MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

8—33

A Visual Support System for Visually Impaired Persons Using Acoustic Interface

Yoshihiro Kawai * and Fumiaki Tomita *
Intelligent Systems Division
Electrotechnical Laboratory

Abstract

In this paper, we outline the design of a visual support system that provides three-dimensional visual information using three-dimensional virtual sound. Three-dimensional information is obtained by analyzing images captured by stereo cameras, and recognizing the objects needed by the visually impaired user. Using the three-dimensional virtual acoustic display, which relies on Head Related Transfer Functions, the user is informed of the locations and movements of objects. The user's auditory sense is not impeded as we use a bone conduction headset, which does not block out environmental sound. The proposed system is expected to be useful in the situation where the infrastructure is incomplete, and when the situation changes in real-time. We plan to experiment with it, for example, to guide users in walking and playing sports.

1 Introduction

Much of the information that humans get from the outside world is obtained through sight. Without this facility visually impaired people suffer inconveniences in their daily and social life. Therefore, much research has been done worldwide on support systems for the visually impaired [4, 7, 8, 15]. However, there are still many problems in representing the real-time information that is changing around a user.

A long cane and a seeing-eye dog are widely used as walking support devices in action operator support system. However, the range that a user can sense with a long cane is limited, and with seeing-eye dogs, there are still problems of availability and practicability. Support devices using electronic technologies have been developed, but considerable training is needed to use them. Support devices using electronic technologies have been developed, but considerable training is needed to use them. It is important to prepare the infrastructure so that users can easily understand the circumstances in their periphery [10]. Surface bumps, Braille panels, and audio traffic signals for visually disabled persons are in practical use. However, economic realities limit their use. Many problems cannot be solved only by infrastructure maintenance and development. Therefore, we aimed to develop an action support system that would provide this access by using acoustic interface three-dimensional spatial information surrounding the user.

Among other visual aid systems using sound so far investigated are a system using ultrasonic waves [16], one that displays images by differences of frequency pitch and loudness [5, 11], one which utilizes a speaker array [12], and one which utilizes stereophonic effects [9]. However, the target of these systems is mainly two-dimensional space. On the other hand, three-dimensional sound can provide more real-world information because it includes an intuitive feeling of depth and the feeling of the front and back. There is a GUI access system using three-dimensional sound [1, 3], but it shows only the relative position between windows, and three-dimensional sound is not being utilized fully.

We are developing a support system that displays three-dimensional visual information using three-dimensional virtual sound. This is unique in that three-dimensional environmental information is obtained for the task that the user sets, and it is represented by three-dimensional virtual sound. Images captured by small stereo cameras are analyzed in the context of a given task to obtain the three-dimensional structure, and object recognition is performed. The results are then conveyed via three-dimensional virtual sound to the user. This system is expected to be useful in the situation where the infrastructure is incomplete, and when the situation surrounding the user is changing in real-time. In addition, this system can be used without much learning because it provides information by a virtual sound superposed to the actual environment sounds. This method would not replace or impede their existing auditory sense. We are assuming that it could be used, for example, to assist in walking and in playing sports, as shown in Figure 1. In this paper, we describe the details of our prototype system, three-dimensional visual information processing methods, experiments of three-dimensional sound interface.

