Mini-Review of Technologies for Mobility of the Visually Impaired

Guillaume Tatur, Edwige Pissaloux

Mini Review

Open Access 🔓

Mini-Review of Technologies for Mobility of the Visually Impaired

Guillaume Tatur¹ and Edwige Pissaloux*²

¹*Institut de la Vision, CNRS - INSERM - University Pierre & Marie Curie, France*

²*LITIS, Université de Rouen Normandie, France*

Corresponding author: Edwige Pissaloux, LITIS, Université de Rouen Normandie, 76000 Mont Saint, Aignan, France

Abstract

As mobility is significantly correlated with autonomy and quality of life, a clear majority of the developed approaches and devices dedicated to visually impaired individuals attempts to augment or support orientation and mobility abilities and maximize independence and safety. Therefore, various methods and technical solutions have been proposed for generating a representation of the visual scene and transmitting this information to the user. As different visual impairments can be distinguished, from congenital blindness to low vision, dedicated solutions have been put forward to accommodate these different needs and abilities. These solutions may be used universally or take advantage of the specificity of the visual impairment (e.g. low vision, late blindness). Through the presentation of sensory supplementation, technical aids and visual neuro-prostheses studies, we will explore some of the approaches and technical solutions that have been proposed or are still in development. Additionally, we will discuss the benefits of providing specialized information dedicated to mobility as opposed to a general-purpose representation.

Abbreviations: VI- Visually Impaired; TVSS- Tactile Vision Substitution System; TDU- Tongue Display Unit; RP- Retinitis Pigmentosa; AMD- Age Related Macular Degeneration; DBR- Depth-Based Representation; LBR- Luminosity-Based Representation

Introduction

As it is significantly correlated with autonomy and quality of life, a clear majority of the developed approaches or devices dedicated to visually impaired (VI) individuals attempt to augment or support orientation and mobility abilities and maximize independence, safety and efficiency of movement. Whether it is applied to sensory substitution devices, visual neuro-prostheses or technical aids, the targeted objective of this representation is one of key points when designing scene representation at a functional level. Two approaches can be considered with respect to the category of information that a representation may provide. The first approach would be the attempt of representing a scene using general information (e.g. raw input from a camera), expecting that the user will extract the relevant cues, pertinent for his or her current task, from the pattern of luminosity values as represented by the image pixels gray level for instance. This approach is highly dependent on the complexity of the represented information and on the user capability to decipher this ambiguous representation through the selected sensory modality. Another approach would be to optimize a representation to provide specific information for a task like mobility. Optimized representation with respect to a specific task or activity is interesting for several reasons. Indeed:

a)   It considers the limited processing capabilities or resolution of the device,

b)   Information can be filtered, simplified and enhanced (or augmented) relatively to the task performed,

c)   As it is seamlessly contextualized by the task and as it uses specific information, this representation should be easier to understand and may be better adapted to end-user needs.

Designing such representation is a challenging task as, for instance, one needs to balance between a generalized or a specialized representation; for example: a device that gives access to the distances to obstacles in the scene through a simple coding into artificial stimuli intensity may be easier to understand and to decode in real time. On the other hand, providing general information such as luminosity information may be ambiguous and may vary with respect to several factors (e.g. lightning conditions, viewpoints, textures). However, the latter approach may allow the user to infer locations of distant and visually salient scene features (e.g. ceiling lights, large contrasted walls...) and allow them to use landmarks for navigation.

This paper, through a state-of-art analysis, tries to identify some guidelines how to find the most suitable and well-balanced representation pertinent for a final exploitation (assistance of mobility). It focuses on devices and techniques mainly based on distance evaluation or depth processing and the presentation

of adapted cues, clues and landmark information for static and dynamic scene objects location. In the subsequent sections are addressed scene representation used in sensory supplementation devices (Section II), in augmented and virtual reality (Section III), in visual neuro-prostheses (Section IV)

Technical Aid and Sensory Supplementations Systems

This section presents an overview of non-invasive sensory supplementation systems, which provide geometric information to the user about its surroundings as well as information pertinent to mobility.

In these systems, information usually acquired from one sensory modality is captured through an artificial sensor (e.g. video camera) and is processed to create stimuli adapted to another modality. Two main categories of these devices can be distinguished to supplement the sense of vision: visuo-tactile and visuo-auditory systems.

