35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A Tactile Vision Substitution System for the Study of Active Sensing*

Brian Hsu, Cheng-Han Hsieh, Sung-Nien Yu, Ehud Ahissar, Amos Arieli, and Yael Zilbershtain-Kra

Abstract—**This paper presents a tactile vision substitution system (TVSS) for the study of active sensing. Two algorithms, namely image processing and trajectory tracking, were developed to enhance the capability of conventional TVSS. Image processing techniques were applied to reduce the artifacts and extract important features from the active camera and effectively converted the information into tactile stimuli with much lower resolution. A fixed camera was used to record the movement of the active camera. A trajectory tracking algorithm was developed to analyze the active sensing strategy of the TVSS users to explore the environment. The image processing subsystem showed advantageous improvement in extracting object's features for superior recognition. The trajectory tracking subsystem, on the other hand, enabled accurately locating the portion of the scene pointed by the active camera and providing profound information for the study of active sensing strategy applied by TVSS users.**

I. Introduction

The visually impaired people usually need to struggle in their daily lives to explore their outside environment. If more information of the environment can be provided in some way with the assistance of recent technology, then they could adapt to the environment more rapidly, via build-in natural learning processes [1, 2].

Sensory Substitution for the visually impaired was first introduced by Bach-y-Rita and coworkers [3-5]. They invented the first sensory substitution system, which is a chair that allow blind to "see" via tactile actuators attached to their back and activated by a video camera. Later versions used other body parts with the latest version including actuator matrix placed on blind's tongue. The information was captured by the camera and fed back to the tactile device. This device is known as a tactile vision substitution system (TVSS) which offers people who suffer from lacking sight an opportunity to make a change.

A TVSS translates visual input, usually from a video camera, into the output of a tactile stimulation array. With the assistance of digital image processing techniques, the most significant features of the image could be extracted and provided as augmented sensation to the visually impaired. As a sequel, the users could more accurately distinguish foreground objects from the background.

*Research supported by the National Science Council and the Ministry of Education, Taiwan, Republic of China and the Ministry of Science and Technology, Israel.

Brian Hsu, Cheng-Han Hsieh, and Sung-Nien Yu are with the Department of Electrical Engineering and the Advanced Institute of Manufacturing with High-tech Innovations, National Chung Cheng University, Chiayi County, Taiwan (corresponding author's phone: +886-5-2720411 ext 33205; fax: +886-5-2720862; e-mail: ieesny@ccu.edu.tw).

Ehud Ahissar, Amos Arieli, and Yael Zilbershtain-Kra are with the Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel (e-mail: ehud.ahissar@weizmann.ac.il; Amos.Arieli@weizmann.ac.il; yaelzilb@gmail.com).

(a) (b) (c)

Figure 1. TVSS system setup. (a) VTMouse ; (b) VTMouse with the active camera and pointer; (c) active sensing experiment.

Since the idea of sensory substitution was introduced, researchers gradually recognized the importance of actively moving the camera in the use of the TVSS. Because tactile resolution in the fingers is far more limited than the visual resolution (and so is the resolution of the arrays of tactile actuators relative to the number of process is termed "active sensing." [6] The knowledge about how people develop effective strategies to actively sense the environment is also an important issue in the development of a TVSS.

This study stem from a cooperation project about sensory substitution between the Laboratory for the Study of Adaptive Perceptual Processing directed by Prof. Ehud Ahissar and the Active Sensing Laboratory directed by Dr. Amos Arieli at the Weizmann Institute of Science in Israel and the Biomedical Signal Processing and System Design Laboratory directed by Prof. Sung-Nien Yu at the National Chung Cheng University in Taiwan. The Israeli team set up experiments for active sensing with a TVSS while the Taiwanese team developed image processing algorithms aiming to enhance active sensing performance of the experiments. A trajectory tracking algorithm was jointly developed to understand participants' active sensing strategies.

This system contains two parts, namely (1) image processing and (2) trajectory tracking. The image processing part converts the color images acquired from the camera into lower resolution binary images with valuable features reserved, which designated to generate adequate output for the tactile device. The trajectory tracking part, on the other hand, tracks the trajectory of the participant on the stimuli and provides information for the study of active sensing strategies to explore the environment, without vision, using only TVSS.

