Using Remote Vision to Navigate Visually Impaired Pedestrians: The Effects of Video Image Resolution on the Performance

Vanja Garaj
School of Engineering and Design, Brunel University
Uxbridge, Middlesex, UB8 3PH, United Kingdom
e-mail: vanja.garaj@brunel.ac.uk

ABSTRACT

This paper presents a study to determine the effects of video image resolution on the performance in recognition of static hazards in the pedestrian travel environment using remote vision. The hazards recognition is a basic activity in the application of remote vision to navigate visually impaired people. As such, the study contributes to the process of development of a novel navigation system for the visually impaired. This system combines a remote vision facility, the GPS, an application of the GIS and a telecommunications unit to provide a platform enabling remote navigation of visually impaired pedestrians by the sighted human guide. The results show that variations in the resolution of video image as applied in the study [704x480 pixels - 4CIF NTSC, 352x240 pixels - CIF NTSC and 176x120 pixels QCIF NTSC] do not cause a significant difference in the ability to recognize static hazards in the travel environment based on video image.

Keywords: Remote Vision, GPS, GIS, Navigation, Visually Impaired People, Video Image Resolution, Video Image Quality

1. INTRODUCTION

A prototype of a system for the navigation of visually impaired people was developed by the Electronic Systems Research Group at the School of Engineering and Design, Brunel University. The system, titled the System for Remote Sighted Guidance of Visually Impaired Pedestrians, integrates a wireless remote vision facility with a positioning and tracking unit based on the GPS and an application of the GIS into a technological platform enabling the remote guidance of visually impaired pedestrians by a sighted human guide [3, 4, 6].

The remote vision facility permits the remote sighted guide to navigate the visually impaired user of the system in the immediate travel environment (micro-navigation; e.g. the assistance in the avoidance of obstacles and other environmental hazards in the path of travel), while the integrated GPS and GIS unit facilitates the navigation through the environment on a large scale (macro-navigation) [7]. The implementation of the system and the consequent availability of the remote sighted guidance service hold the potential to provide mobility assistance comparable to actual sighted guidance [1]. A considerable advantage of remote sighted guidance is that it can offer relative independence of mobility [3].

The prototype of the System for Remote Sighted Guidance of Visually Impaired Pedestrians consists of two terminals [3]. One terminal is designed for use by a visually impaired person receiving guidance while travelling (the user, the user's terminal) and the other by a sighted person remotely guiding the user through means of the system (the remote sighted guide, the guide's terminal).

The user's terminal is a wearable mobile device that includes a portable video camera, a GPS receiver and an electronic compass (Figure 1). The guide's terminal is organised as a stationary personal computer workstation involving a GIS application and a screen with the capacity to concurrently present the digital map of the user's travel environment contained in the GIS application, the video image recorded by the camera in the user's terminal and the user's heading data from the electronic compass (Figure 2). The video camera built into the user's terminal and the video image display in the guide's terminal form the basis of the system's remote vision facility, whereas the GPS receiver and the electronic compass in the user's terminal, combined with the GIS application and the digital map display in the guide's terminal, comprise the system-integrated positioning and tracking unit.

When the system is in operation, the video camera in the user's terminal (positioned on the user's chest and pointed onwards) continuously records the video image of the immediate environment ahead of the user - covering the area extending vertically from the ground up to above the level of the user's body height and horizontally in multiple body widths. At the same time, the GPS receiver captures the radio signals emitted by the GPS satellites visible to the antenna of the receiver at any given moment in time and, based on the information on the position of the satellites in space encoded in the signals, the processing unit in the user's terminal calculates the location of the receiver, i.e. the location of the user, in the user's travel environment. In addition, the electronic compass establishes the data on the user's heading.

In parallel, the video image and the information on the location and heading of the user are transmitted from the user's terminal to the guide's terminal via a wireless link. In the guide's terminal, the location and heading are presented on the screen of the terminal together with the received video image. The process of updating the video image and the location and heading is repeated continuously - for as long as a remote guidance session takes place.