Visuo-Tactile Sensory Supplementation

Various systems have been proposed so far; some of the most representative are presented hereafter.

Scene Representation Through the Tactile Sense

Figure 1: Young visually impaired subject reproducing hand gesture perceived through a head-worn camera and a 6x24 vibrotactile array. The LED monitor in the foreground shows the active pattern of the tactile display [1].

The Bach-y-Rita's TVSS (tactile vision substitution system [1]) is one of the early developed and most well-known visual to tactile sensory supplementation devices. The system converts an image captured by a video camera to generate tactile stimuli through a stimulation device placed on a part of the body (skin) such as the lower back, the abdomen or the fingertip in the first version of the system. Captured color images are transformed into black and white images and reduced spatially to accommodate for tactile device resolution (between 100 and 1032 stimulation points). Figure 1 shows a blind child reproducing hand gesture with the TVSS (placed on the table in front of the child). In [2], extending the initial work of Bach-y-Rita, Collins et al. have developed a 1024 wearable electrotactile system placed on the abdomen and converting images acquired through a 90° field of view camera. As a result,

subjects equipped with such device have been able to safely avoid and navigate around large and contrasted objects. However, these results are only valid in simple indoor environment as subjects, even trained, failed to use the system in real outdoor environment. This setup was later modified into a tongue display unit (TDU), also called Brain Port, which requires reduced electrical current for the stimulation and was more suitable for mobility purpose [3]. Some research teams have developed a "specialized" representation approach to present relevant and adapted information for mobility.

Space Representation in Mobility Support

In [4], the authors propose a novel scene representation method dedicated to mobility. In their approach, the 3D scene captured by a stereoscopic pair of cameras, is projected vertically (i.e. gravitationally/orthographic projected) on a 2D plan. Through the proposed processing, obstacles are represented as non-accessible areas on the plan, whether they are overhanging or lying on the ground. The obtained simplified map of the environment is then presented to the user through a hand-held touch stimulating Braille-like device (Figure 2) and is updated in real-time. In the proposed representation, processing of the 3D data allows a binary partitioning of the space into two zones: obstacle-free and obstacles. The system is aimed to provide a dynamical tactile representation of the scene where raised taxels (i.e. individual tactile stimulators that compose the device) principally show the limits of the obstacles-free space. Furthermore, information is represented on the tactile surface in a user-centered Euclidian reference frame which observation point is indicated by a notch on one of the borders of the tactile surface. Preliminary navigation experiments in standardized test environments indicate that the system can be used to perceive the spatial layout of the dynamical environment (named tactile gist [5]) through this binary tactile representation.

Figure 2: Illustration of the developed scene representation (bottom row) and device (top row). Images of the scene, acquired with a stereo rig, are processed to create a 2D tactile map (right image). This simplified representation (tactile gist) shows dynamically the accessible and non-accessible parts of the environment delimited by raised taxels. The origin of the Euclidian reference frame is attached to the tactile map and is materialized by a notch in one of the device borders. The stimulation device is a Braille matrix of 8 by 8 shape memory alloys based taxels [4,5,6].

In [6], the authors propose a representation of the 3D scene like [4], except that there is no notch on the necklace-like device (overhanged over the neck) what increase the cognitive load for localization in the representation; for these reasons, this device have been replaced by a tactile belt which does not use an external explicit representation of the space. As another approach to mobility, Chekhchoukh et al. [3] propose a guidance technique using also the TDU based on a symbolic scene presentation. The authors mainly propose to generate a 3D path composed of dots that must be followed to travel from one place to another. Stimulus intensity associated to each dot represent the distance of this 3D dots from the user.

Visuo-Auditory Sensory Supplementation

Visual to auditory sensory supplementation systems may be categorized into two approaches:

a)   The auditory coding of the position and distance of the obstacles through the use of an ultrasonic, infrared or laser rangefinder device for instance;

b)   A visual to sound real time translation based on the coding of visual patterns extracted from captured images (e.g. from a head mounted camera).

Range Finder Devices

In this class, the auditory feedback provided by such devices gives the user information about the texture, the distance, the size and position of the scene entities. In these devices, distance is mainly represented by the sound intensity and the positions of objects around the user are coded through binaural disparity.

For instance, the C-5 laser cane [7] is a triangulation-based device, embedded in a white cane. Obstacle detection output information on obstacles in three different zones (at head height, in front of the user and on the ground) by generating a specific tone.

