Obstacle Recognition using a Convolution Neural Network

In Support of Persons with a Visual Impairment

Tomotaka FUJII and Hideaki KATAYAMA*

National Institute of Technology, Maizuru College, 234 Shiroya, Maizuru, Kyoto 625 - 8511, Japan

* katayama@maizuru-ct.ac.jp

Abstract

According to a survey of visually impaired people by The Japan Federation of the Blind, half of the respondents, have collided with still objects, such as guard rails and poles. In order to reduce the occurrence of this problem, we propose the use of navigation software, which helps the visually impaired avoid obstacles. As a first step, obstacles must be recognized from images taken with a camera worn on the bodies of the visually impaired. We focus on the recognition of obstacles using a convolution neural network. This network consists of 10 layers to recognize 5 obstacles. Since the overall recognition rate is 84.4[%], we've found that this method works well in obstacle recognition.

Keywords: *Obstacle recognition, Convolutional Neural Network, Visually impaired people.*

1. Background

When people with visual impairment go out alone, they risk injuries due to potential contact with still objects in their paths, such as a bicycle placed on the sidewalk. According to a survey done by The Japan Federation of the Blind, half of those visually impaired reported colliding with still objects such as guard rails and poles ⁽¹⁾. Previous studies have tested the use of assistive technologies, like sensors and cameras, to help the visually impaired venture out independently with a reduced risk of collision ^(2,3). These technologies recognize obstacles in real time, but don't specify exactly what they are. As a result, the risk of the obstacle is unclear, no specific information about the obstacle is provided.

To solve this problem, we are advancing research which uses an information system that collects images of obstacles. A wearable camera takes photos of the path ahead of a

visually impaired person, and informs them about any obstacles in their way. To realize this system, it is necessary to recognize obstacles from images with locational information taken by a wearable camera. We hope to establish a method using dynamic pictures taken with a camera that is worn on the clothing. In the discipline of image recognition, we use a CNN, which is known to be highly accurate in the field of image recognition, to recognize obstacles.

2. The Development Process

We will now discuss the development process of our system. This system aims to enable visually impaired people to go out alone, just like healthy ones. According to the results of a survey on the outings of able-bodied persons, they most often go out in the daytime from 10 a.m. to 4 p.m. ⁽⁴⁾. According to a research ⁽¹⁾, visually impaired people think that the risk increases in the rainy weather or in the evening and at night, and they tend to refrain from walking alone. For these reasons, we will first develop a system that can be used during the daytime. Also, walking without assistance, a person with visual impairment wears a camera. When the destination is input into the wearable terminal, it takes dynamic pictures of the path ahead of the person, sever side recognizing obstacles in the walker's path. The worn camera follows the route, while the person walks towards the destination receiving audio messages from a small, the wearable device. So, the visually impaired person can walk unassisted and avoid injury due to collision with unseen objects. For example, **Fig. 1** shows the route from home to a nearby station. This research considers the use of wearable devices like glasses as used in other researches ⁽⁵⁾.

The technology can successfully search and identify obstacles along a route using dynamic images taken by a camera worn on the clothing of a pedestrian. The images go to a server, which attempts to recognize objects, analyze them, and locate the objects on a map. Then we can develop an additional map to navigate based on the map with location and obstacles information. **Fig. 2** shows how the wearable device works to identify obstacles. To realize this system, it is necessary to have the map with the location information of obstacles.

3. Recognition of obstacles using CNN

CNN is an algorithm for an advanced type of machine learning, which has deeper layers than Neural Network (NN). In the first section, we describe the obstacle recognition method using NN, and in the second section, the effectiveness of CNN compared with NN.

3.1 Obstacle recognition using Neural Network

Obstacle recognition using NN is divided roughly into

two steps. The first step is of extracting feature quantities from respective learning images and generating the classifier necessary for recognizing obstacles from those feature quantities. **Fig. 3** shows the construction of the classifier. NN consists of an input layer, a hidden layer, and an output layer. Nodes indicated by circles in **Fig. 3** are set in each layer. These nodes are linked with edges.

