HAPTIC ACCESS TO CONVENTIONAL 2D MAPS FOR THE VISUALLY IMPAIRED

Konstantinos Kostopoulos, Konstantinos Moustakas, Dimitrios Tzovaras, Georgios Nikolakis

Informatics and Telematics Institute, P.O. Box 361, 57001 Thermi-Thessaloniki, Greece

ABSTRACT

This paper describes a framework of map image analysis and presentation of the semantic information to blind users using alternative modalities (i.e. haptics and audio). The resulting haptic-audio representation of the map is used by the blind for navigation and path planning purposes. The proposed framework utilizes novel algorithms for the segmentation of the map images using morphological filters that are able to provide indexed information on both the street network structure and the positions of the street names in the map. Next, off-the-shelf OCR and TTS algorithms are utilized to convert the visual information of the street names into audio messages. Finally, a grooved-line-map representation of the map network is generated and the blind users are able to investigate it using a haptic device. While navigating, audio messages are displayed providing information about the current position of the user (e.g. street name, cross-road notification e.t.c.). Experimental results illustrate that the proposed system is considered very promising for the blind users and has been reported to be a very fast means of generating maps for the blind when compared to other traditional methods like Braille images.

Index Terms— Morphological filters, connected operators, haptic interaction, multimodal maps, blind users.

I. INTRODUCTION

The human brain utilizes complex, still unknown procedures in order to perform intuitive tasks such as to decode the information stored in maps. These procedures are very difficult to imitate using computers. It is obvious that all common maps are perceived using the visual modality thus making maps inaccessible for special population categories like the visually impaired. Moreover, since maps are the major means of navigating into unknown spaces, it is more than clear that the visually impaired are not able to use this means.

There have been many research efforts dedicated to the assistance of the navigation of the visually impaired. The University of Michigan provides the visually impaired students with a full tactile map of the University Campus [1]. This specific tactile map, based on the Braille code, provides not only navigational assistance but also useful information concerning campus functionality like the courses program.

Moreover, in [2] a PDA able to recognize road signs or even text from natural figures is proposed. The PDA captures natural figures and detects the text areas. Characters are segmented and a respective aural message is composed and displayed to the blind user, providing him/her with significant information about the surrounding space.

Furthermore, in [3] an automated indoor navigation system dedicated to the visually impaired is proposed. The above system utilizes a camera mounded to the user to capture images from the surroundings. Then through image process the obtained data are analysed by a computer which constructs the local map. Simultaneously the computer follows user's movement and notifies the latter about the presence of POIs like door handles or turning points.

Additionally, in [4] a system developed for the training of the visually impaired is proposed. This system allows visually impaired, to study and interact with various objects in specially designed virtual environments, while allowing designers to produce and customize these configurations. Cane simulation [4] is one of the scenarios that the above system has been used to.

Finally, in [5] a framework for converting 3D maps into haptic representations has been proposed.

The goal of the framework proposed in this paper is to provide the visually impaired with an easy to use means of accessing conventional 2D maps. The user can interact with the produced 3D model and examine the properties of the virtual replica of existing maps.

II. FRAMEWORK OVERVIEW

Figure 1 illustrates the general architecture of the developed platform. It uses as input a still image of a conventional 2D map. The output of the system is a haptic-aural representation of the 2D map. Four main modules can be identified in the system. In particular:

1) Recognition of street names and points of interest (POIs).
2) Recognition of the road network structure
3) Correspondence estimation between roads and names
4) Generation of the multimodal map.

A number of map prerequisites that the civil map should meet have been carefully chosen in order to design a system that will not depend on the map provider. In particular:

Color constraints: Street names should be represented using a dark color in order to be discernible from the rest of the map and thus advancing the process of their recognition.

Positioning: Street names should be located inside the associated road so as to attain their correspondence.

Resolution: Map resolution should be adequate to utilize an OCR algorithm to retrieve street names.

