Using Remote Vision: The Effects of Video Image Frame Rate on Visual Object Recognition Performance

Vanja Garaj, Ziad Hunaiti and Wamadeva Balachandran, *IEEE Fellow*

***Abstract*—The process of using remote vision was simulated in order to determine the effects of video image frame rate on the performance in visual recognition of stationary environmental hazards in the dynamic video footage of the pedestrian travel environment. The recognition performance was assessed against two different video image frame rate variations: 25 fps and 2 fps. The assessment included a range of objective and subjective criteria. The obtained results show that the effects of the frame variations on the performance are statistically insignificant. This study belongs to the process of development of a novel system for navigation of visually impaired pedestrians. The navigation system includes a remote vision facility and the visual recognition of the environmental hazards by the sighted human guide is a basic activity in aiding the visually impaired user of the system in mobility.**

***Index Terms*—Visually impaired people, Navigation, Remote vision, Teleoperation, Video image frame rate, Visual object recognition, User interfaces**

I. Introduction

A prototype of a new system to navigate visually impaired people was developed by the Electronic Systems Research Centre at the School of Engineering and Design, Brunel University. The system is aimed at supporting travel in unfamiliar environments and assisting in the situations of orientation loss during journeys on familiar routes. As such, the use of the system is to complement the traditional mobility aids - cane and guide dog. Named the System for Remote Sighted Guidance of Visually Impaired Pedestrians (SRSGVIP), the new system integrates a wireless remote vision facility with a positioning unit based on the GPS and an application of the GIS into a multimedia platform enabling the remote guidance of visually impaired pedestrians by a sighted human guide [1, 2, 3]. 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 hazards in the path of travel), while the GPS and GIS unit facilitates the navigation through the environment on a large scale (macro-navigation) [4]. The implementation of the system and the consequent availability of the remote sighted guidance service hold the potential to facilitate comprehensive mobility assistance comparable to actual sighted guidance [5]. A considerable advantage of remote sighted guidance is the independence of mobility that it offers [2].

The architecture of the SRSGVIP is presented in Figure 1 on the following page. The currently established prototype of the system consists of two terminals [1, 2, 3]. One terminal is designed to be utilized by a visually impaired person receiving guidance while travelling (the user, the user's terminal) and the other is meant for utilization by a sighted person remotely guiding the user by means of the system (the remote sighted guide, the guide's terminal).

The user's terminal has the form of a wearable mobile device that involves a video camera, a GPS receiver and an electronic compass. The camera is placed on the user's chest and pointed onwards (Figure 2). The guide's terminal is organized as a stationary personal computer workstation including 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 3). The video camera built into the user's terminal and the video image display in the guide's terminal are the basis of the system's remote vision facility. The GPS receiver and the electronic compass in the user's terminal combined with the GIS application and the digital map display within the guide's terminal comprise the system-integrated positioning unit.

When the system is in operation, the video camera in the user's terminal 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 in-built 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

Manuscript received December 15, 2008.
Dr. Vanja Garaj is with School of Engineering and Design, Brunel University, Uxbridge, Middlesex, United Kingdom (Phone: + 44 (0) 1895 266 964; Fax: + 44 (0) 1895 258 728; e-mail: Vanja.Garaj@brunel.ac.uk).
Dr. Ziad Hunaiti is with Faculty of Science and Technology, Anglia Ruskin University, Chelmsford Campus, Bishop Hall Lane, Chelmsford, CM1 1SQ, United Kingdom (e-mail: Z.Hunaiti@anglia.ac.uk).
Prof. Wamadeva Balachandran is with School of Engineering and Design, Brunel University, Uxbridge, Middlesex, United Kingdom (e-mail: Wamadeva.Balachandran@brunel.ac.uk).