II. System Overview

The TVSS contains a tactile stimulation device and a camera. Tactile stimuli was provided by the VTMouse (Tactile World, Ra'anana, Israel) , as shown in Fig. 1 (a). The VTMouse is a standard size computer mouse for the blind [7], which consists three tactile stimulation arrays of 32 pins each (4x8) and provide tactile stimuli to the fingers at different heights (4 levels).

A miniature video camera (active camera) (VQ25B-P37P; Filtech Corp., Yangchon-Gu Seoul, Korea) was attached to the VTMouse, as depicted in Fig. 1 (b) as the visual input sensor. In parallel to the miniature camera is a red laser pointer (605nm) which provides a marker associated with the location of the TVMouse. A wide view camera (fixed camera; 1280x1024, RGB, 15Hz) was arranged in a fixed location (Fig. 1 (C) to the left of the participant). With this arrangement, the movement of the active camera is identifiable in the fixed wide image taken by the fixed camera for further analysis.

III. Image Processing Algorithm for the TVSS

Figure 2 shows the block diagram of the imaging processing algorithms developed for the TVSS. The color video frames acquired by the CCD camera, originally represented with red, green, and blue (RGB) attributes, were first transferred into hue, saturation, and intensity (HSI) color space [8]. Only the intensity part of the frame was reserved and represented as the grey-levels of the image. Image enhancement with histogram equalization [8] followed to make the foreground objects more separable form the background. The enhanced image needed to be further processed with downsampling, low-pass filtering, thresholding, and morphological process in order to generate suitable output for the tactile stimulation array. These methods are explained separately as follows.

A. Discrete Wavelet Transform for Downsampling

The tactile stimulation array had only 96 (12x8) pins as output. Compared to the image frame with a size of 640x480, the tactile array was far less in size. Therefore, the array could only display a very small part of an image, or, alternatively, the image could be shrunk to smaller resolution, a process termed downsampling. In our previous works [7-9], we have demonstrated the high capability of discrete wavelet transform (DWT) in downsampling an image into a quarter of the original size while preserving the most important features. Therefore, we also applied DWT as a downsampling processor in the study.

Figure 3 (a) shows the procedure of a two-dimensional DWT. With low-pass (*h[n]*), high-pass (*g[n]*) filters, and down-sampling operator (↓2), each row of the input image *x[n]* is firstly separated into the low-frequency (L) and high-frequency (H) parts. The similar column operators proceed to separate each column into low-frequency (L) and high-frequency (H) parts. In this manner, the image is separated into four subbands, namely LL, LH, HL, and HH, with different frequency attributes, as depicted in Fig. 3 (b).

After two dimensional DWT, an image is divided into four subband components of the same size (Fig. 2(b)). The high-frequency bands contain rapidly changing information such as noises and edges. The low-frequency bands, on the other hand, contain the features with low variety such as the shape of an object. The LL part was reserved as the downsampled version of the original image [8]. Moreover, since the property of the low-pass and high-pass filters used in the DWT is determined by the mother wavelet, the mother wavelet Rbio6.8 was empirically chosen to reserve the most image power after downsampling.

Two downsampling processes were used in the study. The first downsampler was used to reduce the size of original image such that the computational load of the following process was reduced. The second downsampler was applied after all the image processing had been done and then further reduced the image into a size of 40x30 to better fit the dimension of the tactile device.

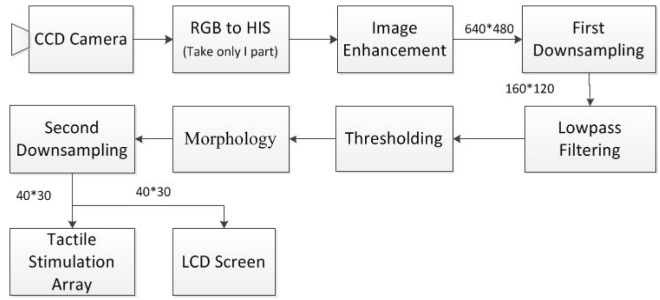
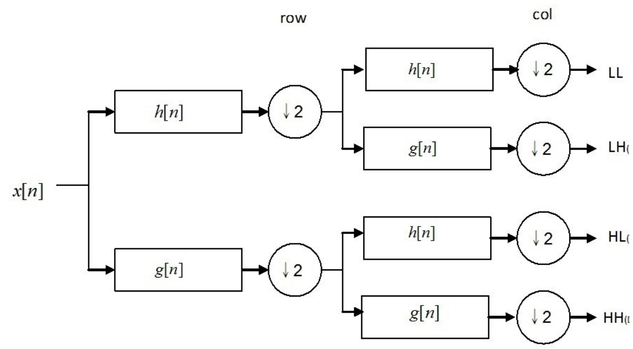
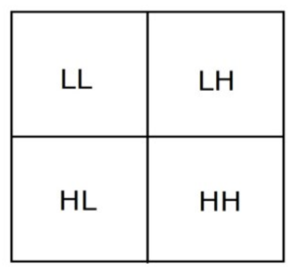


Figure 2. Diagram of the image processors for the TVSS.