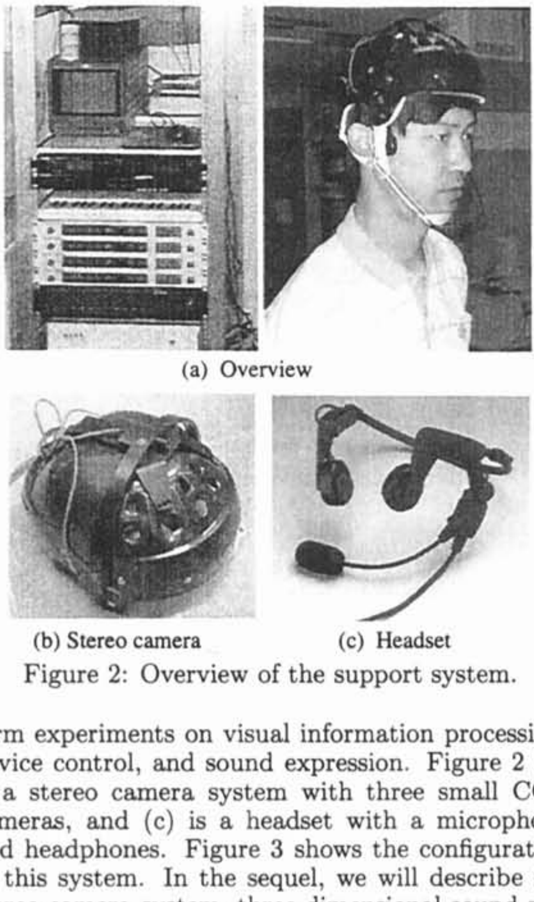
2 System

We have built a prototype system shown in Figure 2 (a) to develop a visual support system to per-

*Address: 1-1-4 Umezono, Tsukuba, Ibaraki 305-8568 Japan. E-mail: {kawai|tomita}@etl.go.jp

(a) Walking (b) Playing sports

Figure 1: Application example.



(a) Overview

(b) Stereo camera (c) Headset

Figure 2: Overview of the support system.

form experiments on visual information processing, device control, and sound expression. Figure 2 (b) is a stereo camera system with three small CCD cameras, and (c) is a headset with a microphone and headphones. Figure 3 shows the configuration of this system. In the sequel, we will describe the stereo camera system, three-dimensional sound system, three-dimensional visual information processing method, system control, and virtual sound expression of visual information.

2.1 Stereo camera system

We use small CCD cameras to obtain information on objects and visual environments. The captured images for recognition are analyzed. Advantages of this method are that it is suited for measurement of very distant objects (e.g., identification of the red/green light of a traffic signal from far away) and for character information readability (e.g., characters on a signboard). Though it is still difficult to analyze images to obtain three-dimensional visual information, there exists a potential use of recent pattern recognition techniques in our application. For example, we have been developing a vision system VVV (Versatile Volumetric Vision), which is a general-purpose system and can be used for many purposes in many fields [14]. This system enables analyzing stereo images captured by three cameras, reconstruction of three-dimensional objects in the target scene, recognition by matching with models, and tracking of moving objects.

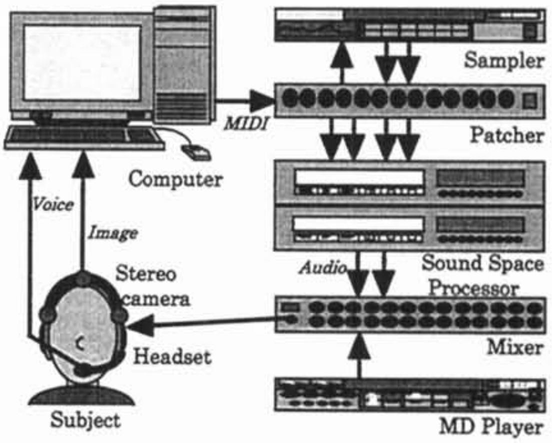


Figure 3: Composition of the system.

It is desirable for the visual information input unit to be small and light because the device will be mounted on the user's head. However, high performance is requested in order to get accurate measurements. As a result, we mounted three small CCD cameras fixed with an aluminum frame on a helmet (shown in Figure 2 (b)). The camera has a 1/4-inch color CCD that captures a 640×480 pixel image. The diameter is 7 mm and the weight, including the 3.5 m cable, is 68 g. The total weight of the helmet is about 650 g. We have set the focus of the lens 3 m, a point to which a long cane can not reach. The reason for using three cameras is to reduce the calculation of the correspondence problems on horizontal lines during the stereo image analysis.