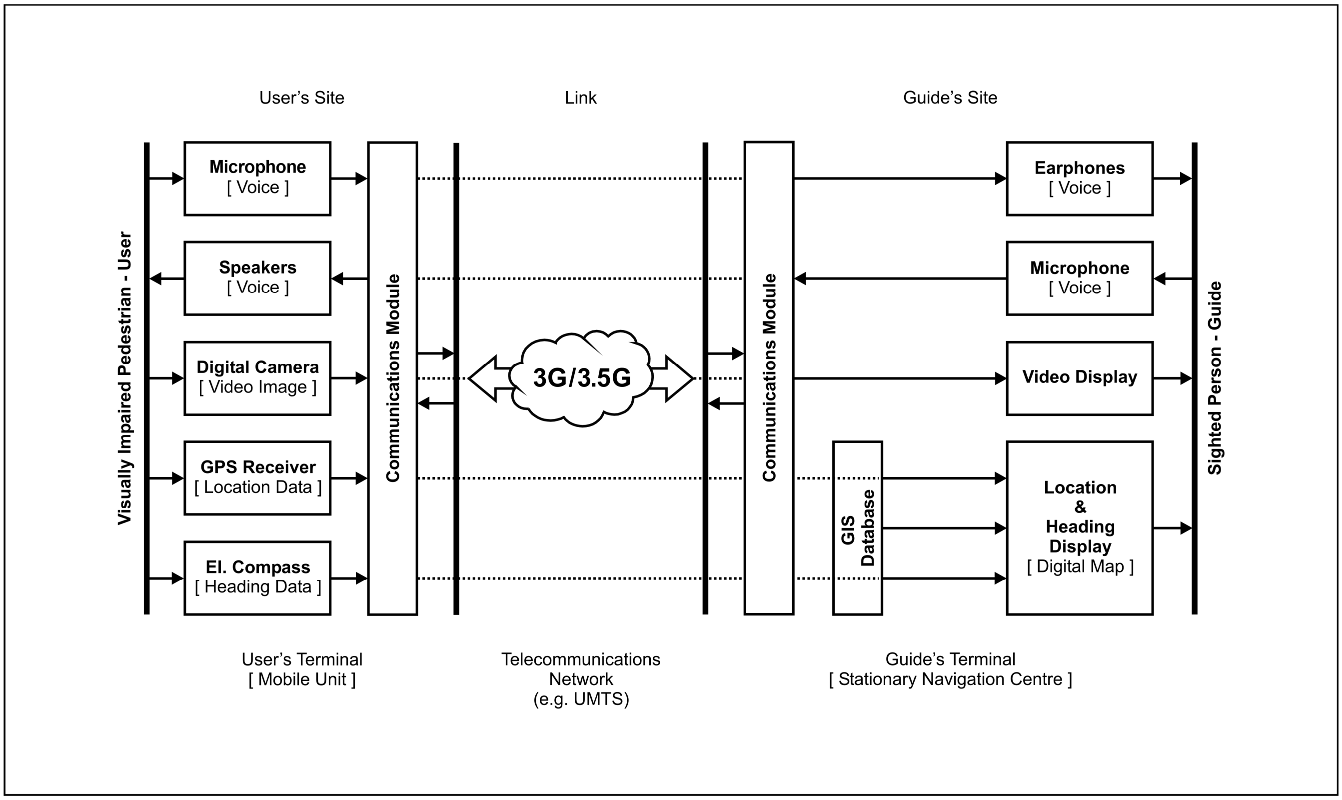

Fig. 1. The architecture of the System for Remote Sighted Guidance of Visually Impaired Pedestrians

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. Moreover, the electronic compass establishes the data on the user's heading (direction).

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 the wireless link connecting the terminals. 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 in micro-navigation and the location and heading updates provide the guide with the spatial information required for the provision of macro-navigational assistance. Micro- and macro-navigational instructions constituting remote sighted guidance

Fig. 2. The User's Terminal

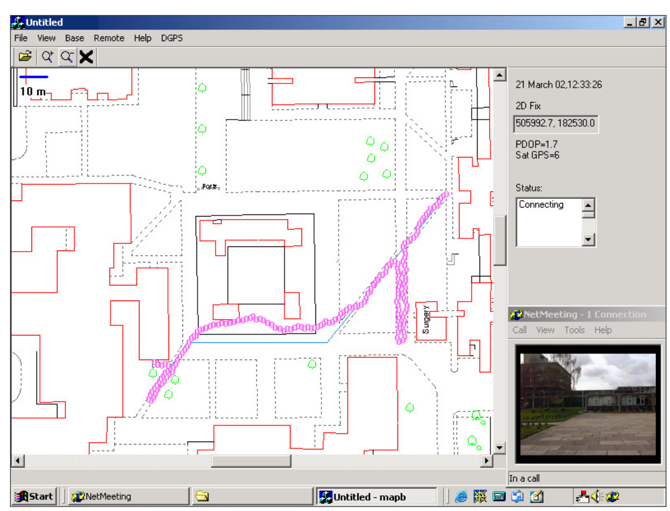

Fig. 3. The screen of the Guide's Terminal

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 journey destination before starting a remotely guided journey, to detail the preferable content and syntax of navigational instructions delivery and to raise any requests that may occur during the journey - e.g. a possible request to swap the originally planned destination for an alternative.

There are two different scenarios being considered for the implementation of the SRSGVIP. The current system prototype, with the user's terminal in the mobile form and the guide's terminal being stationary, has been developed in support of the scenario to establish the system as a specialized navigation centre - to be manned with a number of trained remote sighted guides capable to attend multiple users of the system. Another, more cost-efficient, scenario is to develop a fully mobile system, which would permit the visually impaired system user to receive navigation assistance from a family member or a trusted friend - on the individual basis. This scenario would require the current version of the guide's terminal to be modified so that it is operable on mobile platforms such as PDAs and smart phones.

In both scenarios, the system-enabled navigation assistance should ideally be made accessible any time, any place and for however long it is needed. Whereas some visually impaired users may require the assistance to be provided throughout the journey (e.g. particularly when travelling on unfamiliar routes), a preliminary analysis of the user requirements for the system configuration suggests that most users would need it intermittently - i.e. only on certain journey parts (e.g. in cases of accidental departures from a known route) [2]. Besides the practicality of facilitating travel from Point A to Point B, it is envisaged that the use of the system would contribute by the travel stress-reducing effect it is to afford to the user. This effect stems from the opportunity to share journey-related responsibilities with the remote sighted guide - as opposed to having to rely entirely on oneself.