NN does two operations to extract feature quantities. First, it propagates all the pixels that form images from the input layer to the output layer. Since nodes in different layers are connected by edges, the numerical data stored in the node are conveyed to the next node by means of the edge. At this time, weight is given to the edge. When the data propagate from the input layer to the output layer, the output layer gets a similarity score of the image including obstacles. This operation is indicated by the right-pointing arrow in **Fig. 3**. The second operation is to update the weight of edges. Here the similarity score is compared with the input images to correct the error. The weight of the edge is updated to reduce error and to bring the similarity score obtained in the output layer close to the entire pixels inputted. When the error becomes the smallest after the repetition of these two operations, we get the weight added to the wedge in the hidden layer as NN's feature quantities. Using this weight as feature quantities, a classifier is generated.

The second step is to recognize obstacles included in the input images by using the generated classifier. Images are first input into the classifier. Then the similarity score of each obstacle learned by the classifier is calculated. An obstacle having the highest similarity is finally recognized as an obstacle existing in the input image.

Fig. 3. The generation process of classifier.

3.2 Convolution Neural Network

CNN is structured upon NN whose hidden layers include the convolution layer and the pooling layer. The convolution layer filters images. By so doing, the layer extracts distinctive feature quantities such as the edges of the image. The pooling layer compresses data by taking the maximum value and the average value in the region with respect to the feature quantities extracted from the convolution layer, leaving the important feature quantities. Thus, it is possible to express the feature quantities of data more compactly, which makes it easier to handle. Overall, the convolution layer extracts the local feature quantities of the image, and the pooling layer process to summarize the local feature quantities. The part with higher accuracy than the NN is compared with the classifier in which the feature quantities is extracted with fine granularity of one pixel, and so there is a case in which it cannot be recognized due to the error of one pixel. CNN solves such problems by extracting feature quantities on a region basis. By the way, a hidden layer can extract and highlight strong features, like those noticed by human beings. In other words, feature quantities extraction in this layer is important for improving the accuracy of recognition. The structure of the hidden layers become more complex as the number of layers increases. On layers of images constructed of at least two or more, this algorithm uses deep learning. By increasing layers, complex obstacles can be recognized. However, the algorithm with too many layers has a problem of over-fitting with learning images. It cannot recognize input images other than the learning images. Therefore, it is necessary to determine the appropriate number of layers.

4. Obstacle recognition experiment

To confirm the effectiveness of the obstacle recognition method using CNN, we have conducted an experiment on obstacle recognition. In this experiment, obstacles for recognition are guardrails, utility poles, pole cones, signposts, and a crossing signal. In the survey of Japan Federation of the Blind, guardrails and utility poles are risky obstacles, so these obstacles are chosen as target obstacles. Also, utility poles, signal poles and crossing signals are target obstacles because the visually impaired may collide with them while walking along the road. Pictures taken with a mobile terminal camera are used as experiment images. Fig. 4 shows each obstacle, we have

Fig. 4. Prepared images(example).

used 200 images for learning and 50 images for the experiment.

These pictures are assumed to contain the features of each obstacle. We regard the tip of a guardrail as its feature. The sign plate fastened around a utility pole serves as its feature. Pole cones are featured by their colors of red and white. Signposts are represented by the signal of no parking. A crossing signal is featured by its cross with black and yellow stripes.

Table 1 shows the network configuration of CNN in this experiment. The filter size of each layer is based on that of VGG-16 which achieved a good result at the image recognition meeting ⁽⁶⁾. VGG-16 is CNN developed at Oxford University. This CNN configuration repeats a combination of a 3×3 filter size convolution layer and a 2×2 filter size pooling layer. Finally, the final output is obtained at the output layer. It is effective for the classification of obstacles because it can recognize 1000 kinds of objects. Also, in the network configuration of CNN in this experiment, we refer to the article by Tomine et al. and take into account the fact that the recognition rate was 81[%] in their research on the configuration with three convolution layers and two pooling layers ⁽⁷⁾. We aim at a recognition rate of 90% or more. So we have constructed more layers than theirs, i.e., 3 convolution layers and 3

Table 1. Parameters of each layers.

