利用手機幫盲人辨識危險物體

Recognizing Dangerous Objects for Visually Disabled People with a Mobile Phone

Dyi-Rong Duh, Jui-Yu Hung, Yi-Kang Lan, and Chia-Yu Wu
Department of Computer Science and Information Engineering
National Chi Nan University
{drduh, s3321023, s3321028, s3321035}@ncnu.edu.tw

摘要

在一般生活周遭有許多視覺障礙者，他們無法像健全的人們能夠暢行無阻，即使現今有白手杖與導盲犬來輔助視覺障礙者的生活，但對處於此生活的盲人來說，仍有許多不方便之處。擁有一個可以直接告知視覺障礙者前方危險物體是什麼的輔助系統，可以讓他們不受危險物的威脅。其想法是在嵌入式系統上實作出一個輔助設施，藉以判斷正在接近的物體是否為較具威脅性的樓梯、柱子或逼近物體等。因為手機的方便性、經濟性、體積小與具備照相與可發出立體聲之功能，所以在手機上實作此想法。在透過相機截取畫面與辨認前方的物體後，透過警示訊息，以語音直接告知使用者。希望藉由此系統能幫助使用者更方便於行。

關鍵詞：視覺障礙，威脅性，視覺輔助裝置，嵌入式系統，行動電話

Abstract

In our living, there are many visually impaired persons around us. They can not like the normal people who can be unhindered walking. Even though with White Canes and Guide Dogs, which can assist in their living, there still have many inconveniences. Having the system which can directly tell the visually disabled people what the front objects are can help them to avoid the dangerous objects. The idea is to make an implementation of constructing scenes by judging the larger threatening close objects like stairs, pillars, or moving objects on an embedded system to help the visually disabled people. Because a modern mobile phone is inexpensive, small, and easy to carry, the system is implemented on a mobile phone. This work implements a dangerous object detection system on a mobile phone with a camera and an audio device. Additionally, after detecting and identifying the dangerous objects according to the images captured by the camera, the mobile phone can make a warning message to notify the visually disabled people. With the proposed assistive vision system, visually disabled people can live much convenience and stay away from some

dangers.

Keywords: Visually disabled people, threatening, assistive vision device, embedded systems, mobile phone

1. Introduction

This work develops a mobile embedded system to assist the visually disabled people in improving their lives quality and staying away from dangerous objects. The phrase “visually disabled people” is defined in this study as visually impaired or disabled (even temporarily) people.

Eyes are the windows of the soul. For the visually disabled people, the main reason is that visual deficiency causing inconvenient lives. In addition to the traditional White Canes and Guide Dogs, there is also the latest retinal implant technology in existing auxiliary equipment. Those are both in some limited conditions because the users require a long adaptation period to life with White Canes and Guide Dogs, and the latest retinal implant technology requires high-cost and the visually disabled people must have no damage to the optic nerve. In order to assist more visually disabled people, our idea is to develop an embedded system against above limitations, and it can recognize the front dangerous object and tells them what it is by a direct voice. Considering the convenience of the users, the system must be so small and light enough that the visually disabled people can carry it easily. In existing small and light platform, the mobile was selected for an implementation because of its convenience, universal, low-cost, both with image capturing and stereo functions. In all dangerous objects, stairs, pillars and moving objects are the most common seen. Therefore, our goal is to implement the idea about recognizing dangerous objects on a mobile phone to help visually disabled people.

The initial idea of transforming an image into proper sound revealed by Yeh and Lin in 2004 [9]. They have implemented an embedded system for transforming images into sounds by Intel 80C51 microcontroller and Yamaha YM2413B sound IC in their research. Fig. 1 shows the circuit board they designed [9].

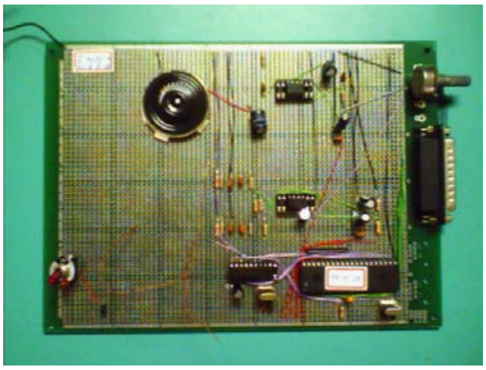


Fig. 1. The embedded system was made by Yeh and Lin for transforming images into sounds [9].