An option being considered as further work is to integrate an automated guidance mode to the existing system set-up. In the automated guidance mode, the system will automatically generate and present macro-navigational instructions in the similar manner as in-car GPS navigation systems (the instructions will be generated based on the location and heading information from the system's positioning unit (Figure 1) and presented to the user as synthesized voice messages). If so agreed with the user, the remote sighted guide will be able to switch the system to the automated guidance mode after helping the user to set off on a journey and before the user approaches the journey sections that are, for example, less demanding in terms of macro-navigation and where micro-navigation is not required (a reliable micro-navigational assistance would be very difficult to achieve in the automated navigation mode due to limitations of the existing digital image processing technologies). The automated guidance mode is expected to add to the versatility and efficiency of the system.

II. The Study

As a part of the process of the system development, a study was carried out to determine the effects of video image frame rate on the ability of the remote sighted guide to recognize the stationary environmental hazards that are important for the micro-navigation of visually impaired pedestrians. Successful recognition of the environmental hazards appearing in the remotely transmitted video image of the pedestrian travel environment is the basic step in the provision of micro-navigational assistance utilizing the system's remote vision facility. The hazards recognition extends to the delivery of the navigational instructions on how to avoid them.

The knowledge of the relationship between the frame rate and the visual recognition performance is crucial to support the decision as to which video image frame rate to apply in the remote vision facility in the future implementation of the system. Herewith, there are two main points to consider:

1) The operational frame rate should enable the remote sighted guide to recognize the hazards and deliver guidance with the maximum effectiveness and minimum effort;

2) The frame rate must be achievable within the technological context of the existing telecommunications infrastructure (3G or 3.5 G) (Figure 1).

In planning the study, it was hypothesized that the recognition performance level of the remote sighted guide would significantly decrease with the frame rate being reduced. This hypothesis was formed based on the following reasoning. The frame rate reduction causes the video image to appear jittery, with the objects in the image "moving" in saccades. In turn, monitoring the video input to identify the hazards becomes increasingly demanding because the remote sighted guide's visual system must constantly conduct spatiotemporal interpolations to "fill" the occurring visual gaps in the image [6].

A. Method

The study was based on a simulation of the remote vision facility utilization in the provision of micro-navigational assistance, which was carried out in controlled laboratory conditions.

The study involved 20 sighted participants divided in two groups of 10 (Group A and Group B). All participants were shown a series of seven different pre-recorded video clips representing the video footage 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 SRSGVIP 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 stationary environmental hazards that were appearing in the clips (for example: "a lamp post in front", "a row of cars on the left hand-side", "a hedge on the right hand-side", "a flight of stairs ahead", etc.). An instance of reporting a hazard was considered as the successful recognition of the hazard. The video clips were devoid of any dynamic hazards - this hazard type was explored in another study.

The seven video clips presented to the study participants never differed in terms of their content. Nevertheless, they did vary in video image frame rate. The participants in Group A were shown the seven video clips encoded with the frame rate of 25 fps (Version 1) and the participants in Group B with the frame rate of 2 fps (Version 2). Other perceivable video quality parameters of the clips in Version 1 and Version 2 were exactly the same (the video image resolution for both versions was 176x144 pixels) or very similar (the parameters such as blockiness, blurriness, noise, ringing and colorfulness distortion [7, 8]).

The frame rate of 25 fps was chosen to represent the ideal case (control condition) in the study because it allows the human eye for the jitter-free perception of the observed video image and thus enables comfortable viewing. As such, this frame rate has been used both in the analogue PAL and SECAM television/video systems and in the 576i standard definition digital television/video system [9]. Another characteristic making the frame rate of 25 fps suitable for the ideal case is that it is still impossible to achieve continuously on average 3G and 3.5G telecommunications infrastructure [10]. The frame rate of 2 fps was selected to contrast the ideal frame rate of 25 fps. The streaming of video image at 2 fps is possible on most existing networks [10]; however, this frame rate makes video image unsteady and therefore much more difficult to monitor.

The effects of video image frame rate were established by comparing the hazards recognition performance (hit rate) of the participants in Group A with the performance of the participants in Group B. The recognition performance was measured across four categories of stationary 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 (e.g. within the boundaries of a pavement) 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 user's head level);

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 intentional or unintentional 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 (Figure 4) were defined based on the classification of the spatial information necessary for the micro-navigation of visually impaired people that was developed by the Working Group on Mobility Aids for the Visually Impaired and Blind of the U.S. National Research Council [11].

In addition to the objective measurement of the hazards recognition performance as described above, the frame rate effects were assessed through the application of the following four subjective criteria:

1) *Mental Demand*

The self-perceived (by the study participants) amount of the mental activity required to perform the recognition task;

2) *Temporal Demand*

The self-perceived amount of the time pressure felt during the performance of the recognition task;

3) *Effort*

The self-perceived amount of the work (both mental and physical) invested into the performance of the task;

4) *Frustration Level*

The self-perceived level of the frustration experienced during the task performance.

