A System-Prototype Representing 3D Space via Alternative-Sensing for Visually Impaired Navigation

Nikolaos Bourbakis, *Fellow, IEEE*, Sokratis K. Makrogiannis, *Member, IEEE*, and Dimitrios Dakopoulos

***Abstract*— Offering an alternative mode of interaction with the surrounding 3-D space to the visually impaired for collision free navigation is a goal of great significance that includes several key challenges. In this paper, we study the alternative 3-D space sensation that is interpreted by our computer vision prototype system and transferred to the user via a vibration array. There are two main tasks for conducting such a study. The first task is to detect obstacles in close proximity, and motion patterns in image sequences, both important issues for a safe navigation in a 3-D dynamic space. To achieve this task, the images from the left and right cameras are acquired to produce new stereo images, followed by video stabilization as a preprocessing stage, a nonlinear spatio-temporal diffusion and kernel based density estimation method to assess the motion activity, and finally watershed-based detection of moving regions (or obstacles) of interest. The second task is to efficiently represent the information of the captured static and dynamic visual scenes as 3-D detectable patterns of vibrations applied on the human body to create a 3-D sensation of the space during navigation. To accomplish this task, considering the current limitations imposed by the technology, we create a high-to-low (H-L) image resolution representation scheme to facilitate the mapping onto a low-resolution 2-D array of vibrators. The H-L scheme uses pyramidal modeling to obtain low-resolution images of interest-preserving motion and obstacles-that are mapped onto a vibration array. These patterns are utilized to train and test the performance of the users in free space navigation. Thus, in this paper we study the synergy of these two important schemes to offer an alternative sensation of the 3-D space to the visually impaired via an array of vibrators. Particularly, the motion component is employed as an element for the identification of visual information of interest to be retained during the H-L transformation. The role of the array vibrators is to create a small-scale front representation of the space via various levels of vibrations. Thus, 3-D vibrations applied on the user's body (chest, abdomen) offer a 3-D sensation of the surrounding space and the motion in it. In addition, we present experimental results that indicate the efficiency of this navigation scheme in creating low-resolution 3-D views of the free navigation space and detecting obstacles and moving areas.**

***Index Terms*— 2-D vibration arrays, alternative sensing, assistive technologies, blind's navigation, high-to-low image representation, human computer interface, motion detection.**

Manuscript received November 5, 2012; revised January 29, 2013; accepted March 11, 2013. Date of publication March 19, 2013; date of current version May 29, 2013. This work was supported by NSF and AIIS Inc. for the Tyflos Project. The associate editor coordinating the review of this paper and approving it for publication was Prof. Subhas C. Mukhopadhyay.

N. Bourbakis and D. Dakopoulos are with the Assistive Technologies Research Center, Wright State University, Dayton, OH 45435 USA (e-mail: nikolaos.bourbakis@wright.edu; dakopoulos.2@wright.edu).

S. K. Makrogiannis is with the National Institute on Aging (NIH), Baltimore, MD 21225 USA (e-mail: makrogianniss@mail.nih.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2013.2253092

I. Introduction

In THE last two decades a variety of useful assistive devices have been proposed, designed and developed in an effort to complementarily assist visually impaired individuals during navigation in their living and working environments [1], [19]–[34]. Since most of these devices process visual information captured from the surrounding space, one of the important challenges that arise is how to efficiently detect and track motion from the surrounding environment and effectively represent it to the visually impaired. Thus, this challenge is directly related to image/video understanding research, especially to motion analysis, and to haptic technologies as well. In this work we first deal with the motion analysis and then we proceed with the high-to low representation for 2D vibrations.

The domain of motion analysis is related to the estimation and recognition of the movement of scene objects and the ego-motion of the imaging sensor. Some widely investigated research topics of this field are motion segmentation, object detection and tracking, stereo imaging, 3D motion and image and video registration. The analysis of dynamic scenes in particular is very popular with applications to surveillance, target tracking and navigation systems. Here we first propose a motion analysis scheme as *an assistive vision* tool that can be useful for assisting the navigation of visually impaired people [13], [14]. The motion analysis and segmentation approaches may be generally divided into the categories of image differencing, optical flow-based methods and spatio-temporal methods. The image differencing or subtraction methods presented first in the literature ranging from simple subtraction operations up to more sophisticated schemes like background registration [5]. The motion segmentation in this context is accomplished by detecting the optical flow discontinuities. Several optical flow implementations have been proposed for video processing [7], however the motion field cannot be reliably estimated on the borders of the moving objects and in smoothly textured areas. These pitfalls have led to the introduction of approaches that simultaneously estimate the global motion and its spatial support; these methods are characterized as spatio-temporal [12]. According to these approaches the image areas are grouped by comparing their spatial and temporal properties.

Here, we use a scheme for segmenting images and detecting moving objects in sequence of images [2]. This is obtained by a video stabilization algorithm, followed by a

spatio-temporal diffusion kernel for motion detection, and a watershed algorithm for segmenting the moving objects. The video stabilization is a preprocessing step that compensates for the small displacements caused by the natural gait of the human body. Next, a novel motion activity measure based on spatio-temporal anisotropic diffusion is proposed for the detection of moving areas. The refinement of the moving areas is done by the kernel density estimation method, while the watershed segmentation algorithm segments the moving regions. Note that the motion analysis takes place after the fusion of the image frames from the two cameras by using a well-known depth method from the open computer vision library.

Furthermore, in an assistive human-environment interface a blind user needs to receive visual information in different forms of sensation in order to sense the surrounding space. There are several very interesting approaches with alternative sensing. One very important and at the same time very informative form is 2D vibration in a form of arrays [18]. The applicability of the 2D vibration arrays, however, is limited due to their low spatial resolution. Consequently, the resolution of the information provided by the vibration arrays is lower relative to the one captured by vision cameras. Thus, the efficient representation of the most important parts of an image into a low-resolution array becomes a challenge. As a response to this issue, in this paper, we also propose to use a high-to-low resolution scheme for reliable conversion of high resolution visual information to low resolution 2D vibrations by preserving critical visual information for the blind user's navigation needs. By preserving the critical visual information we create low-resolution representations where *motion, obstacles and borders* are treated as pieces of information of "interest" and during the reduction of the resolution we use criteria to preserve them as much as possible. The high-to-low (H-L) scheme is based on pyramidal reductions including the user's interest as a part of the information reduction process. The H-L representation of resolution in images (pyramids, trees, interpolation, regular-decomposition) has been extensively studied in the 80 s with main application to image communication and video processing. The main algorithmic scheme used for such a purpose is pyramids in combination with regular decomposition in order to obtain the areas of interests. Specifically, our Tyflos navigation prototype uses two cameras to create stereo-vision representations, to which we apply pyramidal image reduction with criteria of interest (close proximity obstacles, motion, and maximum free navigation space). This process results in a reduction of the image information, while retaining information of interest and eliminating or reducing further information of low importance. The outcome of this process produces a low-resolution representation of the visual information to be mapped on an equally low-resolution array of vibrators. In other words, here we combine three methodologies: the 3D visual representation of the scene, the motion detection (since motion is an important component of collision free navigation) and the H-L scheme. The latter is modeled by using a formal array language to generate 3D vibration patterns. The user senses these patterns on the vibration vest.