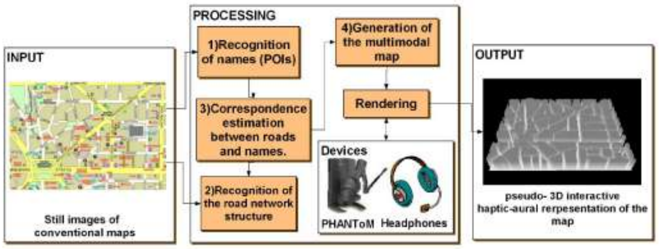


Fig. 1. Architecture of the proposed system.

Special Symbols: Symbols that represent points of interest such as hospitals, churches, parking e.t.c. should be thoroughly defined by the map provider.

III. MAP IMAGE ANALYSIS

The proposed map image analysis focuses on the extraction of the three major map components, street names, road network structure and points of interest(POIs).

III-A. Extracting the semantic information

For the detection of the street names, the primary map (Figure 2a) is successively subjected to erosion (Figure 2b) and dithering to two colors (Figure 2c). Next, the produced image is segmented by applying region growing (Figure 2d). The retrieved regions represent street names, points of interest and noise.

III-A.1. Extracting points of interest (POIs)
The recognition of points of interest is based on the Angular Partitioning of Abstract Image (APAI)[6],[7]. The latter estimates the resemblance between the templates of the POIs and the retrieved regions. Note that the above matching method is scale and rotation invariant. Figure 3 indicatively illustrates the templates of four symbol.

III-A.2. Extracting street names
After recognizing the POIs, the remaining regions represent street names and noise. The system discards all regions that are too small to represent a street name (Figure 2d). Finally off-the-shelf OCR [2] and TTS [8] algorithms convert the visual information of the street names into aural messages.

III-B. Estimating the Road Network Structure

For the estimation of the road network structure the system initially discards all street names from the primary map. Next connected operators [9],[10] and [11] are used to process the image as described in the sequel.

III-B.1. Connected Operators
Consider image I of Figure 4a. In order to discard region B the general idea is to gradually diminish region B while retaining the rest regions of I. Therefore :

I) All regions of I are diminished (Figure 4b) using the dilation operator $\delta_c(I)$.
II) Consider $\epsilon_c(I)$ as the erosion of I. To retain regions A and C but not region B the algorithm applies operator $Max()$ on the images $\epsilon_c(I)$ and I.

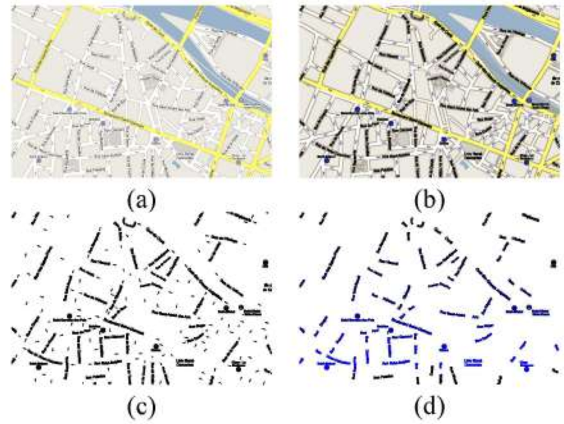


Fig. 2. (a) Primary map image, (b) erosion, (c) dithering, (d) retrieved regions.

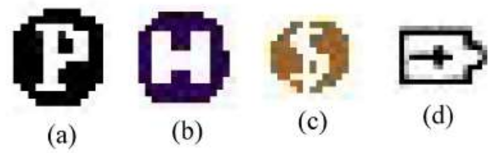


Fig. 3. Symbol templates for : (a) parking, (b) hotel, (c) bank, and (d) church.

Theoretically the second step is iteratively repeated. The intermediate images g_k of the k-th iteration step are recursively computed by equation:

$$g_k = Max(\epsilon_c(g_{k-1}), I). \quad (1)$$

In the map case anti-extensive connected operators [10] are applied so as to enhance roads while diminishing street names.