Participants

All 20 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, mean age and 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 [7, 8]. The mean age of the participants in Group A was 25.2 years (SD = 4.3 years), the age range 21-33 years and the female-to-male ratio 50:50. The mean age of the

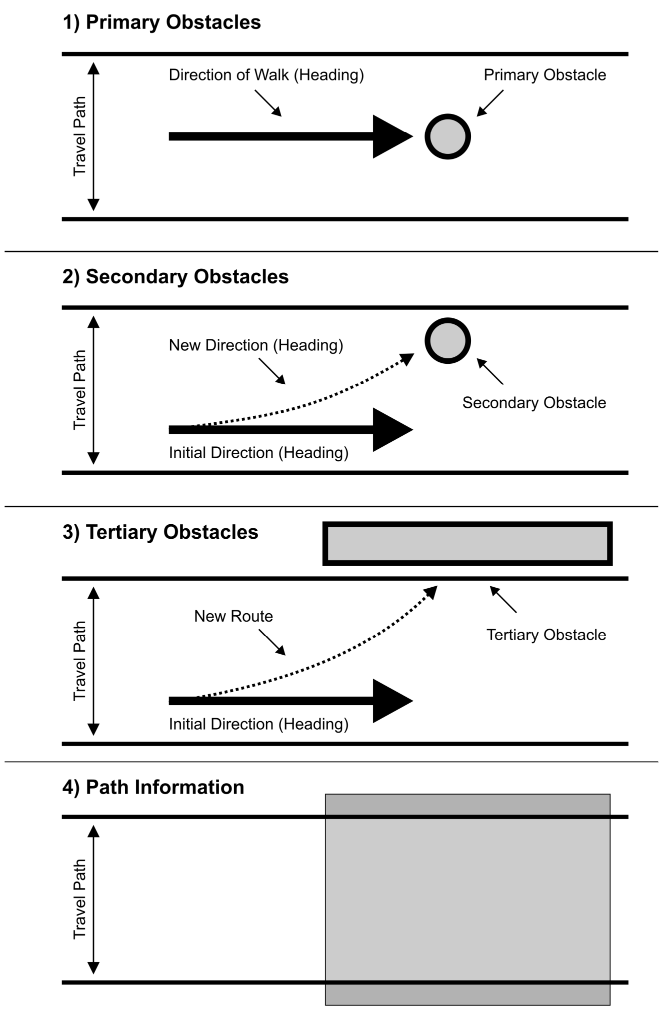

Fig. 4. A diagram of the four categories of hazards in relation to travel path

participants in Group B was 23.6 years (SD = 3.7 years), the age-range 21-31 years and the female-to-male ratio 40:60.

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 AVI format using a Sony DCR-TRV 110 video camera (Hi-8). Subsequently, the clips were edited and the Version 1 and Version 2 frame rate 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 SRSGVIP. The clips were presented in the resolution of 176x144 pixels (Quarter Common Intermediate Format - QCIF).

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 SRSGVIP is located when the terminal is worn by the user (Figure 2). The walking speed during the recording was always kept at around 1 m/s. This speed is the average speed of walk in sighted people as well as in visually impaired people who travel supported by a guide dog. The speed was controlled by a speedometer built into the portable pedometer that was carried along.

The seven video clips present a range of diverse environmental settings and, as such, include a variety of stationary environmental hazards that can be encountered on everyday pedestrian journeys through the urban environment. Whereas the number of secondary obstacles, tertiary obstacles and the environmental hazards 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 5 shows a sequence of image captures from one of the video clips used in the study (VC.1). The total number of hazards existing in the clips, broken down according to the four categories of the assessment, is shown in Table I.

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 experience difficulties when they employ a long cane or a guide dog to support micro-navigation. The duration of the clips ranges from 21 seconds (VC.7) to 71 seconds (VC.5). The specification of duration for each of the clips is provided in Table II. The overall duration of the footage is just below 5 minutes (297 seconds).

Fig. 5. A sequence of images from one of the study video clips

TABLE I
THE NUMBER OF HAZARDS IN THE VIDEO CLIPS

Hazard Category	VC 1	VC 2	VC 3	VC 4	VC 5	VC 6	VC 7	Total
Primary Obstacles	1	1	1	1	1	1	1	*7*
Secondary Obstacles	0	0	4	0	0	4	3	*11*
Tertiary Obstacles	11	4	3	0	4	5	8	*35*
All Obstacles	*12*	*5*	*8*	*1*	*5*	*10*	*12*	*53*
Path Information	13	8	9	7	4	10	7	*58*
All Hazards	*25*	*13*	*17*	*8*	*9*	*20*	*19*	*111*

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. A capture of the screen with one of the video clips used in the study (VC.1) is shown in Figure 6. The software that was utilized to run the clips is Windows Media Player. Throughout the engagement in the performance of the hazards recognition task, the study participants were positioned approximately 50 cm in front of the monitor. This viewing distance is recommended by the International Telecommunication Union (ITU) standards for video image quality assessment [12, 13].

Procedure

The study consisted of 20 individual sessions that took place over the period of 10 days (two sessions per day). Every session involved one of the 20 study participants and lasted approximately 1 hr and 15 min (including the informal conversations at the beginning and end). All sessions were conducted following identical procedure described as follows.

At the start of their participation in the study, each of the participants successfully underwent a simple visual acuity test. In the test, the participants were asked to read a 9 mm high line of text from the distance of 6 m. The test text line represented the bottom text line on the Snellen chart. Accordingly, the ability to read the test text line from 6 m away meant the possession of at least 6/6 (20/20) visual acuity; i.e. normal vision if the reading was carried out unaided or corrected-to-normal vision if it was done with the aid of glasses or contact lenses.

