.

(a) Slope

Position (2) in Figure 6 is the place where the slope label was given. All up and down slopes were divided into different clusters with all participants. It is thought that the DCNN caught the difference in acceleration in the case of the up and down direction driving; the difference in the number of times of rowing and padding duration in the case of manual wheelchairs. Because the distance is short, and the slope is gentle, position (3) in Figure 6 is the point where the slope labels were not given beforehand. Seven of nine people were included in the same cluster as the uphill slope when this point was traveling in a direction going up the slope. In the case of E1 with the highest classification accuracy, it was classified into independent clusters. This suggests that the DCNN gains a more detailed characterization of the road surface than the labels.

(b) Curb

Depending on participants and places, curbstones showed a difference in classification accuracy. In case of E1, the curb at position (1) in Figure 6 was classified into three clusters. With this data, it is considered that the reason why the curbs were clustered in detail is that each has a characteristic of vibration when passing through the curb stone, and the DCNN can classify the minute difference of the vibration.

(c) Tactile block

Points across the tactile block were classified into various clusters. The data labeled with haptic blocks was rich in diversity. Because the tactile block is installed just before the pedestrian crossing, there is a curb or slope in the vicinity, so the data is likely to be affected by them. Also, the contact condition with the block, such as the direction of the wheel or the one wheel/two wheels, and the difference in the contact time can also affect the data. However, the point including the lateral cutting of the tactile block and the point before and after the haptic block were classified into different clusters. It was found that the characteristics of the data including and excluding the haptic block were correctly captured by the DCNN.

6. Future Tasks

In the previous section, we confirmed that the learned DCNN had acquired the characteristics of the road surface more than the labeled. In this section, we discuss the similarity of the road surface conditions of these points while confirming by images, using the method of searching points where the Euclidean distances of the reaction pattern of the DCNN were close.

The point used as a similar data search input is hereinafter called "Query" in this paper. For the arbitrary points, the top five places where the Euclidean distances of the reaction pattern of the DCNN were close were extracted as similar sites with Query. Using the raw data of acceleration as the comparison target, the extraction of the top five places where the Euclidean distances were similar was performed in the same way.

Figure 7 shows a similar road surface search result with three queries; uphill slope, curbstone, and a place crossing a tactile block. For the color of the frame of the image, black was used when having the same road feature, as Query was readable from the image, and red was used if it cannot be read that it had the same road feature. In an uphill search example, in the case of the raw data of the acceleration, only the fourth rank among the top five corresponds to uphill slope, and in the case of using the feature quantity of the DCNN, all the top five corresponded to uphill slope. In a curbstone retrieval example, when using the feature quantity of the DCNN, all the top five corresponds to the first and third rank, and in the case of using the feature quantity of the DCNN, all the top five corresponds to the acceleration, it was judged to be similar to the data of the point where the wheelchair was stopped. In the case of using the feature quantity of the DCNN, only the first place among the top five cases corresponded to the place where the tactile block was crossed. The misjudged second to fourth were roads with continuous unevenness composed of small tiles. It can also be said that these were the results that the similarity with the feature of Query, when traveling on the tactile block, was correctly captured.

By analyzing the similar road surface search in addition to the clustering analysis of the feature amount, it was shown that the DCNN acquired more detailed road surface conditions than the four types of labels given in advance for classification learning. Through these analysis results, this paper confirms the possibility that the reaction pattern of the fully connected layer of the learned DCNN is effective as a means of quantifying the road surface condition as the feature quantity from the acceleration sensor data during the wheelchair running.

On the other hand, it was found that the classification accuracy of the road surface condition varied in the feature quantity obtained for each wheelchair user. In general, inter-class diversity due to users' differences is recognized as one of the main problems for developing human behavior recognition models [41]; how to deal with that problem is an important research topic. The simplest solution to this problem is to develop a large-scale human behavior database. Kawaguchi et al. proposed the concept of collecting a human behavior corpus for understanding real world activities [42]. A large-scale behavior corpus is important not only for developing individual estimation models but also for understanding personal characteristics, and it is technically possible to develop impersonal models applicable to all users. Also, in this research, by collecting and analyzing more data sets including the new environment, it was possible to reduce variations in feature quantities obtained for each user and variations in classification accuracy of road surface conditions. We will collect and analyze more data sets including new environments and qualitatively evaluate the acquired features and propose a method to acquire more accurate feature representation that can reduce the influence of each user.