(a)



(b)

Figure 3. Two dimensional discrete wavelet transform (DWT). (a) concept of 2D DWT; (b) different subband components after 2D DWT.

B. Low-pass Filtering and Thresholding

A 5x5 mask average filter [6] was used as the low-pass filter to eliminate noise and smooth abnormal edges in the image. A thresholding operator followed to convert a grayscale image into a binary image. The threshold value was empirically determined to be 0.6 which resulted in the best performance.

C. Morphology Methods

After down-sampling, low-pass filtering, and thresholding, the acquired color image was transformed into a down-sampled binary image. However, shape defects and missing areas of objects were to be fixed. Morphology methods [10], including dilation and erosion, were employed to tackle this problem.

Dilation is the process to gradually enlarge the boundaries of regions of foreground pixels. The function is expressed as

$$D = B \oplus S = \{x | (\hat{S})_x \cap B\} \quad (1)$$

where B represents structuring elements of dilation operator and S represents the image which is to be dilated. With this operator, the areas of foreground pixels grow in size and holes within those regions become smaller.

Erosion is the inverse of dilation. It erodes away the boundaries of regions of foreground pixels. The function is expressed as

$$E = B \ominus S = \{x | (S)_x \subseteq B\} \quad (2)$$

where B represents structuring elements of erosion operator and S represents the image which is to be eroded. With this method, areas of foreground pixels shrink in size and holes within those areas become larger.

The dilation operator was first applied to connect broken lines and holes inside objects. The erosion operator followed to erode the boundary regions back to the original size. With these operators, holes inside the image were filled.

IV. TRAJECTORY TRACKING ALGORITHM

A. *Diagram of the Trajectory Tracking Algorithm*

Figure 4 is the block diagram of the trajectory tracking procedure. Before tracking, the acquired images from the two cameras were downsampled to reduce the computational load of the following process. The laser pointer was used for identifying the location of the participant on the wide view image. Since the laser light may sometimes scatter and cause errors in tracking, we firstly identify the location of the red spot recorded by the fixed camera based on the difference image of the present and previous frames. The location was considered the probable center of the images acquired by the active camera. We then searched for the real center in the vicinity of the location for the most similar image region acquired by the fixed camera compared to the active camera image using the maximal index calculated from the crosscorrelation function [8]. As a result, the strategy used by the participants could be analyzed through tracking the trajectory of the participants on the stimuli during active sensing tasks [12].

B. *Correlation Coefficient as a Similarity Measure*

In trajectory tracking, we had to measure the similarity between the active camera image and image blocks of the fixed camera image. Pearson correlation coefficient [6], which measures the linear correlation of two variables, was chosen as a similarity measure. The correlation coefficient index is defined as

$$r = \frac{\sum_m \sum_n (\mathbf{A}_{mn} - \bar{\mathbf{A}})(\mathbf{B}_{mn} - \bar{\mathbf{B}})}{\sqrt{(\sum_m \sum_n (\mathbf{A}_{mn} - \bar{\mathbf{A}})^2)(\sum_m \sum_n (\mathbf{B}_{mn} - \bar{\mathbf{B}})^2)}} \quad (3)$$

where variables **A** and **B** are two different arrays, m and n are the indexes of the array elements in the row and the column, respectively, and $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are the means of **A** and **B**, respectively. The value of the correlation coefficient index r is between -1 and 1. The higher the r value is the more similar **A** and **B** are.