TABLE II
THE VIDEO CLIPS DURATION

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

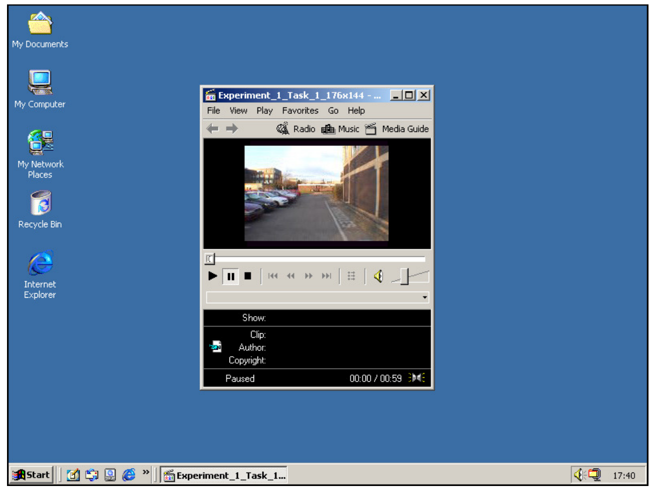

Fig. 6. A study video clip relative to the resolution of the display

On the completion of the visual acuity test, 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 SRSGVIP. 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. This video presents an illustration of all four categories of environmental hazards that were meant to be recognized in the seven study video clips (Primary Obstacles, Secondary Obstacles, Tertiary Obstacles and Path Information). Additionally, a spatial diagram was presented that illustrates the hazard categories in relation to the path of travel (Figure 4) and the instructions given to consider only the hazards within the boundaries of the travel path. When the introduction finished, the participants were permitted time to ask any questions they had regarding the study and the questions were answered in detail.

The participants were then submitted to a hazards recognition training. Based on the same set of actions as the actual hazards recognition carried out in the main part of the study, the training involved the participants in recognizing the environmental hazards important for the micro-navigation of visually impaired people in two training video clips. After the training was completed, the participants were made aware of the committed errors, i.e. the hazards they failed to recognize - in order to enable them to improve their performance. The video clips presented to the participants in the training session were encoded in the same way (frame rate, resolution, etc.) as the clips they saw subsequently in the main part of the study.

The main part of the study followed a 5 minutes break after the completion of the training session. In this part, each of the participants was shown all seven video clips (the participants in Group A - Version 1 frame rate form of the clips and the participants in Group B - Version 2). As mentioned above, in order to assess their hazards recognition performance, the participants were asked to verbally report the presence (the type and location) of stationary 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 later used in documenting and analyzing the participants' performance.

As the final step in the study sessions, the participants were requested to rate the self-perceived mental demand, temporal demand, effort and frustration level they associated with the performance of the hazards recognition task. These ratings were applied to assess the frame rate effects from a subjective point of view. The ratings had to be expressed on the scale between 1 and 20, with the value of 1 depicting the minimum mental demand, temporal demand, effort and frustration level and the value of 20 the maximum.

The participants provided the ratings by filling in the rating sheet they were given by the experimenter. Before providing the ratings, the participants were supplied with the written definitions of the four criteria. The subjective ratings provision concluded the sessions.

Results

After all 20 study sessions were completed, the audio recordings of the verbal reports by the study participants were analyzed to gather the performance data for the hazards recognition task. The performance data gathering was based on careful listening of the recorded verbal reports given by the 20 study participants. While listening, all stationary 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 that were devised earlier by monitoring the video clips. The performance record sheets contain a list and the total number of all hazards appearing in the video clips and a set of seven sheets (a sheet per video clip) 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"[1] and "All Hazards". 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. Subsequently, using the calculated total number of the reported stationary hazards (hits) and the previously established total number of the hazards existing in the clips, which was acquired from the hazards lists included in the performance record sheets, the ratios (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 the four basic and the two cumulative hazard categories. As reported above, the ratio of the reported and the existing hazards, i.e. the hazard 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 III (numerical) and Figure 7 (diagrammatic). The table and the figure present the mean values (M) and the standard deviations off the mean values (SD) for the hit rates achieved by participants in the two groups.

TABLE III
THE HAZARD HIT RATES: MEAN VALUES AND STANDARD DEVIATIONS

Hazard Category		Group A Version 1: *25 fps*	Group B Version 2: *2 fps*
Primary Obstacles	M	0.93	0.89
	SD	0.07	0.09
Secondary Obstacles	M	0.64	0.64
	SD	0.15	0.10
Tertiary Obstacles	M	0.50	0.49
	SD	0.10	0.08
All Obstacles	*M*	*0.59*	*0.57*
	SD	*0.08*	*0.06*
Path Information	M	0.39	0.43
	SD	0.06	0.06
All Hazards	*M*	*0.48*	*0.50*
	SD	*0.06*	*0.04*

[1] "All Obstacles" = Primary Obstacles + Secondary Obstacles + Tertiary Obstacles; "All Hazards" = Primary Obstacles + Secondary Obstacles + Tertiary Obstacles + Path Information