By monitoring the video image update as the user is engaged in locomotion, the remote sighted guide can assist the user in micro-navigation and the location and heading updates supply the guide with the spatial information required for the provision of macro-navigational assistance. Micro- and macro-navigational instructions constituting remote sighted guidance are delivered by the remote sighted guide to the user in verbal form. The delivery occurs through the two-way voice communication channel established as a part of the wireless link between the user's and the guide's terminals. The voice communication channel also enables the user to explain to the remote sighted guide the location of the desired journey destination before starting a remotely guided journey, to detail the preferable content and syntax of the navigational instructions delivery as well as to raise any requests that may occur during the journey - for example, a possible request to swap the originally planned destination for an alternative.

As a part of the process of the system development, a study was carried out to determine the effects of video image resolution on the ability of the remote sighted guide to recognise static environmental hazards important for the micro-navigation of visually impaired pedestrians. The recognition of the environmental hazards based on the video image of the environment is the initial step in the provision of micro-navigational assistance utilising the system's remote vision facility[1]. The study is presented in the remainder of this paper.

Figure 1. The User's Terminal

[1] In the micro-navigational assistance provision using the System for Remote Sighted Guidance, the recognition of an environmental hazard is followed by the delivery of navigational instructions as to how to avoid the hazard.

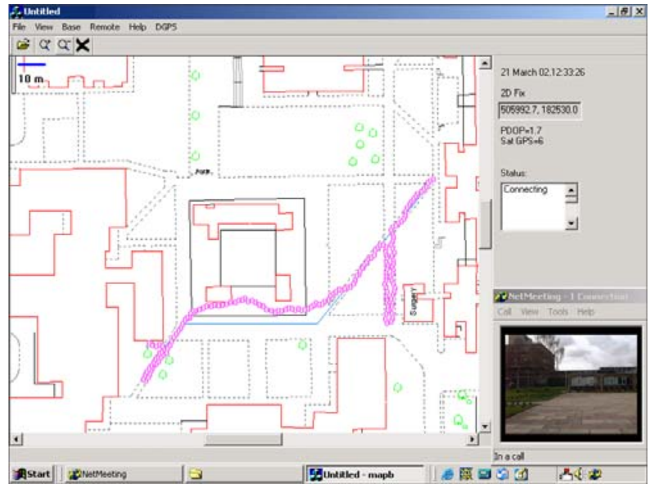

Figure 2. The Screen of the Guide's Terminal

2. THE STUDY

The information on how video image resolution influences the performance in the recognition of environmental hazards is required to support the decision as to which resolution to apply in the remote vision facility in the future process of the system implementation. Herewith, there are two main points to consider: a) the operational video image resolution should enable the remote sighted guide to deliver guidance with the maximum effectiveness and minimum effort; b) the resolution must be achievable within the technological context of the existing 3G-telecommunications infrastructure.

In planning the study, it was hypothesised that a reduction in video image resolution would have a negative effect on the performance level of the remote sighted guide. This hypothesis followed the thought that the resolution reduction would make the hazards more difficult to detect due to a decrease in size in which they appear on the guide's terminal screen.

2.1 Method

The research methodology that was employed in the study is entirely in consistence with the experimental research methodology originating in the science of ergonomics [9, 10]. The study was based on a simulation of the remote vision facility utilisation in the provision of micro-navigational assistance, which was carried out in the laboratory conditions.

The study involved 30 sighted participants divided in three groups of ten (Group A, Group B and Group C). All participants were shown a series of seven different pre-recorded video clips representing the video image of the user's immediate environment ahead that could, in the real world, be captured by the camera in the user's terminal of the System for Remote Sighted Guidance of Visually Impaired Pedestrians while the user of the system is engaged in locomotion.

As the video clips were playing, the participants had to verbally report the type and location of the static environmental hazards that were appearing in the clips (for example: “a lamp post in front”)[2].

The seven video clips presented to the study participants never differed in terms of their content. However, they did vary in video image resolution. The participants in Group A were shown the seven video clips encoded in the resolution of 704x480 pixels (Version 1; 4CIF NTSC format), the participants in the Group B in the resolution of 352x240 pixels (Version 2; CIF NTSC format) and the participants in the group C in the resolution of 176x120 pixels (Version 3; QCIF NTSC format).