Summarizing, the road sketch retrieval is modulated as follows (Figure 6) :

a) Consider the primary map M (Figure 5a), $\delta_c(M)$ (Figure 5b) as the dilation of M and the binary image of the street name's regions $T^{-1}[M]$(Figure 5c). The image g_0

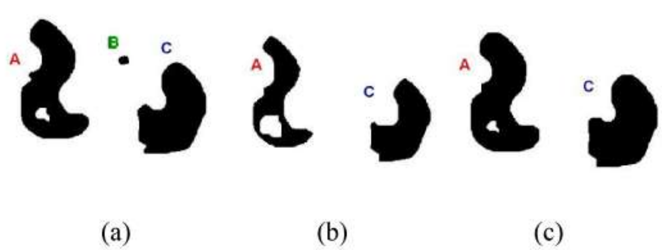


Fig. 4. a) Primary image, b) region B is eliminated as desired. As a side effect, regions A and C have shrunk, c) desired areas are restored to their primary size.

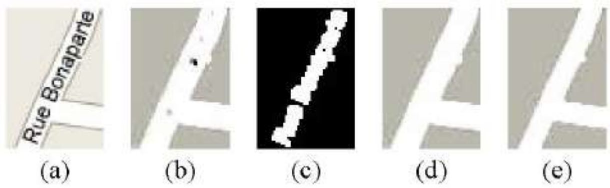


Fig. 5. a) Primary image f, b) $\delta_c M$, c) $T^{-1}[M]$, d) g_0 term, e) street names have been removed while the road has been retained at his primary size.

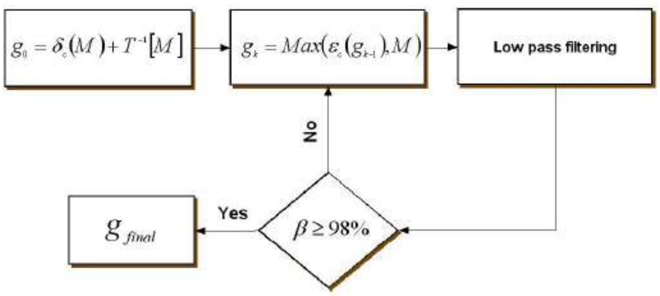


Fig. 6. Anti-extensive connected operators block diagram.

(Figure 5d) is calculated according to equation:

$$g_0 = \delta_c(M) + T^{-1}[M]. \tag{2}$$

b) If we represent as $E[M]$ the erosion of M, then each term g_k produced after the k-th iteration is recursively computed according to the equation:

$$g_k = Max(\epsilon(g_{k-1}), M) \tag{3}$$

.

c) Next, the system examines the last two sequential terms g_{k-1} and g_k. If

$$\beta = \frac{N_V}{RC}100\% = \begin{cases} \geq 98\%, & \text{end} \\ < 98\%, & \text{procceed in step b.} \end{cases} \tag{4}$$

where, N_V is the amount of elements of set V,

$$V = \{(x, y) \in g_k \mid g_{k(x,y)} = g_{k-1(x,y)}\} \tag{5}$$

and R,C are the image dimensions.

Note that an additional step (c) is introduced in the algorithm so as to ensure convergence after finite number of iterations. Figure 5e illustrates the result of the aforementioned iterative procedure.

III-B.2. Color-Based Clustering

After eliminating street names a color-based clustering algorithm classifies every pixel of the resulting image g_{final} to two classes, namely: "Roads" and "Buildings". Thus, we assume that for every type of road a mean color value is predefined. Let us assume that η random variables z_i , where $i \in \{1, 2, \ldots, \eta\}$, represent road colors and follow 3D Gaussian distribution in the color space. More precisely :

$$f_i(\boldsymbol{r}) = \frac{1}{2\sqrt{\pi \boldsymbol{C}}} exp(-\frac{\boldsymbol{r} - \boldsymbol{r}_i}{2\boldsymbol{C}^2}) \tag{6}$$

where $\boldsymbol{r}_i$ the mean value of the i-th road color, matrix $\boldsymbol{C}$ is the covariance matrix and is diagonal.