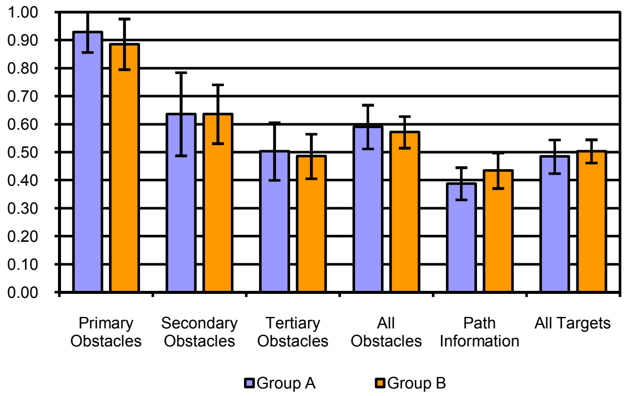

Fig. 7. The hazard hit rates for Group A and Group B: Mean values and standard deviations

When the hit rates were arranged, ANOVA analysis was carried out on the data in order to enable the determination of the video image frame rate effects on the performance in the hazards recognition task by means of interferential statistics. The results of the ANOVA analysis for the hit rates data are presented in Table IV.

Following the hit rate data analysis, the ratings were analyzed that the study participants provided to quantify the mental demand, temporal demand, effort and frustration level. The summary of the subjective ratings data is shown in Table V. The table presents the mean values (M) and the standard deviations (SD) for each of the four ratings criteria for Group A and Group B. The presentation of the ratings data in diagrammatic form is given in Figure 8. The results of the ratings ANOVA analysis are given in Table VI.

TABLE IV
ANOVA ANALYSIS RESULTS: THE HAZARD HIT RATES

$F_{1,18}(\alpha = 0.05) = 4.4100$

Category	F	P
Primary Obstacles	1.3279	> 0.05
Secondary Obstacles	0.0000	> 0.05
Tertiary Obstacles	0.1738	> 0.05
All Obstacles	*0.3843*	*> 0.05*
Path Information	2.9541	> 0.05
All Hazards	*0.6721*	*> 0.05*

TABLE V
THE SUBJECTIVE CRITERIA: MEAN VALUES AND STANDARD DEVIATIONS

Subjective Rating Category		Group A Version 1: *25 fps*	Group B Version 2: *2 fps*
Mental Demand	M	14.40	15.20
	SD	5.18	3.43
Temporal Demand	M	12.20	14.10
	SD	5.18	3.90
Effort	M	12.30	16.10
	SD	5.58	2.38
Frustration Level	M	9.60	12.50
	SD	6.28	4.09

III. DISCUSSION

The assessment of the stationary environmental hazards recognition performance across Primary Obstacles, Secondary Obstacles, Tertiary Obstacles and Path Information categories of the environmental hazards 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 and Group B.

As observable from the summary of the hit rates data in Table III and Figure 7, the mean values of the hit rates in the Secondary Obstacles category do not differ at all, whereas the

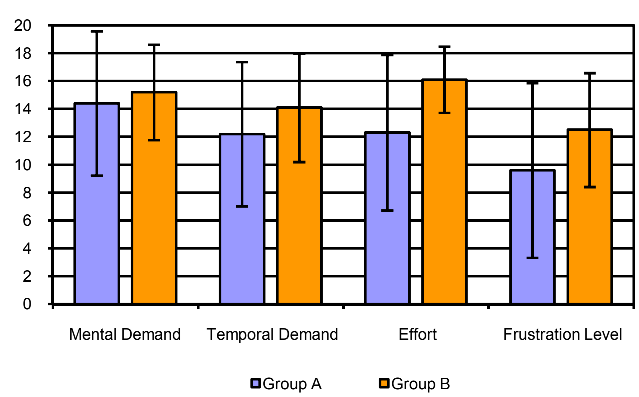

Fig. 8. The subjective criteria ratings for Group A and Group B: Mean values and standard deviations

TABLE VI
ANOVA ANALYSIS RESULTS: THE SUBJECTIVE CRITERIA

$F_{1, 18} (\alpha = 0.05) = 4.4100$

Category	F	P
Mental Demand	0.1655	> 0.05
Temporal Demand	0.8584	> 0.05
Effort	3.9263	> 0.05
Frustration Level	1.4991	> 0.05

mean value differences in other three categories do not exceed the margin of 4% (present in the Primary Obstacles and Path Information categories). Logically, only small differences also exist in the two cumulative categories of “All Obstacles” (2%) and “All Hazards” (also 2%).

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

A very similar situation is present in the ratings for Mental Demand, Temporal Demand, Effort and Frustration Level criteria of the assessment. The maximum between-groups difference of 19% was obtained for the criterion of Effort, the difference in Frustration Level is 14.5%, in Temporal Demand 9.5% and in Mental Demand only 4% (Table V, Figure 8). Even though in all four criteria there is a tendency for the ratings to increase with the frame rate reduction, the ANOVA analysis that was carried out on the ratings (Table VI) shows that the increase is not statistically significant.

According to the study results, the reasoning is feasible that the variation in the frame rate of the video clips shown to the study participants (Group A - Version 1: 25 fps, Group B - Version 2: 2 fps) does not have a considerable influence on the stationary hazards recognition performance. To conclude, the study evidence indicates that the hypothesis (Section II) about the reduction in the video image frame rate possibly causing a negative impact on the performance in the stationary hazards recognition may be rejectable. Nevertheless, the final judgment regarding the hypothesis rejection cannot be made before additional tests are completed. These tests should involve a larger sample of participants and a wider selection of the travel environment types and environmental hazards represented in the test video clips.