Thus, in this paper we offer a synergy of methodologies (image stabilization, image segmentation, motion detection, and depth maps) to detect and track motion in images and followed up by a method that reduces the resolution of the visual information (previously detected) and mapping it into a low resolution vibration array in order the blind user to sense the depth of the free 3D space in front of him/her via vibrations during his (her) navigation. The diagram depicts the overall idea of the methodology.

The overall structure of the paper is organized as follows: Section II briefly presents the navigation system (Tyflos). Section III explains the motion detection method, including the frame stabilization, the motion activity measure, the kernel-based density estimation and the watershed-moving object detection scheme. Section IV describes the high-to-low representations that include pyramidal schemes and vibration patterns. Section V presents testing scenarios with rules that reduce the number of patterns. Section VI provides some experimental results and a discussion of the applicability of the method, and the final Section VII contains the conclusions about the proposed method and our future goals regarding the Tyflos project.

II. Tyflos Wearable System-Prototype

A. Brief System Description

The main role of the Tyflos assistant is to capture environmental data using cameras and range sensors and map the extracted and processed content onto an array of vibrators (vest) offering a 3D sensation of the open navigation space to the user. The Tyflos prototype integrates various hardware components such as: cameras, range sensors, a microphone, a vibration vest and earphones for voice-audio and haptic interfacing with the user. The user along with the main processing unit can wear all these components. The processing unit is a hand-held computer (such as a PDA). Data collected by the sensors is processed by the Tyflos' software methods.

The preliminary design and development of the Tyflos prototype has already been carried out in our lab [15]–[17]. The latest version of the prototype (Fig. 1) consists of two cameras, an ear-speaker, a microphone, the 2D vibration array vest and a portable computer. Students with visual disabilities have evaluated this device and their feedback has been used in the design requirements. As mentioned earlier, the prototype is based on the integration of several software components that reflect to some of the methodologies presented here. Thus, in this paper, we will focus on the Tyflos' navigator module and its role as a mobility assistant to the blind user.

B. Vibration Array

The vibration array (VA) is the critical hardware component of the Tyflos Navigator. The lower level components are the 16-miniature coin-type vibrator motors (C1030B028F; 3 mm thick, 1 cm diameter) arranged in a 4×4 array. Two vibrator motors can be driven, independently from each other, by the PCB vibrator module in square-pulse-type frequencies in the range of 1.25 Hz to 10.5 Hz using one digital potentiometer and two analog timers. A microcontroller controls via a

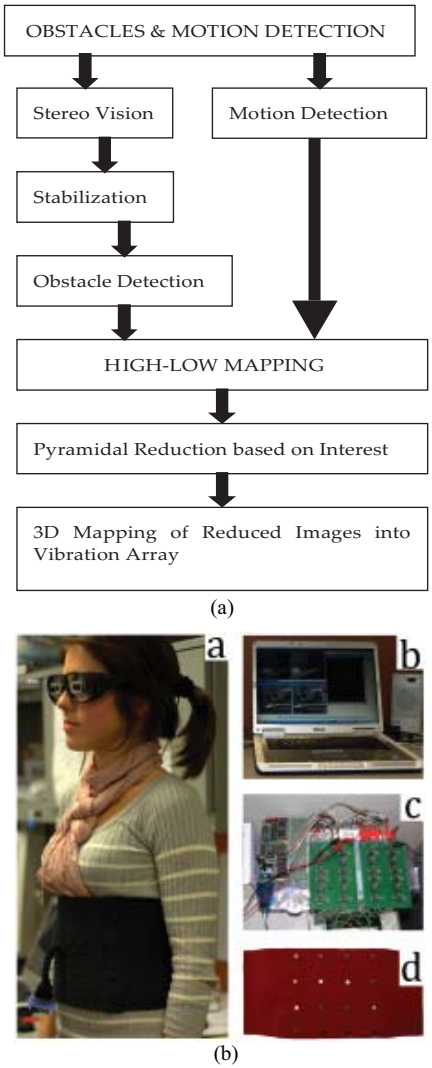


Fig. 1. (a) The overall methodology. (b) Tyflos prototype: The user is wearing the vibration array vest and the glasses with the stereo cameras while a microphone and an ear-speaker are also available for interfacing and controlling the system's functions (a). The cameras capture environmental information, the portable computer (b) processes it and through the microcontroller and the vibration array modules (c) the information is sent to the user via the vibration array vest worn on the abdomen (a). A vest with the 16 vibrating cells arranged in a 4 × 4 manner is shown in (d).

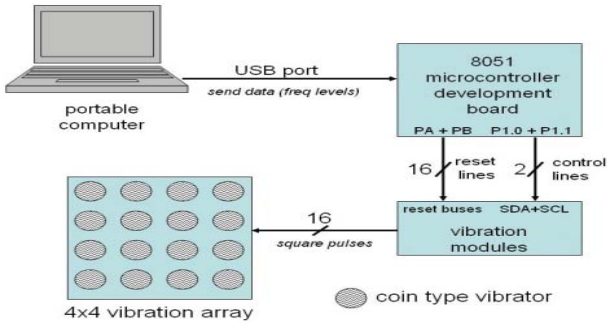


Fig. 2. Overall design of vibration array system.

2-wire interface the vibration modules that can drive all the 16 vibrator motors. The user can also operate the vibration array through a USB port, using the portable computer. The overall system design is shown in Figure 2.

We can also scale it up by adding for example more vibration modules that can "drive" more vibrators and achieving a higher vibration array resolution.

III. Motion Detection Method

The detection of motion is an important component to safe navigation, thus it is one of the main software modules of the Tyflos prototype that provides visual information in the form of motion activity maps. Thus, it is important for the motion scheme to be presented and analyzed in order to explain how activity in image frame is detected and represented. The motion scheme here consists of the following main stages: frame stabilization; estimation of motion activity; motion segmentation.

A. Frame Stabilization

The stabilization of frames in a video sequence is a process that attempts to estimate the small translations of the field of view due to unintentional ego-motion of the camera, which usually comprises of pan operations.