Figure 7. Similar road surface search results using DCNN feature output obtained by inputting acceleration data and raw acceleration data.

7. Conclusions

For providing road accessibility information which is helpful for all walkers, especially for wheelchair users, this paper proposed the system and the method of extracting detailed information on the road surface by machine learning from acceleration sensor data installed in a wheelchair. From preliminary analysis of the relationship between the barrier of the sidewalk and the vibration value and of the relationship between the physical burden of the wheelchair user and the vibration value, the characteristics of the wheelchair sensing data were identified and shown. This paper acquired traveling data of nine wheelchair users, labeled them, classified them by the DCNN, and confirmed their accuracy. Focusing on the reaction pattern of the learned DCNN for data, cluster analysis and similar road surface search were performed. As the result, it was confirmed that the DCNN learned more detailed road surface conditions than the four types of labels preliminarily attached for learning by the DCNN as the feature quantity. The reaction pattern of the fully connected layer of the learned DCNN was shown to be effective as a means of quantifying the road surface condition as the feature quantity from the acceleration sensor data during the wheelchair running.

The first contribution of this paper was to demonstrate the fundamental method of the road accessibility information extraction from wheelchair running data using machine learning with real data of wheelchair users. To develop impersonal models applicable to all users, we will examine new methods to acquire more accurate feature representation that can reduce the influence of each user

by collecting and analyzing more data sets including the new environment. The second contribution of this paper was to confirm the possibility of reducing the labeling cost of supervised learning by showing that the learned feature quantity acquired the road features in more detail than the teacher data. The road surface features extracted from the learned network were able to be utilized as the model for pattern matching. The method of weak supervised leaning by using a simple teacher label that can be automatically assigned without human power will be applied in the next step.

Author Contributions: Conceptualization, I.E.Y. and Y.I.; Funding acquisition, I.E.Y.; Investigation, I.E.Y., H.T., T.W., K.N., Y.F., Y.M., and Y.I.; Methodology, I.E.Y., H.T., K.N., Y.F., Y.M., and Y.I.; Project administration, I.E.Y., Y.F., and Y.I.; Software, H.T., K.N., Y.F., and Y.I.; Writing—original draft, I.E.Y., H.T., T.W., K.N., and Y.F.; Writing—review & editing, I.E.Y.

Funding: This research was funded by Research Grant of Tateishi Science and Technology Foundation in FY2011—12, Research Grant from Chiyoda ward of Tokyo named "Chiyodagaku" in FY2014—16, and Grant-in-Aid for Scientific Research(B), 17H01946, Japan Society for the Promotion of Science.

Acknowledgments: We would like to show our best gratitude to all participants who helped our sensing data collection and data labeling.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, or in the decision to publish the results.