Vision to Sound Scene Translation

In this class, image is converted appropriately into sound. One of the most well-known devices of this class is the voice developed by Meijer et al. [8]. It proposes an image to sound conversion of the captured video frame pixel arrays into successive sound signals. Each signal is generated according to the pixel position and brightness: pitch is proportional to pixel height (vertical position), amplitude (loudness) to brightness. To give information on the horizontal position of the stimulus, a left to right scanning technique has been implemented ("synchronization" sound click indicates the start of a frame scan). There are very few visually impaired people using the voice as the proposed representation is very complex and learning phase requires several years.

Using Residual Vision: Augmented and Virtual Reality Devices

Several visual disorders such as retinitis pigmentosa (RP), age related macular degeneration (AMD) or glaucoma may induce severe vision loss. Indeed, persons suffering from low vision may experience significant visual field reduction (like in RP patients) or central vision loss in the case of AMD for instance, visual defects (e.g. scotoma) and other visual impairments (e.g. reduced contrast sensitivity, impaired color vision). Augmented and virtual reality devices make possible, to a certain extent, to compensate these visual impairments by:

a)   Providing additional information in the remaining field of view. Without such technical aid, this information may not be directly accessible. For instance, in the case of visual field reduction, laterally located static or moving obstacles may not be seen directly.

b)   Enhancing the residual vision, by displaying reinforced contrasts and colors for instance.

One of the most well-known methods based on augmented reality technique was proposed by Eli Peli's works [9]. Peli's approach is based on visual field restitution, as visual field is one of the key components for mobility. In [9], the author presents his approach on "vision multiplexing". The principle of vision multiplexing, in the case of persons with a visual field of view strongly restricted, can mainly be described as a spatial multiplexing technique: using a wide-angle head-worn camera and a see-through head-mounted display, the method is based on the presentation of captured images contours (through real-time edge detection) displayed as white lines and superimposed on the natural view. AUREVI [10] is a research project dedicated to the development of a technical aid for low vision individuals (mainly persons with RP) based on an immersive approach. The currently proposed device concept is composed of head-mounted cameras and a binocular display with screens that prevent the external light from reaching the user eyes. The later point is important, as the chosen approach must have a complete control over the presented luminosity and visual aspect of the scene representation displayed on the screens. Controlling luminosity in real-time should for instance provide optimal display physical output regarding the user's dynamic light adaptation capability (e.g. smoothed light transition between indoor to outdoor scenes). In addition to light adaptation, other system parameters such as color and contrast are driven by the user's residual vision screening. Interestingly these parameters may be updated to consider user's disease progression.

Visual Neuro-Prosthesis

Visual neuro-prostheses are an invasive technique to restore some rough form of vision in individuals suffering from late blindness due to degenerative diseases of the retina. These devices are based on the stimulation of still functional neurons in some part of the visual pathway. By means of this stimulation, mainly electrical, it is possible to elicit visual perceptions in the implant recipient visual field, called phosphenes [11].

Prosthetic Vision for Mobility

In, Tatur et al. [12,13] proposed a method to generate a depth-based representation (DBR), where the brightness of each phosphene is defined by the distance to the surrounding objects, its intensity increasing as distance decreases, as opposed to the luminosity-based representation (LBR) where each phosphene

appearance is driven by the luminosity information in the observed scene. The information obtained from the DBR is independent of the texture, reflectivity of entities and of lighting conditions. Despite the advantages of depth-based representation (i.e.: direct ego-centered distance information presentation and possibility of distance-based filtering or processing), some difficulties remain. Indeed, as only discrete intensity levels can be perceived by the implantee [13] and thus only a discrete number of different phosphene brightness values can be generated. Therefore, such representation must cope with a trade-off between the range of presented distances and the resolution at which it will be displayed (i.e. number of brightness levels representing a range of distance values). Furthermore, even though this representation seems efficient for a safe perambulation, a navigation task requires to be also able to orient oneself. For self-orientation purpose the luminosity information can be used to acquire visual cues such as light sources (e.g. windows, ceiling lights...) as well as contrasted areas (e.g. marked pedestrian crossing). The authors propose a method to create a composite representation that combines both types of information in a unique representation based on temporal scanning of the depth layers (Figure 3). In this scene representation, the initially displayed phosphenes pattern correspond to the LBR; then, depending of their distance to the user, a successive highlight of the objects occurs until a previously defined maximum scanning distance have been reached. Implementation details and considerations are presented in [13].
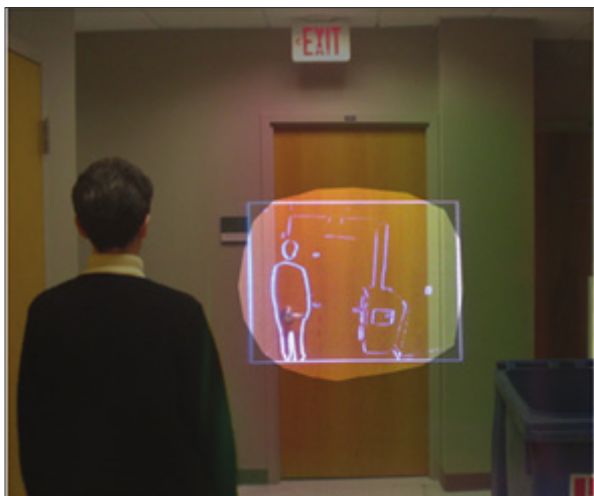