In our application, the camera is part of a wearable device and it follows the translational lateral motion that is caused by the human gait. Thus, a frame stabilization process is performed to keep the camera views constant, which will facilitate the subsequent motion analysis of the scene. The frame stabilization is obtained by transforming the original image plane to horizontal and vertical projections and subsequently estimating the motion in these 1D domains. These projections are estimated by the Radon transform. Assuming a function $f(x, y)$, the Radon transform is defined in the continuous case as follows:

$$T(\rho,\theta) = R_\theta[f(x,y)]. \\ = \int\int f(x,y)\cdot\delta(\rho - x\cos\theta, \rho - y\sin\theta)dxdy \quad (1)$$

In (1), ρ and θ are spatial variables in the radial coordinates; ρ is the radius and θ the angle; and δ is the Dirac function defined in (1)

$$\delta(i,j) = 1, \quad i = j \\ \delta(i,j) = 0, \quad i \neq j. \quad (2)$$

Here we compute the horizontal and vertical projections and for the discrete signal $I(x, y)$ we have

$$T_{0^\circ}(y) = \sum_{x}^{Nr} I(x,y) \quad (3)$$

$$T_{90^\circ}(x) = \sum_{y}^{Nc} I(x,y) \quad (4)$$

where, N_r is number of rows and N_c is the number of rows.

The result of this process is the decomposition of the initial 2D signal into two 1D signals. It was indicated that the affine motion can be efficiently estimated in the transformed space using the principles of optical flow.

Here the motion vector components are estimated by means of statistical measures. In our experiments we have used

the cross-correlation, cross-covariance and sum of square difference. For two 1D signals I and J these measures are respectively estimated as follows:

$$Corr(u) = \sum_x I(x) \cdot J(x-u) \tag{5}$$

$$Cov(u) = \sum_x (I(x) - \bar{I}) \cdot (J(x-u) - \bar{J}) \tag{6}$$

$$SSD(u) = \sum_x (I(x) - J(x-u))^2. \tag{7}$$

In the above equations, u is the estimation point, x is the spatial coordinate, $\bar{I}$ is the mean of I, and $\bar{J}$ is the mean of J. Cross-correlation and cross-covariance are normalized by the sum of the signal energy (squared intensity). The sum of squared difference produces minima at the borders due to the small intensity values (darker regions). This can be resolved when divided by the number of elements to produce the final similarity measures. The compensation for stabilization corresponds to the maximum of cross-correlation or cross-covariance, or the minimum of sum of squared differences. The estimated translations Δu_{Corr}, Δu_{Cov} and Δu_{SSD}: are given by

$$\Delta u_{\text{Corr}} = \arg\ \max Corr(u) - N_{PA}$$
$$\Delta u_{\text{Cov}} = \arg\ \max Cov(u) - N_{PA}$$
$$\Delta u_{\text{SSD}} = \arg\ \max SSD(u) - N_{PA}$$

where N_{PA} is the number of elements along the projection axis (horizontal, vertical or arbitrary).

An example of the stabilization process is displayed in Figure 3, where the natural body movements cause displacements of a wearable camera. This is a case of vertical displacement. In the first row are depicted two consecutive frames. The second row depicts the vertical projections respectively. The similarity measures are depicted in rows 3, 4 and 5 and compensated frame in the last row. Correlation and covariance maximum values indicate the displacement. The latter is also displayed by the minimum of sum of squared errors.

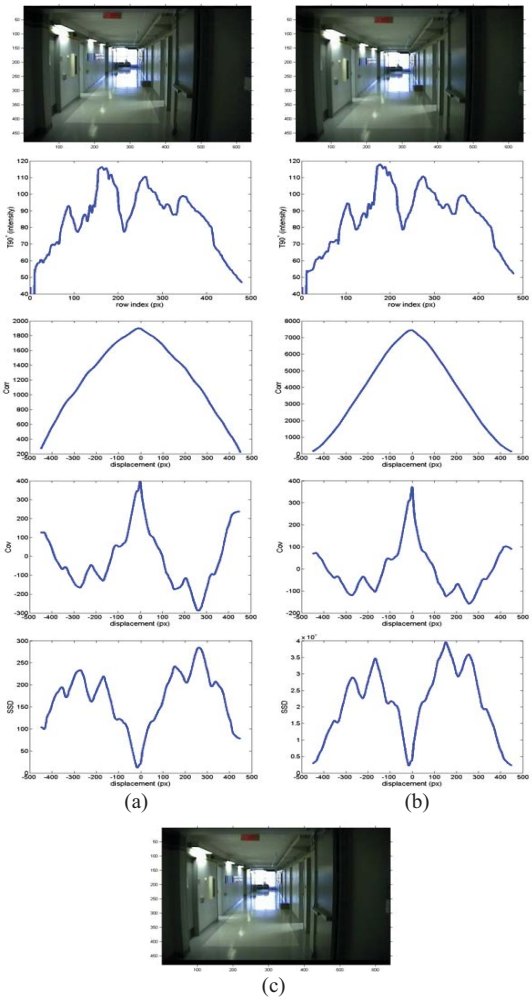


Fig. 3. First row: two consecutive frames in test sequence. Second row: vertical projections of input frames. Third row: normalized (left) and not normalized (right) correlation. Fourth row: normalized (left) and not normalized covariance. Fifth row: normalized (left) and not normalized sum of squared error measure. Sixth row: stabilized frame using sum of squared error.

B. Estimation of Motion Activity

A spatio-temporal anisotropic diffusion method is first applied that uses the information of the current, previous and next frames in the sequence. This method extends the anisotropic diffusion operator equation described in [9] by inserting a temporal variable in the anisotropic diffusion equation. A pixel motion activity measure is also introduced by estimating the accumulation of diffusion at the spatially and temporally homogeneous areas. In addition, a kernel density estimation approach is proposed to address the presence of outliers and disconnected areas, and produce more robust motion detection results.

1) Perona's Anisotropic Diffusion: Perona has proposed an anisotropic diffusion filtering for scalar images that avoid blurring and localization problems of the linear diffusion filtering. They applied an inhomogeneous process that reduces the diffusivity at those locations, which have a larger likelihood to be edges. This likelihood is expressed by the squared gradient. The proposed filter is based on the following PDE equation

$$\delta_s \cdot I = div(g(|\Delta I|^2) \cdot \Delta I) \tag{8}$$

where δ_s is the temporal gradient operator, $g(.)$ is a function that determines the amount of diffusion that is called diffusivity, $\bigtriangledown$ is the gradient operator and I is the image intensity.

In this work an implementation that uses finite differences, originally proposed in [9], was adopted and extended for the case of color images using the Euclidean vector distance to estimate the spatial and temporal gradients. In addition, the fraction type diffusivity function was employed, which is more suitable for region-oriented applications

$$g(\|\Delta I\|) = \frac{1}{1 + \left(\frac{\|\Delta I\|}{K}\right)^2} \tag{9}$$

where K denotes the conductance parameter for which a value from 4.5 to 5 is regularly selected for the purpose of segmentation.

Equation (8) can be equivalently written as

$$\frac{\partial I(x,y,s)}{\partial_S} = g(x,y,s)\Delta I(x,y,s) + \nabla_g(x,y,s)\cdot\nabla I(x,y,s)$$

where, Δ is the Laplacian operator and $g(x, y, s)$ is a constant. The discretization of this equation leads to the following PDE for the proof of 2D case (eq. 10):

$$I_{i,j}^{s+1} = {I_{i,j}}^{s} + \lambda\cdot[g(N)\cdot N + g(S)\cdot S + g(E)\cdot E + g(W)\cdot W]_{i,j}{}^{t}. \quad (10)$$

Here *N, S, E, W* denote the differences between the central pixel and its north, south, east and west nearest neighbors respectively. Furthermore $g(k)$ equals to the value of the diffusivity function in each direction.