Other perceivable video quality parameters, including video image resolution, jerkiness, blockiness, blurriness, noise, ringing and colourfulness distortion [5, 8] were exactly the same or very similar in all three groups.

The effects of video image resolution were established by comparing the hazards recognition performance (Hit Rate) across the three groups.

The recognition performance was measured across four different categories of static environmental hazards with relevance to micro-navigation. The four categories in question are:

1) *Primary Obstacles* - Environmental features with the potential to obstruct the walk (e.g. street furniture, traffic signs, road and pavement works, cars parked on the pavement, kerbs and steps) that are positioned in the travel path directly “in line” with the user’s body (“in line” = in the direction of the user’s heading - anywhere within the width of the user’s body and from the ground level up to the level of the user’s head);

2) *Secondary Obstacles* - Environmental features with the potential to obstruct the walk that are positioned in the travel path, but not directly “in line” with the user’s body (the obstruction may occur in cases of sudden changes in the walking direction);

3) *Tertiary Obstacles* - Environmental features with the potential to obstruct the walk that are positioned along both the left and the right border of the path of travel;

4) *Path Information* - The type of travel path surface and the surface of the area bordering the travel path (both on the left and on the right border of the travel path).

The four categories were defined based on the classification of the spatial information necessary for the micro-navigation of visually impaired people that was drawn up by the Working Group on Mobility Aids for the Visually Impaired and Blind of the U.S. National Research Council [2].

[2] The video clips were devoid of any dynamic obstacles. The recognition of dynamic obstacles has been explored in another study.

2.1.1 Participants

All 30 participants in the study were recruited from the population of undergraduate and postgraduate students in various schools based at Brunel University. The participants were paid £6.00 each to take part in the study.

While assigning the participants to the two groups, the effort was made to keep the age range and mean age as well as the female-to-male ratio similar across the groups. If it did exist, a large between-group difference in these parameters would act as a variable that could have a negative impact on the study outcome [11].

The mean age of the participants in Group A was 23.4 years (SD=4.1 years), the age range 19-29 years and the female-to-male ratio 40:60. The mean age of the participants in Group B was 22.2 years (SD=3.1 years), the age-range 19-28 years and the female-to-male ratio 60:40. The mean age of the participants in Group C was 25.2 years (SD=4.3 years), the age range 21-33 years and the female-to-male ratio 50:50.

2.1.2 Apparatus

The seven video clips used in the study (VC.1, VC.2, VC.3, VC.4, VC.5, VC.6, VC.7) were originally recorded in the AVI video format using a Sony DCR-TRV 110 video camera (Hi-8). Subsequently, the clips were edited and Version 1, Version 2 and Version 3 image resolution forms of the clips were created by Adobe Premiere 6.0 video-editing software. The video format in which the clips were presented to the study participants is MPEG-2 format. The same format is applied to display video image on the screen of the guide’s terminal in the developed prototype of the System for Remote Sighted Guidance of Visually Impaired Pedestrians.

The video clips present a range of diverse environmental settings and, as such, a variety of static environmental hazards that can be encountered on everyday journeys through urban environment. Whereas the number of secondary obstacles, tertiary obstacles and the environmental features falling into the Path Information category varies between the clips, in each of the clips there is only one primary obstacle. The primary obstacle always appears in the ending part of the clip and all the clips finish exactly at the point of contact with the obstacle. Figure 3 shows a sequence of image captures from one of the video clips used in the study (VC.7).

While the video clips were being recorded, the video camera was continuously held in the position that resembles the chest position at which the camera integrated in the user’s terminal of the System for Remote Sighted Guidance of Visually Impaired Pedestrians is located when the terminal is worn by the user (Figure 1). The walking speed during the recording was kept at around 1m/s. The speed[3] was controlled by a speedometer built into the portable pedometer that was carried along.

[3] 1m/s is the average speed of walk in sighted people and in visually impaired people who travel supported by a guide dog.