They used MATLAB in a desktop computer to produce simple images, like lines, circles and triangles. Additionally, they used novel methods to identify some specific objects. For example, the system produces a special sound, if there is a vertical line on an image. Therefore, a vertical line in any image can be identified via the special sound. They also developed some efficient algorithms in frequency domain for the judgment. However, the system is too complex to identify objects and each user of their system need to be trained for recognizing simple objects.

The other idea of constructing scenes by hearing is proposed by Duh, Lu and Wang in 2005 [1], and Duh, Wu, Yang and Jiang in 2007 [2]. The visually disabled people can use the system implemented by Duh, Lu and Wang (DLW system) to sense the information outside the body via their ears. The DLW system first transforms the color image captured by a webcam into a binary image named sketch. Second, the sketch is traversed from left to right per column at a time. Finally, a stereo sound composed of different frequencies is made according to the position of each pixel in the sketch. The DLW system was implemented on a notebook and the stereo sound could be presented on loudspeakers. Fig. 2 shows the DLW system [1] and Fig. 3 demonstrates the DLW system transforming some original image into a sketch [1]. However, the DLW system is too big and too heavy so that it is not easy to be carried by a user.

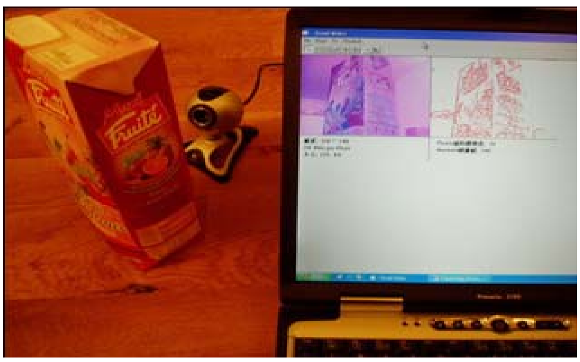


Fig. 2. The DLW system [1].

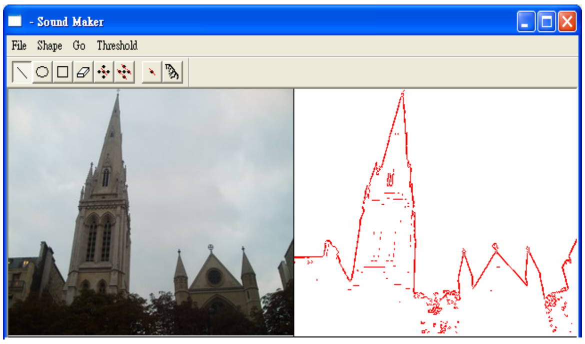


Fig. 3. One of results of the DLW system [1].

The revised system was implemented on NOKIA N70 by Duh, Wu, Yang and Jiang (DWYJ system). The system is smaller and lighter than the DLW system, and it can achieve the same purpose of DLW system. Additionally, a moving object can be detected and identified as follows. First, the DWYJ system takes two continuous pictures and determines features in each image. Second, it gets the first 40 features in each image [3], [4]. Third, the matched features between two images are eliminated and a distinct moving object is located. Finally, present the alert sound for the most obvious object which just detected by the system. The photo of the DWYJ system is shown in Fig. 4 and Fig. 5 demonstrates how to detect and identify a moving object [2].

Fig. 4. The DWYJ system was implemented on a Nokia N70 mobile phone [2].