2) Estimation of Pixel Motion Activity Using Spatio-Temporal Diffusion: In this section we explain the reasoning behind the proposed motion activity measure. This measure is based on the fact that after the spatio-temporal diffusion process the background areas are diffused more than the moving areas.

The main idea for the spatio-temporal non-linear filter is to extend the non-linear diffusion equation using the additional variable of time. Thus the PDE equation now becomes a straightforward extension of (11) except for the notations *PF* and *NF* that symbolize the difference between the central pixel and the previous and next frame nearest neighbors respectively. Other differences are the parameters λ_s and λ_t, which control the amount of spatial and temporal diffusion respectively. In this application it was experimentally found that the temporal diffusion must be significantly greater than the spatial for motion diffusion.

$$\begin{aligned}{I_{i,j,t}}^{s+1} = {}& {I_{i,j,t}}^{s} \\ &+ \lambda_s\cdot[g(N)\cdot N + g(S)\cdot S \\ &+ g(E)\cdot E + g(W)\cdot W]_{i,j,t}{}^{s} \\ &+ \lambda_t\cdot[g(PF)\cdot PF + g(NF)\cdot NF]_{i,j,t}{}^{s} \end{aligned} \quad (11)$$

where *PF* denotes the difference between the central pixels of the current and previous frame and *NF* is the difference between the central pixels of the current and next frame.

Let now s_1 be the initial scale and s_2 the final scale of spatio-temporal diffusion of a frame I. The scale $s = s_1 + 1$ is generated as follows:

$${I_{i,j,t}}^{s_1+1} = {I_{i,j,t}}^{s_1} + \Delta {I_{i,j,t}}^{s_1}. \quad (12)$$

Similarly to (12), for $s = s_1 + 2$ we have:

$${I_{i,j,t}}^{s_1+2} = {I_{i,j,t}}^{s_1+1} + \Delta {I_{i,j,t}}^{s_1+1}. \quad (13)$$

By substituting (12) into (13) we produce

$${I_{i,j,t}}^{s_1+2} = {I_{i,j,t}}^{s_1} + \Delta {I_{i,j,t}}^{s_1} + \Delta {I_{i,j,t}}^{s_1+1}.$$

The final diffusion between the first scale s_1 and final scale s_2 is calculated as

$${I_{i,j,t}}^{s_2} = {I_{i,j,t}}^{s_1} + \sum_{i=0}^{s_2-s_1} \Delta {I_{i,j,t}}^{s_1+i} \quad (14)$$

$$\sum \Delta I_{i,j,t} = {I_{i,j,t}}^{s_2} - {I_{i,j,t}}^{s_1} \Rightarrow \sum \Delta I_{i,j,t} = \sum_{i=0}^{s_2-s_1} \Delta {I_{i,j,t}}^{s_1+i}. \quad (15)$$

The quantity $\sum \Delta I_{i,j,t}$ expresses the amount of total diffusion for each pixel of the image. This quantity is increased in spatio-temporally homogeneous areas. This observation leads to the introduction of a measure of the pixel-wise motion activity as

$$PMA_{i,j,t} = 1 - \frac{\sum \Delta {I_{i,j,t}}^{2}}{\max(\sum \Delta {I_{i,j,t}}^{2})}. \quad (16)$$

According to our previous analysis, greater values of this measure denote higher motion activity areas.

3) Parzen Kernel Density Estimation: The previous process produces a pixel map of the motion activity. The pixels with higher values denote the moving areas. However the PMA index in (16) is subject to several temporal variations that do not correspond to motion. These variations may happen because of background noise and pattern noise caused by the camera, or local illumination changes that are more frequent when acquiring videos in outdoor environments.

Here we propose to resolve this pitfall by means of non-parametric density estimation that makes use of Parzen kernels. The employed feature space includes the horizontal and vertical spatial coordinates and the motion activity values. According to our assumption the areas of higher probability density correspond to the moving objects. The Parzen [8] density estimation belongs to the non-parametric density methods i.e. methods that do not impose any initial assumptions about the shape of the probability density functions. Its operation is based on placing at each observation sample a probability mass and producing a potential according to a Gaussian kernel. The contributions of all the sample points are averaged to estimate the density value at every point of the image. This process is known in the literature as non-parametric density estimation using Parzen kernels. The density value $f_h(x)$ produced by the N sample vectors x_i in position x, is computed as follows:

$$f_h(x) = \frac{1}{N\cdot h^p}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \quad (17)$$

where, h is the bandwidth, $K(.)$ is the kernel function and p is the kernel order. A common choice is the multivariate Gaussian kernel of order 2

$$K(x) = (2\pi)^{-p/2}\cdot\exp(-x^2/2).$$

Bandwidth h determines the kernel's decrease rate with distance and it turns out that the choice of h is much more important for the quality of the estimate than the choice of K. The practical consequences of bandwidth selection are obvious. If h has a big value the estimate will be too

smooth and might not reveal structural features like an existing bimodality. If it is too small the estimate $f_h(x)$ will suffer from statistical variability. In our experiments the h parameter was empirically set to 20, (Fig. 4). The Parzen kernel bandwidth value was determined as the largest bandwidth that still preserves the spatial localization of the more salient PMA features corresponding to real motion.

C. Motion Segmentation

The outcome of the previous process is a probability density map that denotes the areas of high activity. The next step is to apply spatio-temporal segmentation to detect the moving regions.

Based on the observation that the probability map forms a topographic relief that consists of peaks and valleys, it was concluded that a watershed-based approach may be used to identify the moving regions.

The watershed algorithm segments a grayscale or color image into disjoint regions by interpreting the image as a topographic relief. This analysis has emerged from mathematical morphology and was implemented by a series of pure mathematical morphology operators in its early versions. Several implementations have presented since then, such as iterative, sequential, arrowing, flow line oriented and flooding approaches. A flooding implementation is employed here similar to [13].

This operation can be divided into the stages of minima piercing and flooding. More specifically in the first stage the regional minima of the topographic surface of the image gradient magnitude are pierced and the water floods through them. The water progressively floods the catchment basins and corresponding barrages are subsequently built up at the points where water from adjacent minima would be mixed. This process is terminated when the whole surface is flooded and the barrages that were formed throughout this process correspond to the watershed lines.

According to the previous paragraphs the "center" of a moving area is represented by the maximum of probability density. In order to facilitate the watershed flooding process we use the reciprocals of the probability density values. The lowest minima are subsequently selected and flooded to form the moving areas.