This outcome is consistent with several other studies conducted to test the effects of video image frame rate on visual object (target) recognition performance. For example, Chen *et al*. [14] tested the military targets recognition performance of the remote human operator of an unmanned aerial vehicle in a simulated military environment. The study concluded that the frame rate degradation from 30 fps to 5 fps did not significantly affect the performance. In another teleoperation-related study, French *et al.* [15] assessed the frame rate effects on the military targets recognition performance in operating an unmanned ground vehicle. This study included the frame rates of 2, 4, 8 and 16 fps and found that even the lowest rate of 2 fps did not have a significant impact on the performance. Cai [16] investigated the frame rate effects on the recognition of human figure in video surveillance applications. The conclusion was that a significant drop in the performance occurs only when the frame rate is reduced below 1 fps.

The effects of frame rate have been assessed with similar results in the context of some other types of human activity. The visual object (target) acquisition performance-related study by Bryson [17], looking specifically into the task of placing the cursor to a particular position on a personal computer system screen, found that the task difficulty increases significantly only when the screen frame rate drops below 4 fps. Liu *et al.* [18] carried out another study in the visual object acquisition domain. This study was centered around the pick-and-place task in 3D virtual environments (telemanipulation), using a head mounted display. The study findings revealed that the experienced operators’ errors did not show a significant increase until the display frame rate was reduced below 2 fps.

Several studies have explored the frame rate effects on the visual object (target) tracking performance. In a study involving the remote control of a ground vehicle and the related visual tracking task of maintaining the vehicle’s position within a specific path by monitoring the driving lane markings on a remotely transmitted video, Van Erp and Padmos [19] found that the control was significantly degraded with the video frame rate being reduced below 5 fps. A similar study by McGovern [20] did not find any significant control degradation within the range between 30 fps and 7.5 fps and the study by Day [21] concluded that the indirect driving-related visual tracking performance becomes negatively affected when the frame rate is below 4 fps.

An example of human activity whereby the frame rate is a critical factor contributing to the performance is the speech (lip) reading from the remote video (a form of visual recognition task). As demonstrated by Vitkovitch and Barber [22, 23], the frame rates below 12.5 fps may result in a great loss of visual information, thus causing a significant decrease in the speech reading performance.

The information on the relationship between video image frame rate and the environmental hazards recognition performance that was established in the study presented in this paper is bound to be seen as useful by those responsible for the future implementation of the SRSGVIP. As demonstrated in the study, the application in the remote vision facility of the video image with the frame rate of 25 fps is to allow the remote sighted guide for the equal level of the stationary hazards recognition performance as the video with the much smaller frame rate of 2 fps. This fact enables a substantial flexibility in terms of engineering the wireless data link to transmit video image between the user's and the guide's terminals of the system [10]. Besides the application in the context of the SRSGVIP, the results of the study may be of use to the developers of other types of teleoperation systems involving remote vision. Such systems are in recent years being increasingly developed for deployment in the areas of space exploration, military operations, emergency services and health care [24].

REFERENCES

[1] V. Garaj, R. Jirawimut, P. Ptasinski, F. Cecelja and W. Balachandran, "A System for Remote Sighted Guidance of Visually Impaired Pedestrians". British Journal of Visual Impairment, vol. 21 (2), pp. 55-63, May. 2003.

[2] V. Garaj, "Design of a System for Remote Sighted Guidance of Visually Impaired Pedestrians". Ph.D. Dissertation, School of Engineering and Design, Brunel Univ., Uxbridge, UK, 2006.

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

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

[5] L.W. Farmer and D.L. Smith, "Adaptive Technology". In "Foundations of Orientation and Mobility", 2nd ed., B.B. Blasch, W.R. Wiener and R.L. Welsh, Eds. New York, USA: AFB Press, American Foundation for the Blind, 1997, pp. 231-259, ch. 7.

[6] P. Richard, G. Birebent, P. Coiffet, G. Burdea, D. Gomez and N. Langrana, "Effect of Frame Rate and Force Feedback on Virtual Object Manipulation". Presence, vol. 5 (1), pp. 95-108, 1996.

[7] PictureTel Corporation, "iPower Video Technology" (White Paper). Andower, MA, USA: PictureTel Corporation, 2001.

[8] Genista Corporation, "Video Quality Metrics". Tokyo, Japan: Genista Corporation, 2002.

[9] B. Furht, S.W. Smoliar and H. Zhang, "Video and Image Processing in Multimedia Systems. Norwell, Massachusetts, USA: Kluwer Academic Publishers.

[10] A. Richter, "Mobile Video Telephony: Test of 3G Telephones". Vällingby, Sweden: Hjälpmedelsinstitutet (HI)/The Swedish Handicap Institute (SHI), 2007. Available: www.hi.se/publicerat.

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

[12] International Telecommunication Union (ITU), "Subjective Video Quality Assessment Methods for Multimedia Applications". ITU, Geneva, Switzerland, ITU-T Recommendation P.910, 1999.

[13] International Telecommunication Union (ITU), "Interactive Test Methods for Audiovisual Communications". ITU, Geneva, Switzerland, ITU-T Recommendation P.210, 2000.