Layer name	Filter size(length×width×sets)
I1	56×56×3
C1	3×3×3
P1	2×2×3
C2	3×3×3
P2	2×2×3
C3	3×3×3
P3	2×2×3
N1	7×7×128
N2	1×1024×1
O1	1×5×1

pooling layers. Our network of hidden layers therefore has 10 layers. Let I1 be the input layer, C1, C2 and C3 be the convolution layer, P1, P2 and P3 be the pooling layer, N1, N2 be the fully connected layer, O1 be the input layer. The network is organized in the order of <I1, C1, P1, C2, P2, C3, P3, N1, N2, O1>.

Fig. 5 shows the shape of each layer and **Table 1** shows the parameters of the layer. As for the input layer: I1, because the input layer in the above article has the input image whose length and width are equal in size, we have changed the sizes of our input image into 56×56 pixels. In addition, since the color image is being learned, the image is decomposed into the three channels of RGB components. The convolution layer and the pooling layer are processing for these three channels. The fully connected layer is a layer which combines three-dimensional feature amounts in two dimensions in order to recognize obstacles. By performing the pooling process three times on 56 pixels, both the length and width have become 7 pixels. There are 128 sets of feature quantities of 7×7 pixels because 128 filters are

Fig. 5. The shape of the CNN

output in C3. In the fully connected layer N2, N1 is connected to 1024 nodes. In the output layer: O1, 1024 nodes are connected to 5 nodes. From these nodes, scores of similarity between the experiment image and each obstacle can be obtained, and the highest score is recognized and classified as having the obstacle in the image.

5. Result of obstacle recognition experiment

We assume it is possible to recognize an obstacle when its similarity score is the highest. **Table 2.** shows similarity scores of obstacle in the experiment. In this table, the overall average recognition rate is 84.4[%]. Obstacles other than utility poles offer a similarly high recognition rate of 88% or more. Pole cones are colored red and white, so it is considered that the recognition rate is 100[%] due to this strong feature. However, the recognition rate for utility poles is as low as 42%, which is too low to be safely used by persons with a visual impairment. **Table 3.** shows the results of the recognized obstacles in the image including the utility pole. From this table, it is understood that the obstacles with the highest recognition rate are guardrails and utility poles. In 42 [%] of the experiment images, utility poles were erroneously recognized as guardrails.

In this experiment, we have used images containing parts that we think are feature quantities of obstacles for CNN learning. However, the images have included feature quantities not only of the object itself but also of its background. So we assume that each obstacle has had

Table 2. The results of obstacles recognition.

obstacle	recognition rate [%]
Guardrails	96
Utility pole	42
Pole cone	100
Road signal	88
Crossing signal	98
overall	84.4

Table 3. Misrecognition rate.

obstacle	Similarity rate of obstacles [%]
Guardrails	42
Utility pole	42
Pole cone	2
Road signal	10
Crossing signal	4

Fig. 6. Utility poles erroneously recognized as guardrails.

Fig. 7. Utility poles recognized correctly.

feature quantities different from its feature part.

Fig. 6 shows images of utility poles erroneously recognized as a guardrail, and **Fig. 7** shows images of utility poles correctly recognized. By comparing **Fig. 6** and **7** with **Fig. 4**, we can induce a common feature. Since the utility pole was far away, the colors of the line on the road and the road itself were both recognized as a guardrail. In addition to the foreground area which was featured, the background area was also noted as features. From this, it is inferred that the recognition rate will improve by letting the system learn only the feature qualities part.

Also, despite photographing in the same time zone, there was a difference in recognition. We focus on the value histogram of pictures recognized correctly and those recognized erroneously. **Fig. 8** shows utility poles recognized as guardrails correctly and erroneously. **Fig. 9** and **Fig. 10** also show these histograms.

In these figure, the horizontal axis shows lightness, and the vertical axis indicates their number. The higher the peak value of the graph is in the horizontal axis, the brighter the image becomes. **Fig. 9** is the histogram of correctly recognized images, and **fig. 10** is the histogram of incorrectly recognized images. Comparing these histograms, it can be said that the histogram of the image which could

Fig. 9. The histogram of image which correctly recognized as guardrails.