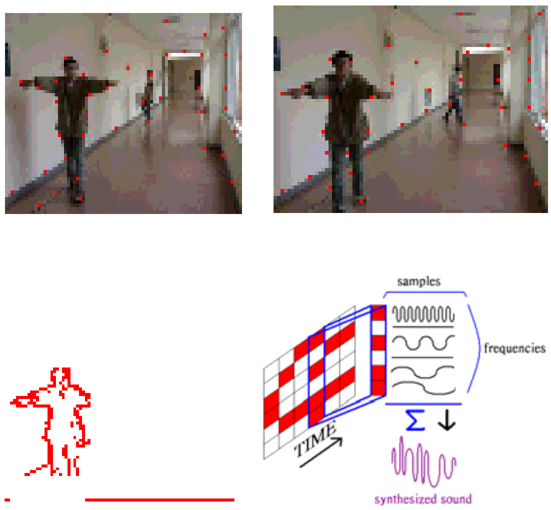


Fig. 5. Detecting and identifying a moving object by the DWYJ system [2].

Even though the DWYJ system is smaller and lighter than the DWL system such that the user can carry it easily, the system can not explicitly express via a voice and the dangerous objects are difficult to be recognized by the untrained user.

"The vOICe" team developed a translator which converts images into sounds for the visually disabled people [4]. The translator could transform an image into proper sound via the camera on a special headset and a notebook computer in a backpack, and then the produced sound is send to earphones as shown in Fig. 6.

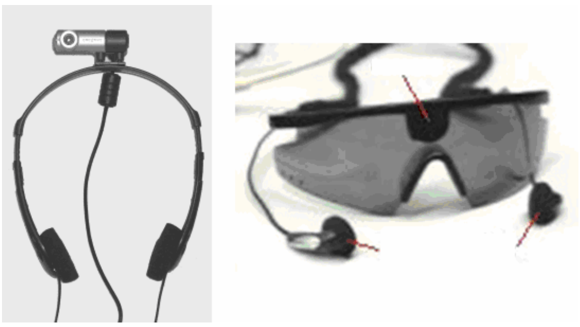


Fig. 6. Headphones used by "the vOICe" team [4].

A cell phone version was developed, too. Since the translator did not transform images into sketches, the visually disabled people must be trained because the image is very complex and hard to be recognized. Restated, their system can not detect moving objects.

There are dangerous objects like stairs, pillars, or moving objects surrounding the visually disabled people. Therefore, our goal is to detect these dangerous objects and inform visually disabled people directly.

In order to recognize dangerous objects and identify them by a voice for visually disabled people with a mobile phone, a simple and fast algorithm is proposed in this study. The general process of the proposed algorithm is depicted in Fig. 7. Based on the algorithm, an embedded system developed on a mobile phone can detect and recognize dangerous objects appear in a captured image such as stairs, pillars, and moving objects and then warn the visually disabled people by a voice.

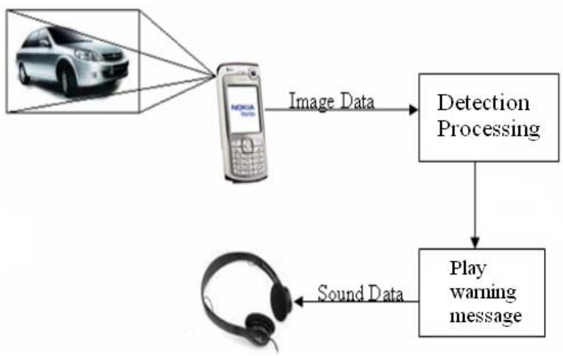


Fig. 7. The general process of the proposed system.

The rest of this paper is organized as follows. Section 2 presents some algorithms for detecting and identifying stairs, pillars and the closest moving object. The algorithms are implemented on a Java-enabled mobile phone NOKIA N70. Section 3 demonstrates some experimental results of the provided system. Conclusions are drawn in Section 4.

2. Principle and Process

This section reveals the principle and process of the proposed system. The work is explained step by step. First, the idea of recognizing pillars, stairs and the closest moving object is presented in detail. Next, how the program works to identify them from the image is demonstrated. Finally, position judgment is also displayed.

2.1 The Idea

The goal is to take pictures by the capture function on mobile phone, and identify whether it is a pillar, a stair or the closest moving object. Based on the result, the system will give a warning message to the user by a voice directly.

First, take two consecutive pictures. Second, gray scale these images, and find the features according to the value of gradient. Third, get the lines which are needed from these features. Finally, recognize the object whether it is a stair or a pillar in front of the user through a series of steps.