IV. High-to-Low Representation of Resolution

This section deals with the loss of visual information due to reduced resolution. In particular, the Tyflos navigation device (section 2) uses two vision cameras to capture visual information from the surrounding environment. The effective resolution of the cameras used in our experiments is 640×480 pixels and we convert it into 256×256 pixels for compatibility reasons with the pyramidal approach. This visual information has to be converted into vibrations on a low-resolution array of 4×4 or 32×32 vibrators. Thus, **H**igh-resolution information is transformed into **L**ow-resolution resulting in unavoidable loss of information. Here we use several criteria during the H-L conversion in an effort to minimize the loss of important visual information (such as *obstacles in close proximity, motion, open space,* etc), of specific interest for the user.

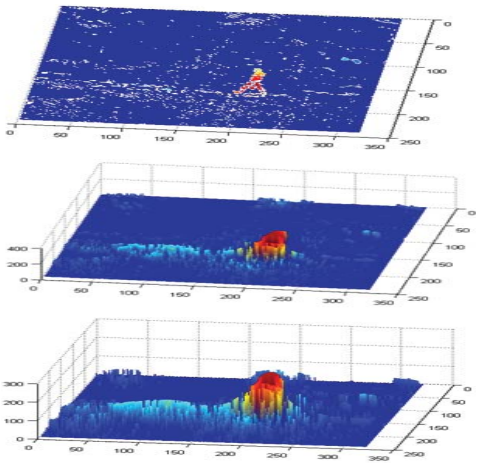


Fig. 4. For estimated density of motion activity map, we use Parzen kernels. This map is displayed in (a) and then represented as topographic relief as seen in (b) and (c). Its treatment as relief permits use of watershed transform for boundary delineation as described in Section III C.

A. Resolution of the Vibration Array

The resolution of a vibration array undergoes many restrictions. Hardware restrictions are: vibrator motors' dimensions (not less than 1 cm, which is the case); power consumption and electronics that are necessary to drive the vibrators. Additional restrictions come from the perception of tactile stimuli, which claim that, on the torso, a distance of minimum 1 to 1.5 cm is necessary for the distinction of two tactile stimulations. Taking all the above into account we decided that the maximum resolution of the vibration array can be 32×32. Our current prototype vibration array consists of 4×4 vibrator motors. The decision for this resolution was made in order to save on power consumption and on the bulkiness of the electronics that accompany the array. Furthermore, the cognitive load for the user is less and it makes us believe that a 4×4 will be a viable solution. The feedback during the experimental phase will direct us for any changes.

B. High-to-Low Representation and Improved Pyramidal Reductions

The cameras used in the current Tyflos prototype can capture images of 640×480. For computational reasons we convert them into 256×256 (stereo and motion detection).

Pyramidal reduction is a basic method to lower a resolution. The disadvantage, of the pyramidal reduction and the interpolation methods, is that they do not take into account any navigational issues i.e. close proximity obstacles, objects of interest, open/safe navigation paths and motion. Here, we use the pyramidal reduction to implement two high-to-low representations. The first will be called visual representation (VR) and the second 3D representation (3DR).

For the VR representation, we define:

Definition 1: The "merging" of the images captured by the two cameras is the so-called *fused-image*.

Definition 2: Pixels of interest are the pixels that belong to obstacles, motion, open-space, or object of interests (like persons).

TABLE I
VIBRATION LEVELS-VIBRATION FREQUENCIES-OBJECT DISTANCES AND COLOR REPRESENTATION

Vibration Level	Vibration Frequency [Hz]	Distance Range [m]	Distance Characterization	Color Representation	
3	10	(0,1)	Very close	White (0 × FF)	Burgundy
2	2	[1,2)	Close	Light gray (0 × C0)	Red
1	1	[2,4)	Far	Dark gray (0 ×80)	Yellow
0	0	[4,∞)	Very far	Black (0 ×00)	Light blue

Definition 3: The *image-of-interests* is an image that contains pixels of interest and has the same size with the fused image.

For visual understanding, a *white* pixel in an image of interest corresponds to a pixel that belongs to a close proximity obstacle, open path, etc), while a *black* pixel corresponds to pixel without interest (background).

For the 3DR representation the depth maps, that the stereo cameras produce, represent only 4 distance ranges, which they correspond to the 4 vibration levels.

Definition 4: The final depth map will be called *image-of-distances* is a 4–tone grayscale image.

The correspondence between vibration levels/frequencies on the VA, distances, and color on the *image-of-distances* is shown in Table I.

Here we provide a visual example of the pyramidal reductions (standard/classical and interest-based) to show the obtained different results. The two modified pyramidal reductions (VR, and 3DR) use criteria of interest (i.e. pixels of open paths, motion, close proximity obstacles, etc), while the standard pyramidal reduction uses only the average intensity criterion for the generation of the pixels at the lower level. Note that pyramids use images of power of 2.

In the 3D representation, the modified pyramidal reductions emphasize on open paths and on proximity objects/obstacle (Fig. 5a), whereas during the visual representation, the modified pyramids emphasize on the objects-of-interest (e.g. motion), as seen in Figure 5b.

C. Simulated Examples

Here we offer two simulated examples for better visual representation of the concepts and the 3D views of the navigation space. We will use a 32 × 32 vibration array. The vibration frequency will be represented in the z-axis. There are 4 vibration levels from 0 to 3, as previously reported in the Table I.

Two simulated cases will be shown covering different possible scenarios during navigation (Fig. 6).

Case 1: Side and vertical obstacles (e.g. person or an obstacle in a corridor).

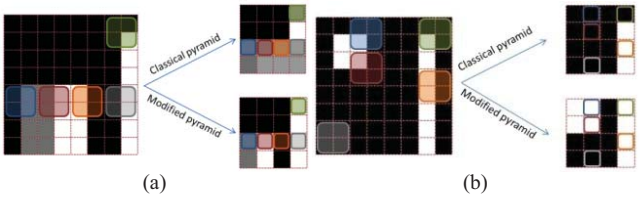


Fig. 5. Classical and modified pyramidal reductions from 8 × 8 to 4 × 4 resolution. Four pixels from higher level create one pixel at lower pyramidal level (shown with same border color). (a) In 3DR, classical pyramid does not keep important information for proximity objects (more white) or open paths (black). (b) In VR, upper left object of interest (white) is preserved when using modified pyramid.

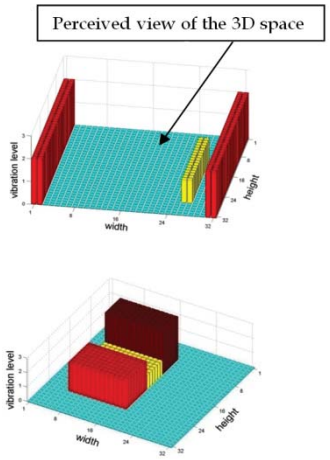


Fig. 6. Simulated examples. (a) case 1. (b) case 2.

Case 2: Side and vertical obstacles (e.g. person and two obstacles around the person in a corridor).

V. TESTING RULES WITH PATTERNS

In this section we will provide results from training of subjects on the (4 × 4) haptic array. Simple training cases were used to prove the concept that users are capable of perceiving information about the navigation space through an advance haptic interface such as the Vibration Array (VA).