Figure 3. A sequence of images from one of the study video clips

The number of hazards existing in the clips, broken down according to the four categories of the hazards recognition performance assessment, is shown in Table 1.

Table 1. The number of hazards in the video clips

Hazard Category	No. of Features per Video Clip							Total
	VC. 1	VC. 2	VC. 3	VC. 4	VC. 5	VC. 6	VC. 7	
Primary O.	1	1	1	1	1	1	1	7
Secondary O.	0	0	4	0	0	4	3	11
Tertiary O.	11	4	3	0	4	5	8	35
All Obstacles	12	5	8	1	5	10	12	53
Path Info.	13	8	9	7	4	10	7	58
All Hazards	25	13	17	8	9	20	19	111

The primary obstacles featured in the clips are a bollard (VC.1), a car parked on the pavement (VC.2), a bicycle barrier (VC.3), a pavement works situation (VC.4), a pole-mounted traffic sign positioned on the pavement (VC.5), a group of people blocking the entire width of the pavement while waiting for a bus (VC.6) and a hedge overgrowing the pavement at the head-height level (VC.7). The majority of the featured primary obstacles belong to the group of obstacles in the detection of which visually impaired people often experience difficulties when they employ a long cane or a guide dog to support micro-navigation.

The duration of the video clips ranges from 24s (VC.4) to 71s (VC.5). The specification of the duration for each of the clips is provided in Table 2.

Table 2. The video clips duration

VC.1	VC.2	VC.3	VC.4	VC.5	VC.6	VC.7
59s	26s	44s	24s	71s	52s	21s

The clips were presented on a 17" CRT monitor (VGA) made by Viglen. During the presentation of the clips, the monitor screen was set at the resolution of 800x600 pixels. The screen capture below (Figure 4) shows the relationships between a study video clip (VC.1) in Version 2 video image resolutions and the monitor screen. The clips were run using Windows Media Player. Throughout the engagement in the performance of the study task (the task of the hazards recognition), the study participants were positioned approximately 80cm in front of the monitor.

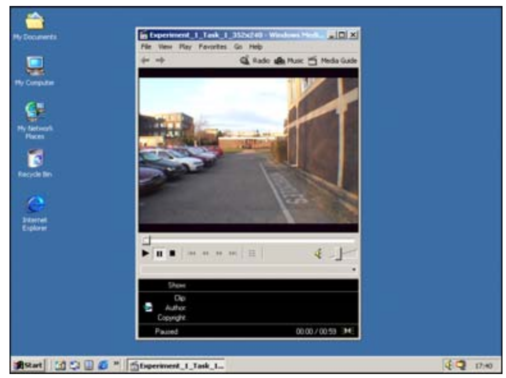

Figure 4. A study video clip in relation to the resolution of the screen: Version 2 (352x240 pixels)

2.1.3 Procedure

The study consisted of 30 individual sessions that took place over a period of 15 days (two sessions per day). Each session involved one of the 30 study participants and lasted approximately 1 hour and 15 minutes. All the sessions were conducted following identical procedure, which is described as follows.

Firstly, the participants were provided with an extensive introduction to the study. The introduction consisted of two parts. In the first part, a short video film was shown that presents the purpose and the modus operandi of the System for Remote Sighted Guidance of Visually Impaired Pedestrians.

The principal reason for showing the film was to allow the participants for the conceptual placement of the study task in the context of the real-world application of the system. The second part of the introduction involved explaining the study task per se. In order to ensure that all the participants received the details of the task in exactly the same way, the explanation was delivered in written form. After reading the explanation, the participants were shown another short video film. This video film presents an illustration of all four categories of environmental hazards that were meant to be recognised in the seven study video clips (Primary Obstacles, Secondary

Obstacles, Tertiary Obstacles and Path Information). As the introduction finished, the participants were answered all questions they had regarding the study and the system.

Subsequently, the participants were submitted to a hazards recognition training. The training was designed for the participants to practice the hazards recognition prior to the actual performance that was carried out in the main part of the study sessions. Based on the same set of actions as the actual tasks performance, the training involved the participants in recognising environmental hazards important for the micro-navigation of visually impaired people in the two training video clips.