For the closest moving object, after taking two pictures and determining the features, compare the difference between these two images. The features on the moving object would change a lot rather than on the background. Choose the features which have the greatest change and form them. The moving object is inside the formed range. Hence, this system can recognize whether the moving object approach the user. Fig. 8 shows what the system does in each round.

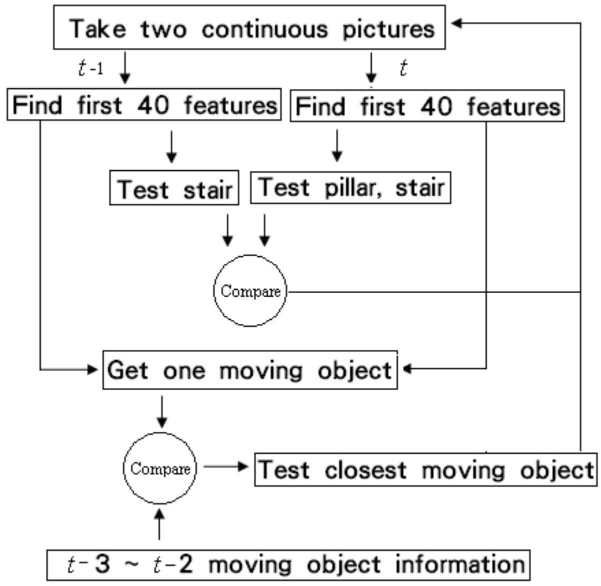


Fig. 8. The flowchart of each round.

2.2 Finding Features of an Image

First, get the value of ARGB for each pixel from the picture captured. Find the gray scale value from the original images according to above values. $Y = 0.3R + 0.59G + 0.11B$, where R is the red value, G is the green value, and B is the blue value of a pixel. As shown in Fig. 9, an original image transforms to gray scaling image [1].

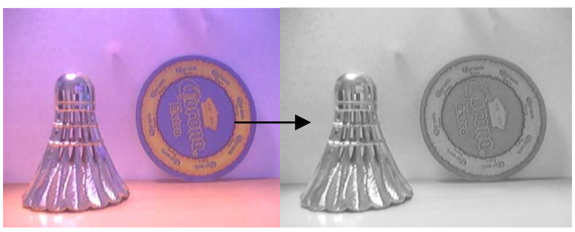


Fig. 9. Gray scale of the original image [1].

Second, calculate the gradient of each pixel. In mathematics, a gradient value is equal to $\sqrt{f_x^2+f_y^2}$, where $f_x = f(x+1, y) - f(x, y)$, $f_y = f(x, y+1) - f(x, y)$, and $f(x, y)$ is the gray scale value of the pixel at coordinate (x, y), that is to say, if a pixel has great gradient value, it will have much difference with its neighbors. According to this result, get contours of the graph by some threshold, as shown in Fig. 10 [1]. Although the threshold is the point of accuracy of the sketching image, this work doesn't focus on it.

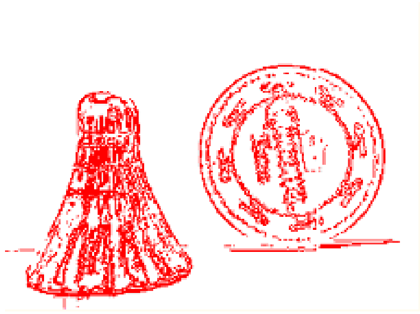


Fig. 10. Sketch the image with a threshold [1].

Third, get the first 40 points by sorting above gradient values as features. When system gets each pixel, it will rule out the possibility of surrounding pixels. This method can avoid that too many features closing to each others. Therefore, this system can select discrete and dissimilar feature points.

2.3 Testing pillars and stairs

This subsection introduces the method of detecting pillars and stairs. Although pillars and stairs are different objects, they are similar in detection. According to pillar's and stair's characteristics, the vertical or horizontal lines are detected which can help us to identify whether the pillar or stair is. The method for detecting vertical and horizontal lines is revealed in the following.