2.2 Three-dimensional virtual sound system

Recently, with the development of virtual reality technologies, the technical progress for acoustics in virtual space is remarkable. We can use three-dimensional virtual sound easily, since some three-dimensional sound equipment has already been produced commercially. We have assembled built our acoustic system using mainly the sound space processor RSS-10 made by Roland corporation (the left side in Figure 2 (a)). This is a device by which an arbitrary three-dimensional virtual sound space is calculated on the basis of Head Related Transfer Functions (HRTFs). For the output device, we selected bone conduction headphones, which do not entirely cover the user's ears, and therefore have little influence for him/her on hearing and understanding environmental sounds.

2.3 Three-dimensional visual information

Measurement and recognition of three-dimensional objects in the target scene is done by analysis of

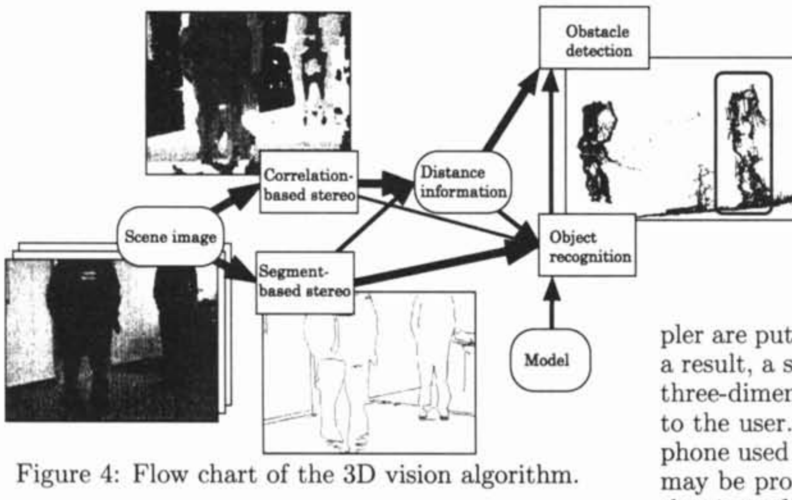


Figure 4: Flow chart of the 3D vision algorithm.

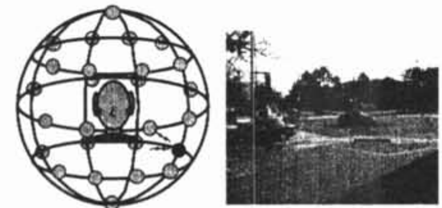
stereo images. The flow of the process is shown in Figure 4, and is an integration of a correlation-based method and a segment-based method.

Distance information is obtained by the correlation-based method, which is a technique to calculate the disparities between stereo images using the fact that correlation values of intensity at the same place are higher. There is a weak point in that it takes a long processing time if the search range is not limited, however, owing to its simple algorithm, it is possible to process it in real time by special hardware. The three-dimensional data obtained are comprised of sets of points, and a structuring process such as segmentation is needed. On the other hand, the segment-based method is an algorithm for reconstructing three-dimensional wire-frames by correspondence of boundary edges [6]. First, some special features, such as segments, are extracted and a correspondence search for them is performed. This is a complicated procedure, but it is a superior method for structure reconstruction and recognition of three-dimensional objects. By combining these two methods, the three-dimensional structured data with surface information is obtained.

After acquiring the three-dimensional data in the stereo vision, the object recognition process follows [13]. The three-dimensional data are matched with object models in a database to identify what objects are present and to know their status. Users can know information on the object needed for performing a task. In addition, gaps or obstacles are detected using depth information. The results of these processes are transmitted to the virtual sound system.

2.4 System control

We will explain the design of the whole system, as shown in Figure 3 . Three images captured by the stereo cameras are sent to a computer, and are then analyzed by the process mentioned above. After having had the three-dimensional data reconstructed, the user's targets are recognized and tracked. Both the status of the objects obtained and the sound that is assigned to each object from a sampler are put input to the sound space processors. As a result, a sound image for each target is mapped in three-dimensional virtual sound space and carried to the user. In addition, we may also have a microphone used by the user to set a task by voice, which may be processed by a voice recognition engine. At the time of writing this manuscript, it is unfinished.