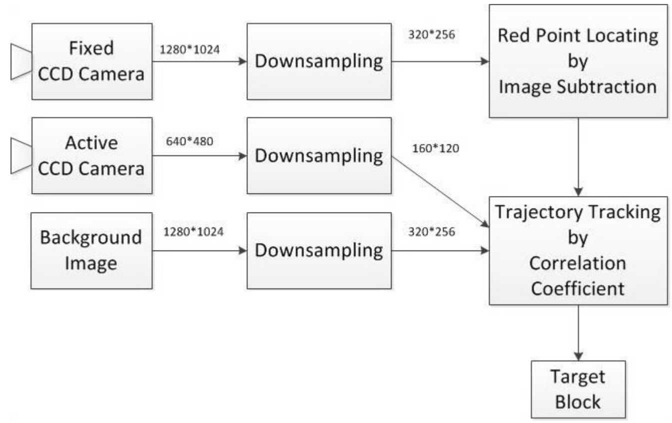


Figure 4. Diagram of trajectory tracking.

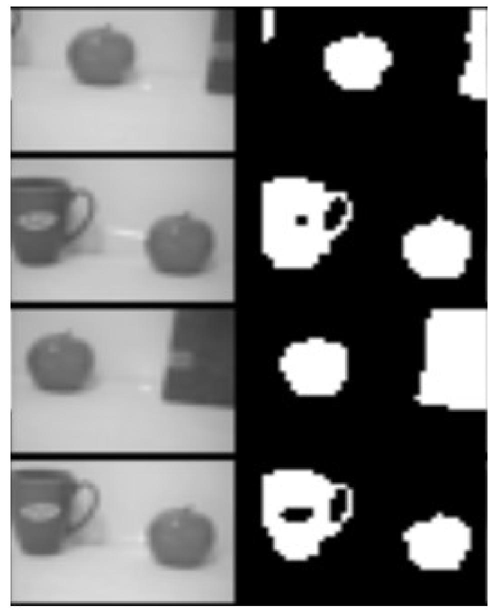


Figure 5. Result of the TVSS image processing algorithm; left column: downsampled grayscale image; right column: binary image after processing.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

To test the performance of the system, active sensing experiments were conducted in the laboratory at Weizmann Institute of Science in Israel. The participants were eye masked and asked to identify the objects placed in front of a blank white wall. The responses of the participants were recorded by the active camera for analysis.

Figure 5 shows one result of the image processing algorithm for the proposed TVSS. The left column shows the downsampled grayscale image frames. The second column shows the binary images after processing, which are fed into the tactile device as input. Three objects, namely a cup, an apple, and a book, were observed in the recorded sequence. It is interesting to note that, although reflection of light on the smooth surface of objects may sometime produce artifacts in the grayscale image, the application of image processing techniques is able to fill the holes in the image and loosen the problem. Although the mark on the cup sometimes appears as a hole in binary image, the size is significantly reduced and can be compensated by active sensing (see below). As a result, all the three objects in the grayscale image were vividly transferred into downsampled binary image with little loss of shape information.

The records from both fixed and active cameras were analyzed by the trajectory tracking algorithm. Figure 6 demonstrates one of the results. The left column shows images acquired from the active camera. The middle column shows the most similar image blocks acquired by the fixed camera, as shown in the right column, which included images acquired from the fixed camera with the most similar regions marked. It can be seen that images acquired from active and fixed cameras differ in some sense because of distinct shot angles and orientations. However, the proposed trajectory tracking algorithm for the TVSS constantly succeeded in locating the most similar blocks from the fixed images.

After locating the image blocks from the fixed camera that is most similar to that from the active camera, the trajectory of the active camera movement could be tracked. Figure 7 shows one of the results. The red points in the figure show the center points of the located most similar images to the active images. The result shows that the participant moves her hand back and forth trying to identify objects in the scene according to the tactile stimuli generated from the device. In this case, the participant moved more frequently in the horizontal direction than in the vertical direction searching for objects. After she "felt" an object, she focused on the edges as an attempt to identify the objects. The trajectory tracking provides an insight into the participants' behavior and strategy in using the TVSS for exploring the surrounding environments.