References

- 1. Laakso, M.; Sarjakoski, T.; Sarjakoski, L.T. Improving accessibility information in pedestrian maps and databases. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2011**, *46*, 101–108. [CrossRef]
- 2. Matthews, H.; Beale, L.; Picton, P.; Briggs, D. Modelling access with GIS in urban systems (MAGUS): Capturing the experiences of wheelchair users. *Area* **2003**, *35*, 34–45. [CrossRef]
- 3. Karimi, H.A.; Zhang, L.; Benner, J.G. Personalized accessibility map (PAM): A novel assisted wayfinding approach for people with disabilities. *Ann. GIS* **2014**, *20*, 99–108. [CrossRef]
- 4. Ponsard, C.; Snoeck, V. Objective accessibility assessment of public infrastructures. In Proceedings of the 10th International Conference on Computers Helping People with Special Needs, Linz, Austria, 11–13 July 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 314–321.
- Hara, K. Scalable methods to collect and visualize sidewalk accessibility data for people with mobility impairments. In Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology, Honolulu, HI, USA, 5–8 October 2014; ACM: New York, NY, USA, 2014; pp. 1–4. [CrossRef]
- Cardonha, C.; Gallo, D.; Avegliano, P.; Herrmann, R.; Koch, F.; Borger, S. A crowdsourcing platform for the construction of accessibility maps. In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, Rio de Janeiro, Brazil, 13–15 May 2013; ACM: New York, NY, USA, 2013; p. 26. [CrossRef]
- Plötz, T.; Hammerla, N.Y.; Olivier, P. Feature learning for activity recognition in ubiquitous computing. In Proceedings of the International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; Volume 22, pp. 1729–1734. [CrossRef]
- Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional neural networks for human activity recognition using mobile sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, TX, USA, 6–7 November 2014; pp. 197–205. [CrossRef]
- Yang, J.B.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, 25–31 July 2015; pp. 25–31.
- Jiang, W.; Yin, Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; ACM: New York, NY, USA, 2015; pp. 1307–1310. [CrossRef]
- 11. Rad, N.M.; Furlanello, C. Applying deep learning to stereotypical motor movement detection in autism spectrum disorders. In Proceedings of the IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016; pp. 1235–1242. [CrossRef]

- Zhang, A.; Cebulla, A.; Panev, S.; Hodgins, J.; De la Torre, F. Weakly-supervised Learning for Parkinson's Disease Tremor Detection. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; pp. 143–147. [CrossRef]
- 13. Mora, H.; Gilart-Iglesias, V.; Pérez-del Hoyo, R.; Andújar-Montoya, M.D. A Comprehensive System for Monitoring Urban Accessibility in Smart Cities. *Sensors* **2017**, *17*, 1834. [CrossRef] [PubMed]
- 14. Yairi, E.I.; Igi, S. Geographic Information System for Pedestrian Navigation with Areas and Routes Accessibility. *J. Inf. Process.* **2005**, *46*, 2940–2951. (In Japanese)
- 15. Development Specification for Spatial Network Model for Pedestrians. Available online: http://www.mlit. go.jp/common/001244374.pdf (accessed on 11 January 2019). (In Japanese)
- 16. Zimmermann-Janschitz, S. Geographic Information Systems in the context of disabilities. *J. Access. Des. All* **2018**, *8*, 161–192. [CrossRef]
- 17. Yairi, E.I.; Nara, H.; Igi, S. The GIS for improving accessibility of routes and areas, and its practical use by community of residents. *Trans. Hum. Interface Soc.* **2005**, *7*, 463–475. (In Japanese)
- 18. Yairi, E.I.; Igi, S. Research on Ubiquitous System for Mobility Support of the Elderly and Disabled and Its Technology Transfer to Industry. *J. Inf. Process.* **2007**, *48*, 770–779. (In Japanese)
- 19. Koga, M.; Izumi, S.; Matsubara, S.; Morishita, K.; Yoshioka, D. Development and verification of navigation system to support wheelchair user activity in urban areas. *IADIS Int. J. WWW/Internet* **2015**, *13*, 43–56.
- 20. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–181. [CrossRef]
- 21. Kuo, T.S.; Tseng, K.S.; Yan, J.W.; Liu, Y.C.; Frank Wang, Y.C. Deep aggregation net for land cover classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018. [CrossRef]
- 22. Rakhlin, A.; Davydow, A.; Nikolenko, S. Land Cover Classification from Satellite Imagery With U-Net and Lovász-Softmax Loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 262–266. [CrossRef]
- Eriksson, J.; Girod, L.; Hull, B.; Newton, R.; Madden, S.; Balakrishnan, H. The pothole patrol: Using a mobile sensor network for road surface monitoring. In Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services, Breckenridge, CO, USA, 17–20 June 2008; ACM: New York, NY, USA, 2008; pp. 29–39. [CrossRef]
- 24. Allouch, A.; Koubâa, A.; Abbes, T.; Ammar, A. Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. *IEEE Sens. J.* 2017, *17*, 4231–4238. [CrossRef]
- Mohan, P.; Padmanabhan, V.N.; Ramjee, R. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In Proceedings of the 6th ACM International Conference on Embedded Networked Sensor Systems, Raleigh, NC, USA, 5–7 November 2008; ACM: New York, NY, USA, 2008; pp. 323–336. [CrossRef]
- Yu, J.; Zhu, H.; Han, H.; Chen, Y.J.; Yang, J.; Zhu, Y.; Chen, Z.; Xue, G.; Li, M. Senspeed: Sensing driving conditions to estimate vehicle speed in urban environments. *IEEE Trans. Mob. Comput.* 2016, 15, 202–216. [CrossRef]
- 27. Lara, O.D.; Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **2012**, *15*, 1192–1209. [CrossRef]
- 28. Liu, H.; Taniguchi, T.; Tanaka, Y.; Takenaka, K.; Bando, T. Visualization of driving behavior based on hidden feature extraction by using deep learning. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2477–2489. [CrossRef]
- 29. Nagamine, K.; Iwasawa, Y.; Yutaka, M.; Yairi, E.I. Physical Strain Evaluation of Manual Wheelchair Driving by Accelerometer. *IEICE Trans. Inf. Syst.* **2017**, *J100-D*, 773–782. [CrossRef]
- 30. Laporte, R.E.; Montoye, H.J.; Caspersen, C.J. Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Rep.* **1985**, *100*, 131. [PubMed]
- 31. Achten, J.; Jeukendrup, A.E. Heart rate monitoring. Sports Med. 2003, 33, 517–538. [CrossRef] [PubMed]
- 32. Sallis, J.F.; Saelens, B.E. Assessment of physical activity by self-report: Status, limitations, and future directions. *Res. Q. Exerc. Sport* 2000, *71*, 1–14. [CrossRef] [PubMed]