(a) Virtual sound source locations. (b) Environment sound recorded at the crossing.

Figure 5: Experiment setup.

2.5 Sound display of visual information

Not all visual information is converted to auditory information. Sounds only about targets, which are needed to do the task, are output. However, for obstacles or in dangerous situations, such as stairs, walls, or cars, an alarm or voice is output to alert the users to the dangerous objects.

Regarding the output sound, the same sound is used if it exists in the real world. Otherwise, a sound that the user can easily recognize is assigned. It seems important to create and superpose a sound space that is the same as a real environment as much as possible. Users can change these settings and parameters as they wish.

3 Experiment

We conducted some basic tests on the auditory localization of three-dimensional sound in a virtual space in order to develop an acoustic interface. First, we conducted some preliminary experiments for four visually impaired persons and six university graduate students with eye masks. In these experiments, we found that (1) open-air type headphones were better than closed-air type headphones and the earphone type as an output device, (2) there were some mis-recognitions between the front and back directions, and up and down directions, and (3) active action, which means moving their head freely, was superior to intuitive judgment in order to recognize sound images precisely.

We performed on two sound localization tests to investigate the influence of environment sound. In these experiments, we used bone conduction headphones as an output device. The subjects were three people (two men and one woman). The sound locations for all directions in the virtual sound space are shown in Figure 5 (a). Twenty-six sound sources were arrayed on the globe with a radius of 3.0 m. They were located at both (north and south) poles, and at 8 points each on a horizontal plane and cross horizontal planes at ± 45 degrees, at 45-degree intervals. The height of the center from the floor was 3.2

m. The sound was presented at random 52 times. Each sound source was assigned a unique number. The subjects answered the number that they felt, after having learned the correspondence between numbers and directions in advance. As a sound source, a similar sound to that at audio traffic signals was used. The environment sounds which included the sounds of automobiles, was recorded at a pedestrian crossing as shown in Figure 5 (b).

We performed two sound position recognition tests, without the environment sound (Case A), and with the sound (Case B). Table 1 (a) and (b) shows the rates of correct answers and recognition time for each of Case A and B. T_n is the rate of correct answers according to the following:

T_1: rate when the answer completely matched the actual direction.

T_2: rate when the answer was taken as correct if its distance was within one position from side to side and up or down.

The average of completely correct answers at T_1 was very low (in Case A only 14.1%, in Case B 9.6% less than A). This result shows that the environment sound disturbed the sound image localization, the two main reasons for which are considered to be:

1. The influence of narrow range of bone conduction headphones. The effective frequency band is only 30 to 3000 Hz at more than 80 dB.

2. The problem of sound source. We used a simple sound, which was similar to the sound at a sound box, instead of 10 kHz pink noise, which is popular with sound image localization, similar to environment sounds, and easy to recognize. The frequency band is narrower than pink noise, so it was more difficult to localize sound images.

Table 1 (c) shows an analysis of subject A's errors. He mis-recognized to almost the same degree in both Case A and Case B. However, the errors between front and back directions in Case A were bigger than in Case B. It is assumed that the influence of the environment sound is large for sound position recognition between front and back directions. It is known that frequency is related to the recognition of front and back directions [2], so the difference of frequencies should be emphasized to improve recognition between the front and back directions.

Table 1: Results of experiments.

(a) Case A: without environment sound

Subject	T1*	T2*	Time
A	11.5%	48.1%	2.75 sec.
B	19.2%	53.8%	3.67 sec.
C	11.5%	38.5%	8.43 sec.
Avg.	14.1%	46.8%	4.95 sec.

*Tn: Tolerance n

(b) Case B: with environment sound

Subject	T1*	T2*	Time
A	13.5%	44.2%	2.94 sec.
B	5.8%	34.6%	4.57 sec.
C	9.6%	36.5%	9.49 sec.
Avg.	9.6%	38.4%	5.67 sec.