A. Pattern Generation

The 16 elements can create $2^{16} = 65{,}536$ different vibration patterns and this is a simplified case, considering only 2 vibration states (vibration and no-vibration) and not the 4 vibration levels that the system is capable of producing. Experimental results in haptic devices [18] show that the users need to learn these patterns to some extent so that they will be able to understand them and distinguish one from another. Thus, the large number of patterns corresponds to a very heavy cognitive load, which has to be reduced.

Here the goal is to select a set of patterns that the user will be able to learn and distinguish from each other. This set will be selected using criteria based on:

1) different possible navigation scenarios
2) safe navigation
3) simplicity.

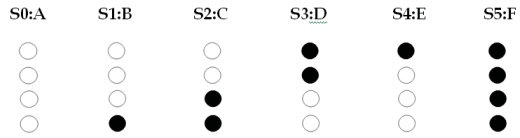


Fig. 7. Six vibration symbols (activated cells are colored black).

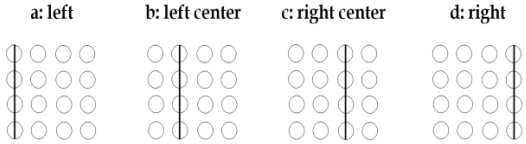


Fig. 8. Four different positions/directions that symbols can appear on vibration array.

The selection of the patterns will be based on two rule-generation approaches: the vertical and the horizontal. These approaches investigate and create a set of rules based on human factors and navigational issues.

B. Vertical Rules

During navigation the most important information is whether there is an open path to navigate, which corresponds to a specific direction (x-coordinates) but the major advantage of the two-dimensional array, compared to the one-dimensional, is that it can inform the user about how high or low an object is (y-coordinate). Thus, if the path is not open then the user can be informed for the position of the obstacle in the y-coordinates. We argue here that this information doesn't have to be very detailed. Thus, some rules are set to reduce the number of patterns that can appear in the vertical dimension (y):

1) *If three cells on the same column are activated then all column is activated*. Explanation: the path is fully blocked and it is not of great importance for the user to know more details.
2) *If two non-consecutive cells in the same column are activated then the middle cell will be activated.* Explanation: Simplifies the obstacle representation.
3) *If the lowest cell is not activated the cell above is activated then the lowest cell will be activated.* Explanation: mid-height obstacles don't leave enough useful space below them so they have to be considered as blocked space too.
4) Similar to rule 3 but for *overhanging obstacles*.
5) *If both overhanging and ground obstacles exist then the path is considered fully blocked and all the cells will be activated.*

The above rules can define 6 types of column symbols (Fig. 7) that can appear in four different positions on the array (Fig. 8) reducing the number of navigation patterns from 64 to 12.

C. Horizontal Rules

In this approach, the rules are set by comparing a symbol with the symbols next to it. The idea is that the information that a symbol carries can be correlated with the neighboring symbols by emphasizing or de-emphasizing it. For example, if a large ground object is between two tall obstacles, it can be de-emphasized and be considered as a small obstacle, keeping its nature as a ground obstacle but emphasizing the nature of the tall ones. We have defined the following rule base for the horizontal dimension (Figure 7):

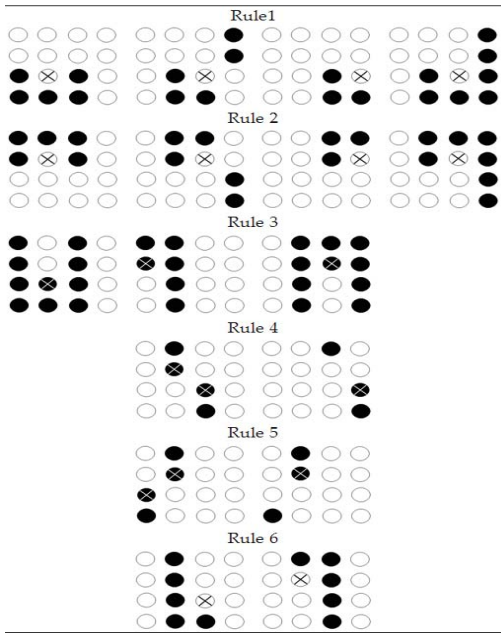


Fig. 9. Patterns where horizontal rules apply. After application of rules white cells with X are enabled and black cells with X are disabled.

1) *If S1 has S2 on one side, AND, the other side is S2/3/5/B then S1 will be transformed to S2 (Note: SB is the pseudo-symbol of the border of the array).*
2) *If S4 has S3 on one side, AND, the other side is S2/3/5/B, then S4 will be transformed to S3.*
3) *If S2/3 has S5 on one side and S5/B on the other, then S2 will be transformed to S1 and S3 to S4.*
4) *If S2 has S3/4 on one side then S2 will be transformed to S1 and the side S3/4 to S4*
5) *If S3 has S1/2 on one side then S3 will transformed to S4 and the side S2 to S1.*
6) *If S1/4 is next to S0 then transform S1/4 to S2/3.*

Examples of the above rules are shown in Figure 9, in the experimental part. By applying the above rules, 998 symbols are excluded leaving 298 symbols available keeping only 0.45% of the initial number of symbols.

D. Test

A series of tests was performed for evaluating the user's ability to recognize symbols and patterns on the vibration array. A number of sighted subjects were used initially to test the vibration array on a variety of patterns on specific locations generated by the method before the actual tests on visually impaired. After the successful testing on typical users, 10 visually impaired subjects (6 males, 4 females) were selected with their ages varying from 14 to 60. They had never been trained with or used the 2D VA or other tactile devices.

TABLE II
USERS' ACCURACY IN IDENTIFYING SYMBOLS
(SYMBOLS AT SAME POSITION)

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
95.0	92.0	79.3	77.3	83.3	88.0	89.3	66.7	82.7	90.7

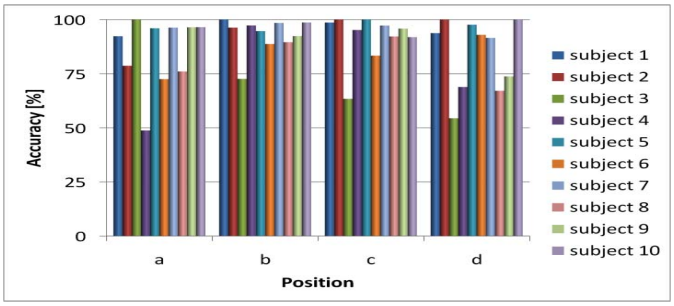


Fig. 10. Users' accuracy in identifying positions (random symbol).

They were asked to wear the vest and we made sure that all the vibrating motors are in contact with their body.

Before every experiment we explain to each user the purpose of it and run a short demo of the experiment so that they will get familiar with the different patterns. During the experiment the subjects were standing so that the vest is making proper contact with their body and we tried to minimize our interaction with the subject for not distracting him/her. We performed 300 measurements/trials for each experiment.