[14] J. Y. C. Chen, P. J. Durlach, J. A. Sloan and L. D. Bowens, "Robotic Operator Performance in Simulated Reconnaissance Missions". U.S. Army Res. Lab., Aberdeen Proving Ground, Aberdeen, MD, ARL Tech. Rep. ARL-TR-3628, 2005.

[15] J. French, T. G. Ghirardelli and J. Swoboda, "The Effect of Bandwidth on Operator Control of an Unmanned Ground Vehicle". In Proc. I/ITSEC, Orlando, FL, 2003.

[16] Y. Cai, "Minimalism Context-Aware Displays". Cyberpsychol. Behav., vol. 7 (6), pp. 635–644, 2004.

[17] S. Bryson, "Effects of Lag and Frame Rate on Various Tracking Tasks". In Proc. SPIE, Stereoscopic Displays Appl. IV, Bellingham, WA, 1993, vol. 1925, pp. 155-166.

[18] A. Liu, G. Tharp, L. French, S. Lai, and L. Stark, "Some of What One Needs to Know about Using Head-mounted Displays to Improve Teleoperator Performance". IEEE Trans. Robot. Autom., vol. 9, no. 5, pp. 638–648, Oct. 1993.

[19] J. B. F. Van Erp and P. Padmos, "Image Parameters for Driving with Indirect Viewing Systems". Ergonomics, vol. 46 (15), pp. 1471-1499, Dec. 2003.

[20] D. E. McGovern, "Experience and Results in Teleoperation of Land Vehicles". In "Pictorial Communication in Virtual and Real Environments", S. Ellis, M. Kaiser and A. Grunwald, Eds. London, UK: Taylor & Francis, 1991, pp. 182-195.

[21] P. N. Day, "Cognitive Effects of Delayed Visual Feedback in Real Time Operator Control Loops". In "Unpublished First Year Report". Heriot Watt Univ., Edinburgh, UK, 1999.

[22] M. Vitkovitch and P. Barber, "Effect of Video Frame Rate on Subjects' Ability to Shadow One of Two Competing Verbal Passages". J. Speech Hear. Res., vol. 37 (5), pp. 1204-1211, Oct. 1994.

[23] M. Vitkovitch and P. Barber, "Visible Speech as a Function of Image Quality: Effects of Display Parameters on Lip-Reading Ability". Appl. Cogn. Psychol., vol. 10 (2), pp. 121-140, 1996.

[24] J.Y.C. Chen and J.E. Thorpp, "Review of Low Frame Rate Effects on Human Performance". IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, vol. 37 (6), pp. 1063-1076, Nov. 2007.

Dr. Vanja Garaj was born in Koprivnica, Croatia. He holds a B.Sc. in Product Design from the University of Zagreb, Croatia (1998) and a Ph.D. in Systems Design (Human Factors) from Brunel University, UK (2006). Since 2006, he works as a lecturer in Multimedia Design at the School of Engineering and Design, Brunel University. Between 2003 and 2006, he worked as a research fellow at the SURFACE Inclusive Design Research Centre, The University of Salford, UK. Before joining academia, he worked as a designer in advertising industry.

Dr. Garaj is a member of the Institute of Engineering and Technology (UK). His research interests span human factors within the context of systems, products, environments and services design.

Dr. Ziad Hunaiti was born in Amman, Jordan in 1976. He received a Diploma in Aeronautical Communications Engineering in 1996 from Queen Noor Civil Aviation College, Jordan, a B.Sc. degree in Electrical and Electronics Engineering in 2001 form Near East University, Cyprus and a Ph.D. degree in Systems Engineering from Brunel University, UK in 2005.

He worked as an instructor between 1996 and 2002 in both Queen Noor Civil Aviation College, Jordan and Near East University, Cyprus. Between 2002 and 2008, he worked as a researcher at Brunel University, UK and Anglia Ruskin University, UK. He currently works as a lecturer at Anglia Ruskin University. He has published over 50 journal papers, conference papers and book chapters.

Dr. Hunaiti is a Fellow of the Higher Education Academy (HEA), UK. His main research interests include networking, satellite navigation systems, wireless networks, mobile information systems, m-health, m-learning, location based services and communication systems applications.

Prof. Wamadeva Balachandran (M '91–SM '96–F'04) received a B.Sc. degree from the University of Colombo, Sri Lanka in 1970 and M.Sc. and Ph.D. degrees from the University of Bradford, UK in 1975 and 1979, respectively. He is a Professor of Electronics Systems and Director of the Centre for Electronic Systems Research at Brunel University, UK. He is a visiting professor at University of Mansoura, Egypt and University of Donguan in China.

He was a visiting scholar in the School of Engineering and Applied Science, University of California, Los Angeles in 2004, sponsored by the Royal Academy of Engineering, UK. He is actively pursuing research on electrohydrodynamics, charge particle dynamics, electroaerosols, GPS navigation systems, m-health, biometrics and lab-on-a-chip technology. He has published over 275 journal and conference papers and filed twelve patent applications in these fields.

Prof. Balachandran is a member of the Electrostatic Processes Committee of the IEEE Industry Applications Society. He is a fellow of the Institute of Engineering and Technology (UK), Institute of Physics (UK), Institute of Measurement and Control (UK) and Royal Society of Arts (UK). He was the recipient of IEEE John Melcher Best Paper Award in 2000.