The training was followed by the main part of the study sessions. In this part, each of the participants was shown all seven video clips (the participants in Group A - Version 1 resolution form of the clips, the participants in Group B - Version 2 and the participants in Group C - Version 3). As mentioned above, in order to assess their features recognition performance, the participants were asked to verbally report the presence (the type and location) of static hazards important for the micro-navigation of visually impaired people that exist in the clips while the video clips were running.

The order in which the clips were presented was the same for all the participants. The verbal reports of the environmental hazards presence were recorded by a portable voice recorder. The audio recordings were subsequently employed in documenting the performance of the participants.

2.2 Results

After all 20 study sessions were completed, the audio recordings of the verbal reports by the study participants were analysed to gather the performance data for the hazards recognition task.

The performance data gathering was based on the careful listening of the recorded verbal reports given by the 30 study participants. While listening, all the static hazards reported by each of the participants for each of the seven video clips were marked (ticked off) on the hazards lists in the corresponding performance record sheets. The sheets were devised earlier by monitoring the video clips[4] (a set of the seven sheets was assigned for every participant).

When the performance record sheets were thus populated, the total number of the hazards reported by each of the participants for each of the seven video clips was calculated across the four assessment categories (Primary Obstacles, Secondary Obstacles, Tertiary Obstacles and Path Information) plus the two cumulative categories named as “All Obstacles”[5] and “All Features”[6]. The calculation was carried out based on counting the hazards in each of the categories that were on the hazards lists in the performance record sheets marked as reported.

Consequently, using the calculated total number of the reported static hazards (the hits) and the previously established total number of the hazards existing in the clips (this number was gathered from the hazards lists included in the performance record sheets), the ratios (the hit rates) were determined for each of the participants of the reported and the existing hazards.

Like the hits, the hit rates per participant were also established for all seven video clips and across all four features categories and the two cumulative categories. As reported above, the ratio of the reported and the existing hazards, i.e. the hazards hit rate, was in the study designated as the measure of the hazards recognition performance.

A summary of the hit rates data is provided in Table 3. The table and figure present the mean values (in the table marked as M) and the standard deviations off the mean values (marked as SD) for the hit rates achieved by participants in the two groups. The mean values and the standard deviations are presented across the four main plus the two cumulative categories.

Table 3. The Hazards Hit Rates for Group A, Group B and Group C - Mean Values per Category

Category		Group A -------------- Version 1	Group B ------------- Version 2	Group C ----------- Version 3
Primary Obstacles	M	0.9286	0.9286	0.9286
	SD	0.0753	0.0753	0.0753
Secondary Obstacles	M	0.6091	0.6273	0.6364
	SD	0.0748	0.1088	0.1485
Tertiary Obstacles	M	0.5600	0.5200	0.5029
	SD	0.1862	0.1311	0.1028
All Obstacles	M	0.6264	0.5943	0.5906
	SD	0.1339	0.0963	0.0781
Path Information	M	0.4362	0.4552	0.3879
	SD	0.1431	0.0969	0.0576
All Hazards	M	0.5171	0.5216	0.4847
	SD	0.1182	0.0845	0.0600

As the hit rates data was organised, ANOVA analysis was carried out in order to enable the determination of the video image resolution effects on the performance in the hazards recognition task. The analysis included the comparison between the mean hit rates for Group A, Group B and Group C across all

[4] The performance record sheets contain a list and the total number of all hazards appearing in the video clips.

[5] “All Obstacles” = Primary Obstacles + Secondary Obstacles + Tertiary Obstacles

[6] “All Hazards” = Primary Obstacles + Secondary Obstacles + Tertiary Obstacles + Path Information

assessed categories of the hazards plus the two cumulative categories. The ANOVA analysis results are shown in Table 4 on the following page.