After evaluating all functions f_i, pixel p is classified as "Roads" if and only if

$$f_i(\boldsymbol{r}_p) > t_i \tag{7}$$

for only one of the functions f_i , where t_i the i-th threshold. Otherwise p is classified as "Buildings".

Crossroads are simply detected as the areas that belong jointly to more streets. Each street segment that lies between two crossroads is then linked to the appropriate street name.

III-C. Generation of the Haptic Map

The final step is the construction of the Multimodal Map. The latter's model supports haptic interaction and provides navigational aural messages. Based on set 'Roads' and 'Buildings' the system constructs a grooved-line-map representation of the map network as illustrated in Figure 7b.

Via off-the-shelf OCR and TTS algorithms user is provided with the necessary aural messages while navigating. For haptic interaction the PHANToM Desktop force feedback device [12] is used.

IV. EXPERIMENTAL RESULTS

IV-A. Experiments

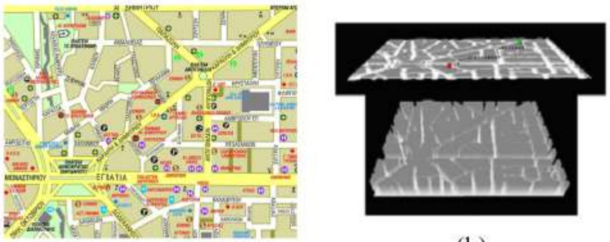
The proposed method was tested in generating Haptic Map models of maps from various providers. Notice that all used samples fulfill all the prerequisites enumerated in Section II.

Figure 7a shows a map image of the center of Thessalonici acquired by [13] as also the obtained Multimodal Map. Figure 8a illustrates a map image of Seattle acquired by [14]. Figure 8b shows the recognized street names. The detected crossroads are shown in Figure 8c with red color. Finally the produced Multimodal Map is illustrated in Figure 8d.

The obtained Multimodal Map contains noise that is in acceptable range. Moreover the majority of crossroads are detected correctly. At this point it must be mentioned that the presented map analysis methods are executed off-line. As also illustrated in Figures 7b and 8d the resulting pseudo-3D representation of the map is very clear and haptic rendering can be easily performed at interactive rates.

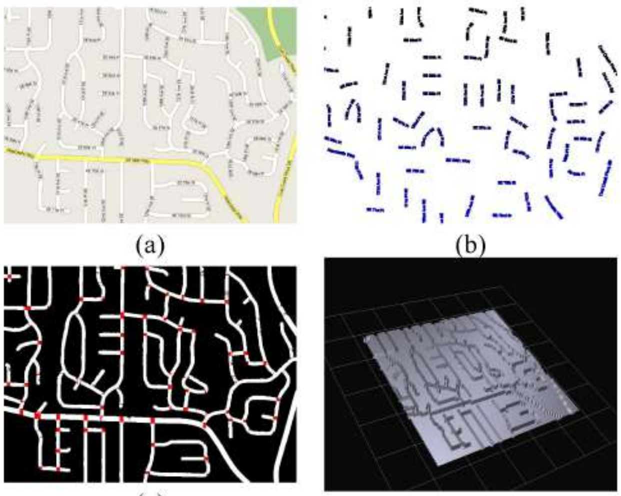
IV-B. Evaluation

The designed Haptic Map has been tested by 19 visually impaired people from the Local Union of the Pan-Hellenic Association of the Blind in Thessalonica, Greece. The overall evaluation was satisfying. Initially some instructions about the system were given to the users. Then their task was to go from a specific start point (red point in Figure 7) to another location in the pseudo-3D map (green point in Figure 7). Users have not faced major difficulties, while they expressed their belief that haptic maps are a very promising means of navigation.