The subjects had very good accuracy (greater than 92%) in identifying the position of a symbol when a specific symbol was used and very good accuracy (in most cases over 80%) in identifying the symbol for a specified position (Table II). On the contrary, when a random symbol was used in random position, the percentages are considerably smaller; position accuracy is shown in Figure 10 and symbol accuracy in Figure 11. Our possible explanations for this difference are the fatigue heavy cognitive load (match haptic information with symbol/position name).

The users had difficulty to distinguish vibrating sensations between one or two consecutive vibrating cells, although the spatial acuity on the torso is 1.5-2 cm [18] higher than the one of the array so we believe that with adequate training this will be overcome.

Despite the problems discussed above, overall the accuracy percentages are still high enough, considering that all subjects were not trained or ever had experience with tactile feedback experiments. For future work we will extend the experiments with a goal to define a set of recognizable patterns by the user represented in a low resolution (4×4) in order for the user to evaluate the open space.

It is also important to report here that the identification of the various levels of frequency by the subjects was not that high (Fig. 12) due to the lack of training with them.

VI. EXPERIMENTAL RESULTS

Three indoor navigation scenarios were chosen to demonstrate the high-to-low representations. In scenario 1 (Fig. 13)

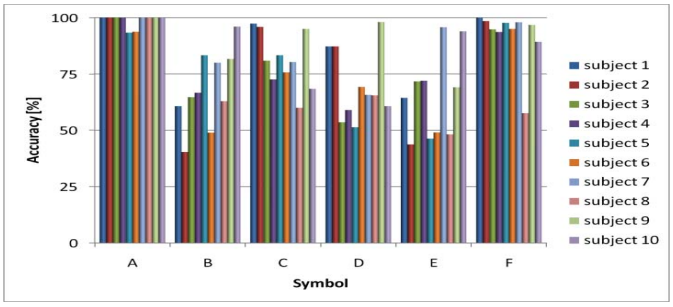


Fig. 11. Users' accuracy in identifying symbols (symbols at random position).

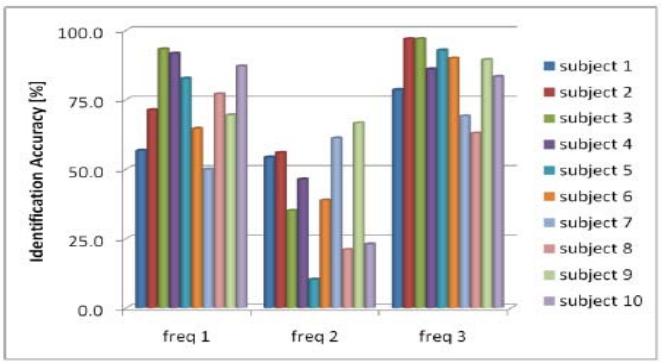


Fig. 12. Identification accuracy for different vibration levels/frequencies.

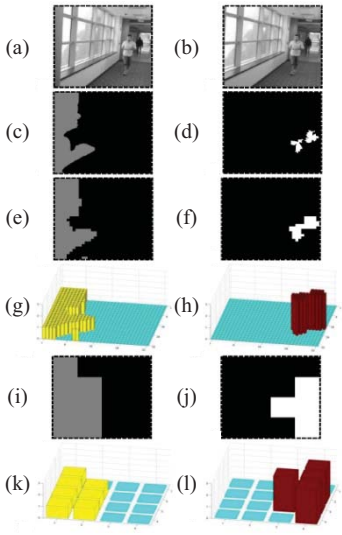


Fig. 13. Navigation scenario 1.

we notice that the moving persons are detected using the proposed motion detection method and they are presented to the user, as they are considered as objects-of-interest. Navigation scenario 2 is a similar case (Fig. 14). In navigation scenario 3 (Fig. 15) we notice that the moving person is detected as a close obstacle but the motion detection method emphasizes it.

In addition, some video results are shown in Figure 17. Each video consists of sequences of groups of 4 frames. The cameras capture images from an inside environment. These images are represented as high-resolution frames at the upper left (left camera) and right (right camera) corners of each

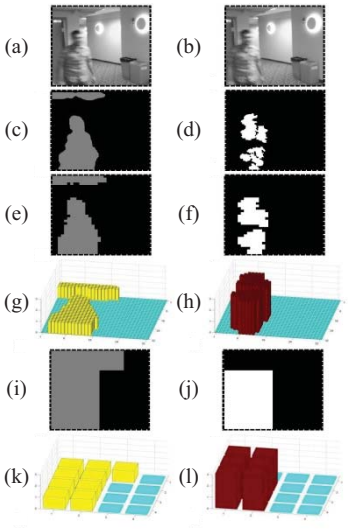


Fig. 14. Navigation scenario 2.

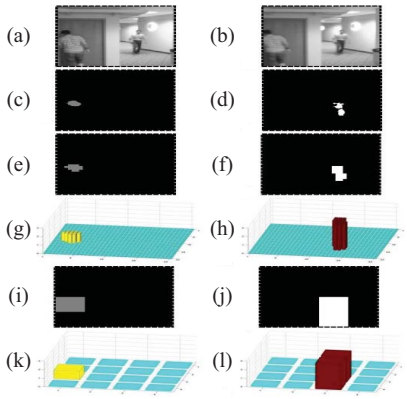


Fig. 15. Navigation scenario 3 (similar to Fig. 13). (c) shows a small gray area and a black region around it as only close object to cameras is human on left of image and in distance $\sim$4 m.

group. At the bottom left side of the group is the corresponding high-resolution disparity map of the image; and at the right lower frame is the 3D vibration-tactile vocabulary pattern sent to the vibration array. Black pixels correspond to no vibrations; dark gray to Figure 13 shows the navigation scenario 1: (a) and (b) are the left and right frames captured by the stereo cameras; (c) it is the simplified *image-of-distances*, which represents the depth-image from obstacles with max distance of 4 m; (d) are the regions of high activity like motion (image-of-interests), which are detected using 2 previous frames; (e) is the low resolution (32×32) image-of-distances and (f) is the image-of-interests after applying the modified pyramidal reduction. (g) and (h) give a 3D view of the vibration array on the user. Finally (i) and (j) are the lowest (4×4) resolutions with the corresponding 3D view of the vibration array on (k) and (l). In Figure 14 the navigation scenario 2 is similar to Figure 13.

In the Figures 16 and 17 we present frames from one of several videos to show the usefulness of the Tyflos device for detecting moving objects (here a human) and how the vibration patterns are presented on the vibration vest in the form of regions with different vibration strength.

Fig. 16. Each frame consists of four views. Upper left and upper right views are left and right images captured by cameras. Lower left view is disparity map-image by using four level of intensity, and lower right is low-resolution image that reflects to four vibration levels of vi-bration array.

More videos are available in http://www.cs.wright.edu/atrc/a%20wearable%20navigation%20aid%20for%20blind.html.