Table 4. ANOVA Analysis Results

$F_{2, 27}(\alpha = 0.05) = 3.3500$		
Category	F	P
Primary Obstacles	0.0000	> 0.05
Secondary Obstacles	0.1465	> 0.05
Tertiary Obstacles	0.4131	> 0.05
All Obstacles	0.3494	> 0.05
Path Information	1.0863	> 0.05
All Hazards	0.4932	> 0.05

3. CONCLUSION

The assessment of the static environmental hazards recognition performance across the Primary Obstacles, Secondary Obstacles, Tertiary Obstacles and Path Information categories of environmental features with relevance to the micro-navigation of visually impaired people that was carried out in the study, revealed either none or only rather small differences in the performance levels achieved by the study participants in Group A, Group B and Group C.

As observable from the summary of the hit rates data in Table 3, the mean values of the combined hit rates in the Primary Obstacles category do not differ at all, whereas the mean value differences in other three categories do not exceed the margin of 5.71% (present in the Tertiary Obstacles category - between Groups A and C). Logically, only small differences also exist in the two cumulative categories of "All Obstacles" (Maximum mean value difference margin = 3.58% - between Groups A and C) and "All Features" (Maximum mean value difference margin = 3.69% - between Groups B and C).

A high degree of similarity in the hazards recognition performance levels between participants in Group A, Group B and Group C was also displayed in the ANOVA analysis that was carried out on the hit rates data. The analysis found that, statistically, the combined hit rates achieved by the participants in the two groups do not differ significantly in either of the four plus two categories (Table 4).

Accordingly, the reasoning is feasible that the variation in the resolution of the video clips shown to the study participants (Group A - Version 1: 704x480 pixels, Group B - Version 2: 352x240 pixels, Group C - Version 3: 176x120 pixels) does not incur any effects on the static hazards recognition performance.

The study evidence allows for the hypothesis about the reduction in the video image resolution causing a negative impact on the performance in the static hazards recognition that was drawn while planning the study to be rejected.

4. REFERENCES

[1] Farmer, L. W. and Smith, D.L. *Adaptive Technology*. In Blasch, B.B., Wiener, W.R. and Welsh, R.L. (Eds.) *Foundations of Orientation and Mobility*. 2nd Edition, Chapter 7, 231-259. AFB Press, American Foundation for the Blind, New York, USA, 1997.

[2] Foulke. E., Bach-Y-Rita, P., Blasch, B.B., Brabyn, J., Enoch, J., Faye, E.E., Goodrich, G., Keeney, A.H., Scadden, L. and Warren, D.H. (Working Group on Mobility Aids for the Visually Impaired and Blind, Committee on Vision, National Research Council). *Electronic Travel Aids: New Directions for Research.* National Academy Press, Washington D.C., USA, 1986.

[3] Garaj, V. *Design of a System for Remote Sighted Guidance of Visually Impaired Pedestrians*. Ph.D. Thesis, Brunel University, Uxbridge, UK, 2006.

[4] Garaj, V., Jirawimut, R., Ptasinski, P. and Balachandran, W. *A System for Remote Sighted Guidance of Visually Impaired Pedestrians.* British Journal of Visual Impairment, 21, 2, (2003), 55-63.

[5] Genista Corporation. *Video Quality Metrics.* Genista Corporation, Tokyo, Japan, 2002.

[6] Hunaiti, Z., Garaj, V. and Balcahandran, W. *A Remote Vision Guidance System for Visually Impaired Pedestrians*. The Journal of Navigation, 59, 3, (2006), 1-8.

[7] Petrie, H., Johnson, V., Strothotte, T., Raab, A., Michel, R., Reichert, L. and Schalt, A. *MoBIC: An Aid to Increase the Independent Mobility of Blind Travellers.* British Journal of Visual Impairment, 15, 2, (1997), 63-66.

[8] PictureTel Corporation. *iPower Video Technology (White Paper).* PictureTel Corporation, Andower (MA), USA, 2001.

[9] Sanders, M.S. and McCormick, E.J. *Human Factors in engineering and Design.* McGraw-Hill, New York, USA, 1992.

[10] Wickens, C.D., Gordon, S.E. and Liu, Y. *An Introduction to Human Factors Engineering.* Longman, New York, USA, 1997.