- 33. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183. [CrossRef]
- 34. Reid, G.B.; Nygren, T.E. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Adv. Psychol.* **1988**, *52*, 185–218. [CrossRef]
- 35. De Luca, C.J. Use of the surface EMG signal for performance evaluation of back muscles. *Muscle Nerve* **1993**, *16*, 210–216. [CrossRef] [PubMed]
- 36. Karvonen, M. The effects of training on heart rate; a longitudinal study. *Ann. Med. Exp. Biol. Fenn.* **1957**, *35*, 307–315. [PubMed]
- 37. Mulder, L.; Veldman, J.; Van derVeen, F.; Van Roon, A.; Rudel, H.; Schachinger, H.; Mulder, G. On the effects of mental task performance on heart rate, blood pressure and its variability measures. *Blood Press. Heart Rate Var.* **1993**, *4*, 153–166.
- 38. Iwasawa, Y.; Yairi, I.E.; Matsuo, Y. Combining human action sensing of wheelchair users and machine learning for autonomous accessibility data collection. *IEICE Trans. Inf. Syst.* **2016**, *E99-D*, 115–124. [CrossRef]
- 39. Esterman, M.; Tamber-Rosenau, B.J.; Chiu, Y.C.; Yantis, S. Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage* **2010**, *50*, 572–576. [CrossRef] [PubMed]
- Hammerla, N.Y.; Kirkham, R.; Andras, P.; Ploetz, T. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In Proceedings of the 2013 International Symposium on Wearable Computers, Zurich, Switzerland, 8–12 September 2013; ACM: New York, NY, USA, 2013; pp. 65–68. [CrossRef]
- 41. Bulling, B.S.; Blanke, U. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 33. [CrossRef]
- 42. Kawaguchi, N.; Ogawa, N.; Iwasaki, Y.; Kaji, K.; Terada, T.; Murao, K.; Inoue, S.; Kawahara, Y.; Sumi, Y.; Nishio, N. Hasc challenge: Gathering large scale human activity corpus for the real-world activity understandings. In Proceedings of the 2nd augmented human International Conference, Tokyo, Japan, 13 March 2011; p. 27.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).