2.3.1 Detecting and composing lines

Use the formula $y = mx+b$ to judge that whether the lines are the same or not. The variable x and variable y are the x-axis and y-axis value of each point of the line. The variable m and variable b are the slope and constant part. When system gets the m and b value of lines, it will detect whether lines are the same or not. For example, if there are two lines, L_1 and L_2, they are the same when m_1 equal to m_2 and b_1 equal to b_2, otherwise, they are two different lines. Then, if the system judge that they are the same and the slope of which are expected, the two lines will be grouped to form a new longest line. Although the slopes are the same, the system will judge that they are not the same lines by the different constant values. Therefore, a new line will be grouped to a new group.

2.3.2 Pillars

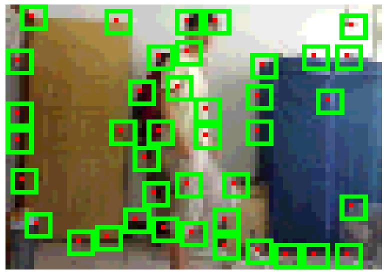
Use the groups generated to find and display the line which is grouped more than 3 times and its length is greater than 1/3 height of the image. That is, at least 4 features should be included in a line and its length must be long enough. If the number of displaying lines is more than one, there might be a pillar in front of the user.

2.3.3 Stairs

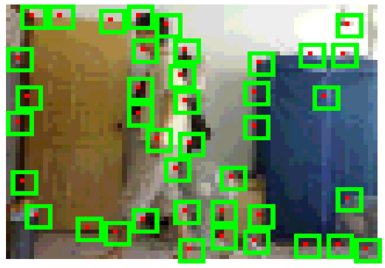
When the light shines the stair, there are bright side and dark side. If the object is a stair, the features will be on the nosing which between the two distinct sides. Then, the discovered lines will be recognized whether the stair is. After taking two pictures according to a series steps presented in Subsections 2.2 and 2.3.1, the system counts the number of features between the two endpoints from the found horizontal line and eliminate the line which is less than four points to avoid producing the coincidence. Thus, based on these feature points in the dispersion of the line, the lines which have less than 50% in the points-dispersion are eliminated. According to the characteristic of a stair, there is some space between ladder and ladder. Because the different horizontals have a different constant part from the formula as shown in Subsection 2.3.1, the constant part is used to compare the horizontals. If there are more than one line which accord with the above conditions and the distance of each other lines are all more than two times of the windows size, this image will be identified as a stair. On the other hand, if there is any condition which does not be accorded, the image will not be identified as a stair. Finally, after judging two images, if they both are judged as stairs, then the front object for the seeing of the visually disabled people is recognized as a stair and a warning message is sounded to them.

2.4 Testing closest moving object

In order to get the closest moving object from these pictures, the system compares the two images after taking two continuous pictures and finds the first 40 features of each picture. The position of features of theses two images are shown in Fig. 11 [2].



(a) Features at time $t-1$



(b) Features at time t

Fig. 11. The first 40 features of the images [2].

There is not much difference in the background and some difference in moving object. Eliminate the features whose differences are smaller than a threshold. As Fig. 12 shown, these remaining features indicate where the moving object is [2]. However, there may be more than one moving object in the image. In order to indicate the moving object concerned, three steps are presented as follows.

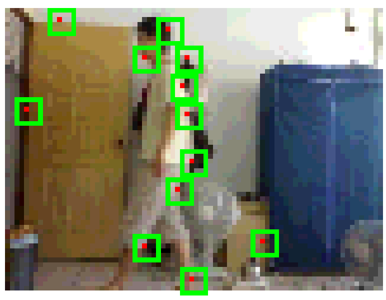


Fig. 12. The remaining features [2].

First, divide features into several groups according to their position as shown in Fig. 13. Second, choose the group which has the most number of features as the distinct moving object because a larger object often have more features. The sketch image of the distinct moving object is depicted in Fig. 14.

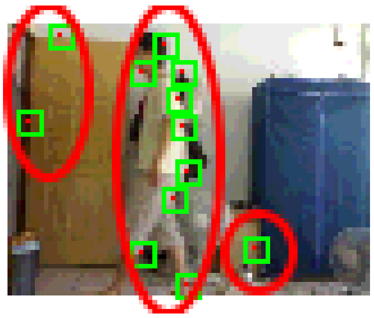


Fig. 13. Three groups in this image [2].