Figure 3: Illustration of the spatial multiplexing principle dedicated to low vision patients with severe visual field reduction, equipped with a head mounted display and a wide field of view camera. As it is a see-trough device, the natural scene can be seen directly (background image). Edge maps are calculated in real-time from the captured camera images and are minified and displayed as white lines. Although the patient has a tunnel vision (illustrated as a circular highlighted aperture in the image center), he now has access to information contained in a wider part of the environment [9].

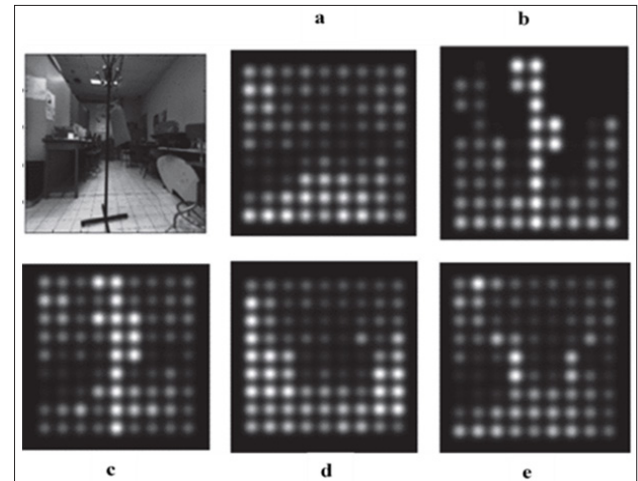This representation has additional advantages (see Figures 4 & 5):

Figure 4: Composite representation as proposed by [13]. While observing a scene using the LBR (a), and based on the DBR (b), the proposed method allows to activate a temporal scanning technique which progressively highlight objects contained in the successive depth layers of the scene (from near distances (b) at a time t0 to a maximum distance (e) at t1 > t0).

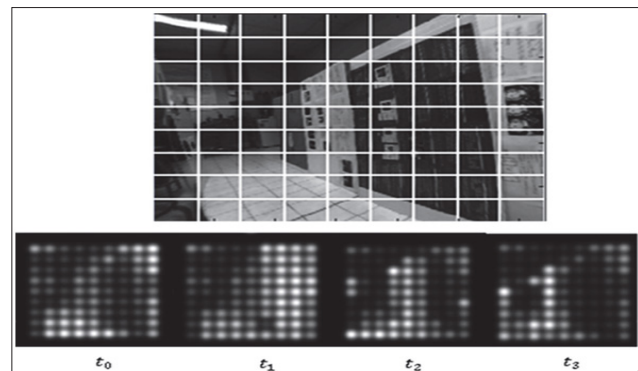**Figure 5:** Representation of a corridor-like scene. Top row: full resolution image (white lines represent phosphenes receptor field areas) of a large room with panels which create a corridor. Bottom row: composite representation for the time t0 to t3. Depth information indicates the presence of lateral obstacles that border what seems to be a corridor. A dark region on the right side of the image could have been interpreted as a passage with only the LBR. However, it appears to be part of the wall and could then be correctly interpreted. Vanishing point may also be detected by observing the convergence of the successive depth layers highlight.

a)   This temporal scanning method should provide a better depth layer discrimination,

b)   Clues on environment geometrical layout may be inferred by the user,

c)   Progressive highlights of the depth layer should lower the ambiguity of the luminosity-based representation by giving local depth information on contrasted areas which

could be misinterpreted otherwise (e.g. a dark wall could be understood as a passage). A similar approach using this depth-based representation have been tested in [15] in navigation task through a simple maze with overhanging obstacles. Results indicate that the depth-base representation was more efficient than the luminosity based one. However, the simulated electrode array dimension (30x30 phosphenes) was far from current epi-retinal prostheses capability.