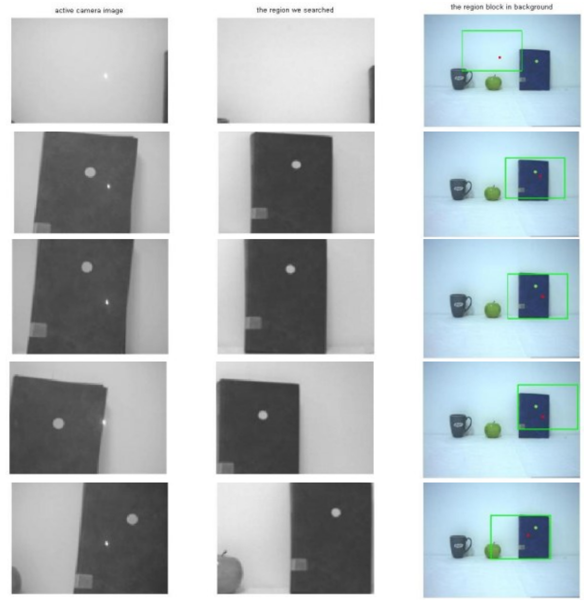


Figure 6. Results of TVSS trajectory algorithm; left: images from the active camera; middle: most similar image blocks in the fixed camera; right: images from the fixed camera with the marked most similar regions.

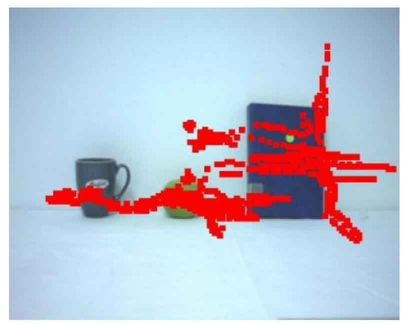


Figure 7. Trajectory of the active camera

VI. CONCLUSION

This paper presents the preliminary result of the Tactile Visual Substitution System developed for the study of active sensing. Two subsystems were developed. The image processing subsystem solves the problems, such as high-to-low resolution transformation, light reflection on the surface of objects, and texture artifacts, usually encountered by traditional TVSS. The trajectory tracking subsystem, on the other hand, enables the researchers to trace the movement of the tactile device which reflects the strategies employed by the TVSS user to explore the environment. The preliminary results of the study demonstrate the capability of the system in offering a more effective TVSS and a more powerful diagnosis system for the study of active sensing of the visually impaired.

ACKNOWLEDGMENT

This study was supported in part by the grant NSC 99-2923-E-194-002-MY2 from the National Science Council, Taiwan, a grant from the Ministry of Education, Taiwan, R.O.C. and the grant 710863 from the Ministry of Science and Technology, Israel.

REFERENCES

[1] E. Ahissar, M. Abeles, M. Ahissar, S. Haidarliu, and E. Vaadia, "Hebbian-like functional plasticity in the auditory cortex of the behaving monkey," *Neuropharmacology*, vol. 37, pp. 633-655, 1998.

[2] V. Ego-Stengel, D. E. Shulz, S. Haidarliu, R. Sosnik, and E. Ahissar, "Acetylcholine-dependent induction and expression of functional plasticity in the barrel cortex of the adult rat," *Journal of Neurophysiology*, vol. 86, pp. 422-437, 2001.

[3] P. Bach-y-Rita, C. C. Collons, F. A. Saunder, B. White, and L. Scandden, "Vsion Substitution by Tactile image projection," *Nature*, vol.221,, pp. 963-4, 1969.

[4] P. Bach-y-Rita and S. W. Kercel,"Sensory substitution and the human-machine interface," *Trends in Cognitive Neuroscience,* vol.7, no.12, pp.541-546, 2003.

[5] P. Bach-y-Rita and K. A. Kaczmarek, "Tongue placed tactile output device," *US Patent* 6,430, 450, 2002.

[6] L Renier and A. G. De Volder, "Cognitive and brain mechanisms insensory substitution of vision: a contribution to the study of human perception," *J Inter Neurosci* 4, pp. 489-503, 2005.

[7] http://www.tactile-world.com/.

[8] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., New Jersey: Prentice Hall, 2007.

[9] S.-N. Yu, and C.-N. Lin, "An Efficient Paradigm for Wavelet-based Image Processing Using Cellular Neural Networks," *International Journal of Circuit Theory and Applications*, vol. 38, no. 5, pp. 527-542, June 2010.

[10] C.-N. Lin and S.-N. Yu, J.-C. Hu, "Image Processing for a Tactile /Vision Substitution System Using Digital CNN," *28th Annual International Conference of IEEE Engineering in Medicine and Biology Society*, Aug. 30-Sep. 3, 2006, New York, USA.

[11] http://homepages.inf.ed.ac.uk/rbf/HIPR2/morops.htm.

[12] A. Saig, G. Gordon, E. Assa, A. Arieli, E. Ahissar, "Motor-sensory confluence in tactile perception," *Journal of Neuroscience*, vol. 32, pp. 14022–14032, 2012.