Fig. 14. Sketch image of the moving object [2].

Third, detect whether the moving object is close to the user. After getting the moving object from two continuous pictures, the system saves the square areas of moving object. In the next round, it may find another moving object from the next two continuous pictures. It can compare these square areas. If the latter area is larger than the former one, then this indicates that the object is close to the user. But only with this information of these square areas is not enough to make a correct decision. For example, a moving object which is on left at time t and another moving object which is on right at time t+1, and the latter area is larger than the former. Is there a moving object close to the user? The answer is no, because time t and time t+1 may find two different objects. Therefore, the system saves the central position of the moving objects and compares them. If the central positions of these two moving objects are very close, then there is a moving object close to user.

2.5 Position judgment

For a stair, the left most and right most endpoints of the lines are used to judge the position. The screen is divided into three parts, left, center, and right. If the two points are both on the right part, left part or central part, the system will judge that the object on the same position. If there is one point on the right and the other one is on the left part, the system will judge the position of object which is on the center. If one endpoint is on the center and another one is on the right (or left part), the system will judge the position of the object on the right (or left).

For a pillar, it is the same as a stair. When all x coordinates of lines of a pillar are in the right, left or central part, users will be warned that a pillar is on their right, left or center. If one of lines is on the center and another line is on the right or left, then they will be told that a pillar is in front of them.

Finally, about detecting the position of the moving object, this system will get the central point of it. Judge the point which is on the right, left, or center of the image.

3. Experimental Results

As Fig. 4 shown in Section 1, the proposed system is implemented on a Nokia N70 platform which has practiced in various outside environments under the mobile device is moved. The pictures captured by Nokia N70 can be processed on J2ME Wireless Toolkit simulator for simulation and debugging. All program of the system are written in 100% JAVA for portability. Three interesting experimental results of detecting a pillar, a stair and the closest moving object are shown in the following.

3.1 Detecting pillars

When the contrast of the pillar is greater than the background, the lines of a pillar can be actually found. Therefore, users could be told that there is a pillar in front of them as shown in Fig. 15.

Fig. 15. An example of detecting a pillar.

When the contrast of the background is greater than pillar, the features which are around the pillar can not be found. Therefore, the pillar in front of user couldn't be detected. An example of the missing pillar is shown in Fig. 16.

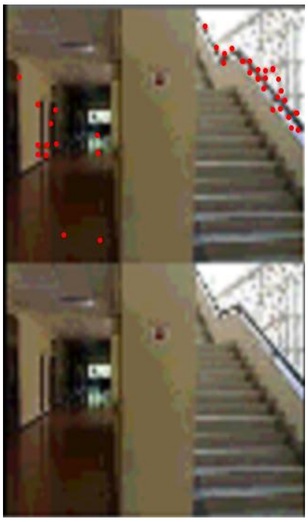


Fig. 16. An example of the missing pillar.

About the false alarm, there are many objects which are like pillar such as elevators, doors or windows. It is easy to detect there are more than one line in this picture. Thus, they may be recognized as pillars as shown in Fig. 17.

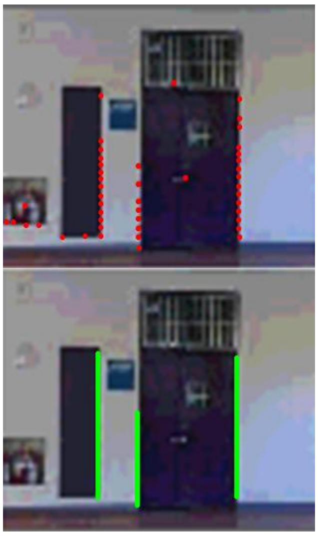
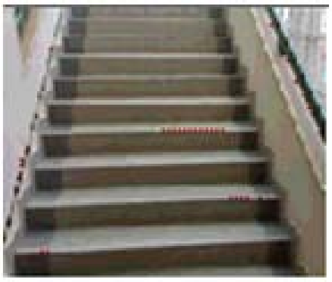


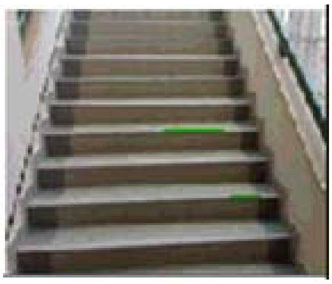
Fig. 17. An example of the false alarm of pillars.