(a) (b)

Fig. 7. (a) Primary map image illustrating the center of Thessalonica city Greece, (b) The obtained Multimodal Map model used for evaluation by visually impaired.



(a) (b) (c) (d)

Fig. 8. (a) Primary map image, (b) recognized street names, (c) red spots indicate crossroads and (d) 3D map model.

V. CONCLUSIONS

During the latest years cartography has been greatly advanced with the GPS being the tip of the iceberg. Unfortunately, up to date common maps are perceived using the visual modality. Therefore, special population categories like the visually impaired cannot access this information. In the present paper research is focused on transforming visual data to haptic representations.

A robust framework has been presented that generates a haptic-aural representation of the 2D map. The blind users are able to navigate in the generated pseudo-3D map using a haptic device, while audio feedback regarding the street names is also provided. Extensive tests have illustrated the efficiency of the approach.

VI. ACKNOWLEDGMENTS

This work is supported by the SIMILAR IST Network of Excellence (FP6-507609).

VII. REFERENCES

[1] Disability INformation Resources (DINF). [Online]. Available: http://www.dinf.ne.jp/doc/english/Us_Eu/conf/csun_98/

[2] Vincent Gaudissart, Silvio Ferreira, Céline Thillou, Bernard Gosselin, "SYPOLE: Mobile Reading Assistant for Blind People.", Faculté Polytechnique de Mons, Laboratoire de Théorie des Circuits et Traitment du Signal Bâtiment Multitel.

[3] Lee Wee Ching and Maylor K.H. Leung , "SINVI : Smart Indoor Navigation for the Visually Impaired", 8th International Conference on Control, Automation, Robotics and Vision Kunming, China, 6-9th December 2004.

[4] Dimitrios Tzovaras, Georgios Nikolakis, Georgios Fergadis, Stratos Malasiotis and Modestros Stavrakis, "Design and Implementation of Haptic Virtual Enviroments for the Training of the Visually Impaired",IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol.12, No.2, June 2004.

[5] K. Moustakas, G. Nikolakis, K. Kostopoulos, D. Tzovaras and M.G. Strintzis, "The Force Field Haptic Rendering Method: Application in Haptic Access to Visual Data for the Training of the Visually Impaired", IEEE Multimedia Magazine.

[6] Abdolah Chalechale, Golshah Naghdy and Alfred Mertins, "Sketch-Based Image Matching Using Angular Partitioning", IEEE Transactions on Systems, Man, and Cybernetics- Part A:Systems and Humans, Vol. 35, No. I, January 2005.

[7] Abdolah Chalechale, Golshah Naghdy and Alfred Mertins, "Sketch-Based Image Retrieval Using Angular Partitioning".

[8] N.Grammalidis, N.Sarris, F.Deligianni and M.G.Strintzis: "Three Dimensional Facial Adaptation for MPEG-4 Talking Heads", EURASIP Journal on Applied Signal Processing, Special Issue on Signal Processing for 3D Imaging and Virtual Reality, Vol. 2002, No. 10, pp. 1005-1020, October 2002.

[9] P.Salembier, L.Garrido and A.Oliveras. UPC Barcelona, SPAIN. "Region-based filtering of images and video sequences: a morphological viewpoint", May 2001.

[10] Philippe Salembier, Albert Oliveras and Luis Garrido. "Anti-extensive connected operators for image and sequence processing.", IEEE Transactions on Image Processing, Vol. 7, No.4, pp.555-570, April 1998.

[11] P.Salembier and F.Marqués. "Region-based representations of image and video: Segmentation tools for multimedia services.", IEEE Transactions on circuits and systems for video technology, volume 9, No 8, December 1999.

[12] SensAble Technologies Inc.. PHANToM Haptic Device.

[13] http://www.031.gr/

[14] http://maps.google.com/