Fig. 10. The histogram of image which erroneously recognized as guardrails

not be recognized was distributed in low lightness as a whole. We infer that we could not recognize it for this reason.

This time we have corrected the contrast of the image by applying γ correction to the input image with reference to the article by Motizuki, et al ⁽⁸⁾. The image which could not be recognized in **Fig. 8 (b)** has been subjected to gamma correction as shown in **Fig. 11**. **Fig. 12** also shows the histogram of this image.

γ correction was applied to 29 images that could not be recognized correctly. **Table 4** shows the recognition results at this time.

The correction of contrast can be said to be effective for obstacle recognition, since 4 of 29 images were recognized

(a) Utility poles correctly recognized as guardrails. (b) Utility poles recognized erroneously.

Fig. 8. Experiment images

Fig. 11. The image processed through γ correction

Fig. 12. The histogram of image processed through γ correction

Table 4. The results of recognition of the images processed through γ correction

Pictures	The pictures recognized correctly
29	4

correctly. In other words, at dusk or cloudy weather, it is considered that recognition rate improves by correcting contrast by an appropriate method.

6. Conclusion

We have considered the method of obstacle recognition using dynamic images taken with a camera on a portable terminal using CNN. This experiment used the images of five obstacles. Since the overall recognition rate was 84.4%, it can be said that it is effective as a method for recognizing obstacles during the daytime in good weather. CNN with 200 learning images and ten layers is effective because the experiment has shown a high rate of recognition.

On the other hand, although the classifier learned the regular part of each obstacle, there was a big difference in recognition rate. This is probably because the feature quantities determined by humans and the feature quantities determined by CNN are different from each other for different obstacles. Furthermore, in order to improve the accuracy of recognition, it is necessary to increase the number of layers and that of images as shown in the reference ⁽⁶⁾. At the same time, from experimental results, we also find it necessary for the system to learn locally by such method as would help to learn only local parts.

In addition, this time we made it learn only images in the daytime, which is the time zone when the visually impaired most often goes out. We have shown that recognition rate can be improved by applying contrast correction to the

input image even if the amount of light changes in the same time zone due to a change in the weather. In order to further raise the recognition rate, we will consider applying contrast correction to the learning image as well.

In the future, we will confirm which range CNN recognizes as a feature. In order to do so, we will use the method of visualizing the feature points focused on by CNN. As soon as this feature is confirmed, it will learn the local image cut from the range. We will re-experiment with this image and aim for the establishment of obstacle recognition method and the development of a system that maintains robustness against light intensity.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 15K01492.

References

- (1) Japan Federation of the Blind, “Research project report for establishing a system that comprehensively guarantees the safety of the visually impaired when going out”, Tokyo Japan Federation of the Blind, <<http://nichimou.org/wpcontent/uploads/2014/02/9aff1483649e8747de10bfb5c8bca55.pdf>> [Accessed May 2017].
- (2) Takeshi Takegami, et al, “Consideration of surrounding situation sensing for visually handicapped person using USB camera and ultrasonic ranging sensors”, ITE Annual Convention 2010, ROMBUNNO.4-1,2010.
- (3) Katsuma Shinohara, et al. “Development of white cane type walking support for visually impaired”, Technical Reports on Information and Computer Science from Kochi Vol. 6, NO.5, 2014.
- (4) Masashi Kuwano, et al, “Longitudinal Change in Aggregated Time Use based on the Individual Out-home Activity Records”, JSCE, Journal of JSCE D3, Vol. 67, No.5, pp.563-pp.571,2011.
- (5) Keisuke Shimakage, “OTON GLASS at ACM CHI 2015”<<https://otonglass.jp>> [Accessed October 2017].
- (6) Keren Simonyan, et al, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv:1409.1556v6 [cs.CV],2015.
- (7) Kota Tomine, et al. “Classification of group action using convolutional neural network”, 2015 JSPE, pp.923-924, 2015.
- (8) Hiroki Motizuki, et al. “Proposal of a method for discriminating garbage on beach at the beach part using image processing”, IPSJ 76, pp.2.237-pp.2.238,2014.