3.2 Detecting Stairs

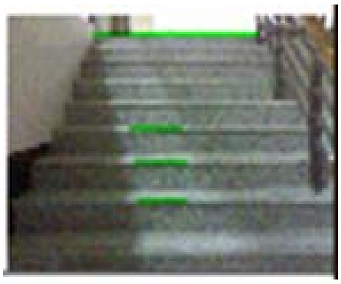
The second experiment tries to detect a stair. If the number of green lines detected by our system in each image is more than one, the system will judge that there is a stair in the image. Additionally, if a stair is detected in both two consecutive images, the system will identify a stair here. As shown in Fig. 18, these results show that our system can detect and identify a stair.



(a)The features found by our system.



(b)The number of lines is 2.



(c) The number of lines is 4.

Fig. 18. Two examples of detecting stairs.

The result of this experiment is obviously correct because the found feature points which are in the nosing. The nosing can help our system to detect the stairs. Restated, the horizontal lines can be detected by the steps presented in Subsections 2.3.1 and 2.3.3.

Although this experiment is correct, some conditions might lead to the formation of miss and false alarm. For the miss example as shown in Fig. 19, the stair material let the features scatter everywhere and the mopboard also let the most features scatter in the edge. Such cases may let the feature points can't scatter in the nosing.

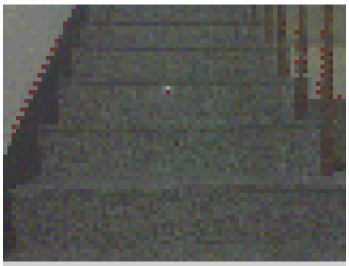


Fig. 19 A miss example of detecting stairs.

A false alarm case may appear if a blonde car (especially a white car) in the sample image. The system will find the lines between the windscreen and the steel plate. This case will also affect the judgment as Fig. 20 shown.

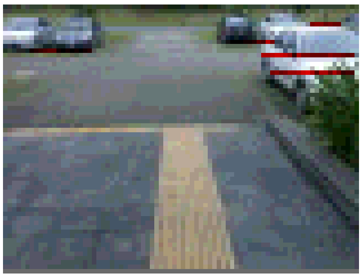


Fig. 20 A false alarm case for detecting stairs.

In addition, the mobile phone must be straight to the stairs to capture pictures or the system may miss stairs.

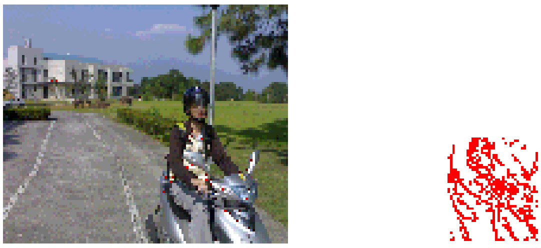
3.3 Detecting a close moving object

Our system takes two continuous pictures to detect a moving object in each time. As shown in Fig. 21, a motorcycle is coming to the user. The system detects a moving object (the motorcycle) and its area gets greater than before. Hence, it recognizes that there is a moving object close to the user.

(a) The two continuous pictures at time $t-1$.

(b) The two continuous pictures at time t.

(c) The moving object at time $t-1$.

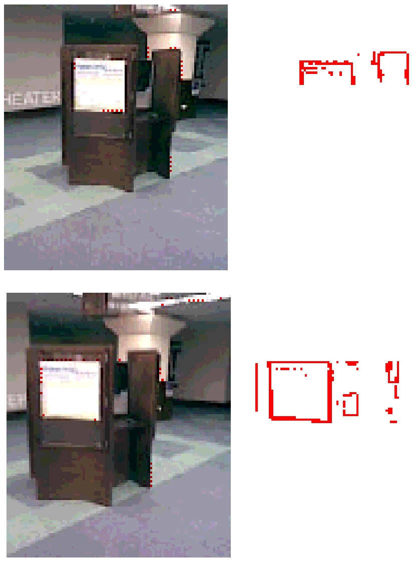


(d) The moving object at time t.