Conclusion

This paper has addressed an overview of the recent presentation of the space supposed to be integrated in devices for spatial data acquisition and mobility of visually impaired people assistance. Some of these representations have been already partly validated with dedicated mobility tasks [16]. Whether it applies to mobility or any other tasks, researches still need to be conducted to define optimal scene representations adapted to the chosen category of technical aid, stimulation method and processing strategy. For the later, as it has been discussed through this chapter, there is still a debate whether the information should be processed or not regarding the performed task for instance, or simply "converted" to imitate the natural visual sense input (e.g. luminosity to stimuli intensity) even if the recent literature tends to support the former approach.

References

1. Bach y Rita P, Kercel SW (2003) Sensory substitution and the human-machine interface. Trends in Cognitive Sciences 7(12): 541-546.

2. Collins C C (1977) Electrotactile visual prosthesis, in Functional Electrical Stimulation. F T Hambrecht and J B Reswick (Eds.) Marcel Dekker, New York, USA pp. 289-301.

3. Chekhchoukh A, Goumid M, Vuillerme N, Payan Y, Glade N (2013) Electrotactile vision substitution for 3D trajectory following. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society, Annual Conference: 6413-6416.

4. Pissaloux E, Velázquez R, Maingreaud F (2009) Multisensor Fusion and Integration for Intelligent Systems: An Edition of the Selected Papers from the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems 2008. In H. Hahn, H Ko, S Lee (Eds.). pp. 349-357.

5. Pissaloux E, Velazquez R, Maingreaud F (2017) A New Framework for Cognitive Mobility of Visually Impaired Users and Associated Tactile Device. IEEE Trans. Human Machine Systems 47(6): 2168-2291.

6. Bourbakis N, Makrogiannis S K, Dakopoulos D (2013) A System-Prototype Representing 3D Space via Alternative-Sensing for Visually Impaired Navigation. IEEE Sensors Journal 13(7):2535-2547

7. Benjamin M J, Ali N A, Schepis AF (1973) A laser cane for blinds. In Proc San Diego Biomedical Symp 12: 53-57.

8. Meijer PBL (1992) An experimental system for auditory image representations. IEEE Trans. Biomed. Eng 39(2):112-121.

9. Peli E (2001) Vision multiplexing: an engineering approach to vision rehabilitation device development. Optom Vis Sci 78(5): 304-315.

10. Maalej A, Tatur G, Lorenzini M C, Dupeyron G, Dumas M, et al. (2014) High Dynamic Range Imaging System for the Visually Impaired. ACVR2014 Second Work Assist Comput Vis Robot 8927: 632-642.

11. Humayun M S, Weiland J D, Greenberg G, Williamson R, Little J, et al. (2003) Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. Vision Research 43(24): 2573-2581.

12. Tatur G, Marc I, Dumas M, Dupeyron G (2010) Conception of a Phosphene based Visual System - A Functionnal Approach. In Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS 2010): 339-344.

13. Tatur G, Marc I, Dupeyron G, Dumas M (2014) Orientation and Mobility with Prosthetic Vision - Combination of Luminosity and Depth Information for Scene Representation. In Proceedings of the International Conference on Bio-inspired Systems and Signal Processing (BIOSTEC 2014): 122-126.

14. Humayun MS (2009) Preliminary results from Argus II feasibility study: a 60 electrode epiretinal prosthesis. Investigative ophthalmology & visual science 50(13): 4744.

15. Lieby P, Barnes N, McCarthy C, Liu N, Dennett H, Walker JG, Botea V, Scott AF (2012) Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions, ARVO 2012, Fort Lauderdale, Florida, USA.

16. Pissaloux E, Velázquez R, Hersh M, Uzan G (2017) Towards a cognitive model of human mobility : an investigation of tactile perception for use in moblity devices. Journal of Navigation 70(1): 1-17.

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

https://biomedres.us/