(c) Analysis of mis-recognized answers of subject A

Subject A	Error	Error contents		
		Front&Back	Up & Down	Others
Case A	51.9%	14.8%	59.3%	33.3%
Case B	55.8%	41.4%	51.7%	20.7%

4 Conclusions

We are developing a recognition support system using three-dimensional virtual sound for visually impaired persons. In this paper, we described the design of our prototype system, three-dimensional visual and auditory information processing methods, and some basic experimental results on our acoustics. Problems in sound image localization were clarified through the experiments. For example, we found that there were many mis-recognitions between back and front, and up and down directions, and an active action of the subject is necessary to recognize the virtual sound locations correctly.

In the future, we will first, complete it as an on-line system. We will take these experimental results into account and will develop an acoustic interface so that the rates of sound image localization are improved, for example, by altering the sound source frequencies for specified locations and directions that are likely to be more mis-recognized. Moreover, we will develop a user interface which allows task setting by voice, and investigate the influence of virtual sound superposed to the environment sounds.

References

[1] T. Aritsuka, N. Hataoka: GUI representation system using spatial sound for visually disabled, Proc. of ASVA'97, pp.415–420, 1997.

[2] J. Blauert: Spatial Hearing (Revised Edition), The MIT Press, 1996.

[3] P. Gorny: Typographic semantics of webpages accesible for visually impaired users – Mapping layout and interaction objects to an auditory interaction space –, Proc. of ICCHP 2000, pp.251–257, 2000.

[4] H. Ichikawa, H. Ohzu, S. Torii, T. Wake: Visual-Sense Disability and Technology to Aid It, Nagoya University Press, 1984 [in Japanese].

[5] T. Ifukube, T. Sasaki, C. Peng: A blind mobility aid modeled after echolocation of bats, IEEE Trans. BME-38, 5, pp.461–465, 1991.

[6] Y. Kawai, T. Ueshiba, Y. Ishiyama, Y. Sumi, F. Tomita: Stereo Correspondence Using Segment Connectivity, Proc. of ICPR'98, I, pp.648–651, 1998.

[7] K. Koshi, H. Kani, Y. Tadokoro: Orientation Aids for the Blind Using Ultrasonic Signpost System, The 1st Joint Meeting of BMES and EMBS, pp.587, 1999.

[8] J. M. Loomis, R. G. Golledge, R. L Klatzky., J. M. Speige, J. Tietz: Personal guidance system for the visually impaired, Proc. of ASSETS'94, pp.85–91, 1994.

[9] J. M. Loomis, C. Hebert, J. G. Cicinelli: Active localization of virtual sounds, J. Acoust. Sot. Am., 88, 4, pp.1757–1764, 1990.

[10] H. Matsubara, K. Goto, S. Myojo: Development of Guidance System for Visually Impaired Persons, Railway Technical Research Institute, 13, 1, pp.31–36, 1999 [in Japanese].

[11] P. B. L. Meijer: An Experimental System for Auditory Image Representations, IEEE Trans. Biomed. Eng., 39, 2, pp.112–121, 1992.

[12] M. Shimizu, K. Itoh, Y. Yonezawa: Operational Helping Function of the GUI for the Visually Disabled Using a Virtual Sound Screen, Proc. of ICCHP'98, pp.387–394, 1998.

[13] Y. Sumi, Y. Kawai, T. Yoshimi, F. Tomita: Recognition of 3D free-form objects using segment-based stereo vision, Proc. of ICCV'98, pp.668–674, 1998.

[14] F. Tomita, T. Yoshimi, T. Ueshiba, Y. Kawai, Y.Sumi: R&D of Versatile 3D Vision System VVV, Proc. of SMC'98, TP17-2, pp.4510–4516, 1998.

[15] D. H. Warren, E. R. Sttelow: Electronic Spatial Sensing for the Blind, Martinus Nijhoff Publishers, pp.35–61, 1985.

[16] http://www.sonicvision.co.nz/