Fig. 21 Detect a close motorcycle.

Our system can also detect the block as shown in Fig. 22 and it will make a voice to warn the user. Although the object does not move, the system can judge the moving object when user gets closer.

(a) User is close to a block.



(b) Detect an object from this block.

Fig. 22 When user is close to a block.

This system could still get miss as shown in Fig. 23. Although there is a person getting closer, the features are too discrete on the pictures having a person. Therefore, when the features are grouped, no proper objects containing the person can be identified. As shown in Fig. 23(b), the system can only get a half of body of the person. That is a situation of miss.

When mobile phone is static, a moving object is close to user. This condition can be detected by this system. If you carry this mobile phone and move around to test this system, you may receive some false alarms.

The reason is that we do not always forward one direction to go. Therefore the features are changed a lot in the pictures. The system will get great performance when it is static.

(a) A person is close to the user.

(b) Detect a moving person.

Fig. 23 A miss from a closing person.

4. Conclusions

According to our implementation, this system can really achieve our goal on the visible mobile platform; it can organize stairs, pillars, and the closest moving object. Even you take this system on walking, it can detect the stairs, pillars and some moving objects. If any mentioned object is detected, the system will warn the user by a voice. A user will hear three kinds of voices to recognize which kind of dangerous objects is close. Therefore, user can easily carry the mobile phone which equipped with the above functions to detect the surroundings and stay away from dangers.

However, the developed system still has some shortcomings. The first one is that it takes two consecutive pictures in each round and this takes about 0.3 seconds because the camera's cycle time. Additionally, the system takes about 3 seconds to detect a dangerous object and voices it. In other words, the system can not warn user immediately. That is, the visually disabled people with the system cannot be notified in time. At this moment, the system user should walk slowly for having enough time to detect and warn the user. The problem can be solved by providing a better detection algorithm or replacing the mobile phone with a high performance one. Another weak point is that the system will make a wrong decision when the camera is big shaking. To cope with the problem, an anti-shake camera should be included in the system.

References

[1] D.-R. Duh, C.-W. Lu, and K.-H. Wang, "Constructing scenes by hearing of the visual disabilities," *Proc. National Computer Symp.*, Yung-Kang City, Tainan Hsien, Taiwan, Dec. 15–16, 2005.

[2] D.-R. Duh, J.-X. Wu, L.-C. Yang, and J.-S. Jiang, "Detecting and identifying a distinct moving object with a mobile phone for visually disabled people,"

Proc. 4th Int. Conf. in Information Technology and Application, Harbin, China, vol. 1, pp. 248–253, Jan. 15–18, 2007.

[3] B. Jung and G. S. Sukhatme, “Detecting moving objects using a single camera on a mobile robot in an outdoor environment,” *Proc. 8th Int. Conf. on Intelligent Autonomous Systems*, Amsterdam, Netherlands. pp. 980-987, Mar. 2004.

[4] P. B. L. Meijer, "Seeing with sound," http://www.seeingwithsound.com.

[5] Nokia Corporation, “MIDP 2.0: Working with pixels and drawRGB(),” May 2003, http://www.forum.nokia.com/info/sw.nokia.com/id/a0116d4e-6f4f-4fb9-b786-802e48710eb9/Working_witrh_pixels_and_drawRGB_v1_0.pdf.html.

[6] Sun Microsystems, “Java technology for the wireless industry specification,” http://jcp.org/aboutJava/communityprocess/final/jsr185/index.html.

[7] Sun Microsystems, “Mobile information device profile 2.0,” http://jcp.org/aboutJava/communityprocess/final/jsr118

[8] C. Tomasi and T. Kanade, “Detection and tracking of point features,” *Technical Report CMU-CS-91-132*, Carnegie Mellon University, 1991.

[9] H.-C. Yeh and T.-C. Lin, “An embedded system for the visual disabilities seeing scenes by ears,” Project Report, Department of Computer Science and Information Engineering, National Chi Nan University, Oct. 2004.