It is important to be mentioned here that the method presented here and its vibration array (vest) has some similarity with the Braille code approach, but at the same time is very different in the way that offers 3D vibration sensations for all or part of vibrators, where the operation of one vibrator may have an impact on its neighbors. Actually this is a topic that our team is working with for larger scale arrays of vibrators for separating (or segmenting) 3D vibration patterns with noise.

A. Various Navigation Scenarios

The last experimental part involves testing with real navigation scenarios. Eleven videos were captured from inside the school of engineering. The videos include different possible navigation scenarios, including moving or static people, overhanging obstacles, low height obstacles, doors etc.

The videos from the left and right cameras were processed off-line (stereo, high-to-low, pattern matching) and a selection of the final vocabulary frames (179 in total) from each scenario was presented to the users. Some characteristic frames including the left/right camera and the high-resolution disparity map are shown in Fig. 16. The users were asked to do describe the vibration pattern sent to them. This includes position of open paths and obstacles with distance/frequency information.

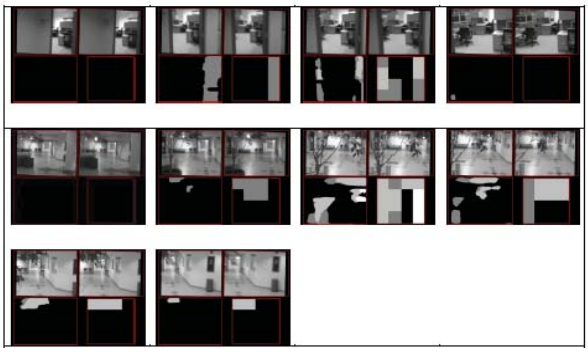
For better understanding of the sequences of frames presented below, we explain what the four views of each frame represent in Figure 16.

Note here that for the examples above the selected frames are arranged in groups of 4. Top left and top right are left and right high-resolution camera frames. Bottom left is the high-resolution disparity map. Bottom left is the (4×4; low resolution) tactile vocabulary pattern sent to the vibration array. Black pixels correspond to no vibrations; dark gray to vibration frequency 1; light gray to frequency 3 and white to frequency 3 (the highest).

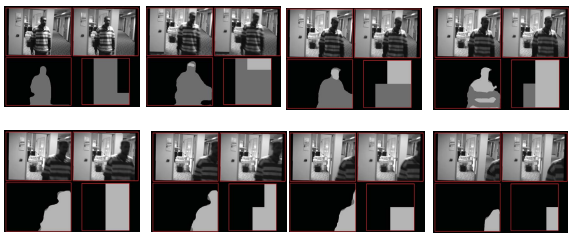
B. Discussion on Patterns and Navigation Scenarios

The experimental results from vibration patterns and navigation scenarios are presented in Tables III and IV and they give us some important feedback for the evaluation and further improvement of the Tyflos Navigation prototype but before further discussion we want to emphasize on two points.

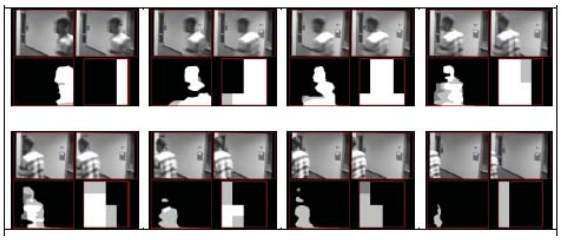
Video Frames:
Video -1: The user wearing the Tyflos device is navigating through a door into an open space (four top frames).
Video-2: The user is navigating in an open space with various obstacles (small trees, boxes, corners).



Video-3: A walking person crosses the view of the user (who is wearing the Tyflos device). The video sequence shows how the moving obstacle (walking man) blocks the view and how the low-resolution image offers such pieces of information to the user via these 4 levels of vibration.



VIdeo-4: Here is another example of a walking person in front of the Tyflos device, and how the visual information is represented at the 4 levels in order to inform the user via vibrations, as it was described in previous sections.



Video-5: In this sequence there is an open space for navigation.

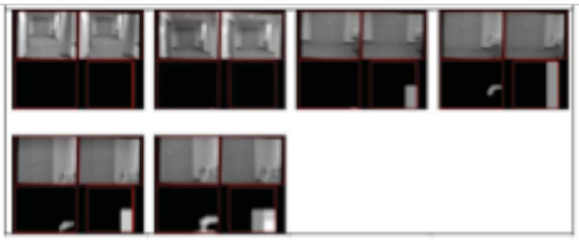


Fig. 17. It shows five video sequences of six to ten group-frames, processing results from H-L representation, results reflecting motion and free navigation space. In this video, human subject is moving toward and in front of Tyflos device and consecutive vibration frames (lower level) show detection of motion (white areas) and distance from device.

The users did not receive any training with our prototype or any other system with tactile interface. Also, the navigation scenarios are captured from real-life indoor environments, not synthetic or simulated.

TABLE III
CORRECT IDENTIFICATION OF PATTERNS WITHOUT VIBRATIONS, OPEN PATHS WITHIN PATTERNS, AND NONZERO FREQUENCIES

Subject	Empty images (%)	Open paths (%)	Nonzero freqs
0	97.0	87.4	26.6
1	100.0	96.4	34.4
2	97.0	93.2	32.5
3	97.0	86.5	27.6
4	97.0	93.7	35.2
5	100.0	89.6	18.5
6	90.9	95.0	26.7
7	97.0	91.4	30.9
8	100.0	86.9	42.6
9	100.0	94.6	30.5
AVERAGE	***97.6***	***91.5***	***30.6***

TABLE IV
AVERAGE CONFUSION MATRIX FOR DIFFERENT FREQUENCIES

	Guess			
Actual	0	1	2	3
0	**91.6**	4.0	3.2	1.3
1	36.8	**27.6**	31.8	3.8
2	34.1	15.0	**35.8**	15.1
3	30.4	4.6	29.3	**35.7**

The most important issue for a visually impaired user, during navigation, is whether there are open safe navigation paths and in which direction. For our current prototype a safe navigation path is determined by a column that is free of vibrations. Table IV shows that in 97.6% of the frames the users were able to detect a fully open frame (scene without obstacles), which is a relatively easy task but important to notice. We also see that in 91.5% of the cases the users were able to identify correctly the open paths in images containing obstacles.

We test the same scenarios with several videos taken from real cases and the blind users were tested to determine if they have sense open-space for navigation or moving obstacles in front of them. From the videos the low-resolution images were converted into array patterns and these patterns were transferred onto the vibration array and the blind users respond to them. The results from these experiments (Figure 17) have shown that the users, although had no previous training, were able to correctly detect open space, obstacles and moving obstacles. More specifically, in video-1, all the users detect open space in the frames 1 & 4 and obstacles for the frame 2. However, only four users have detected partially open space in frame 3. In video-2, all the users have detected open space in frames 1and obstacles in frames 2-6, three users, however, detected partial open space in frames 2, 5, & 6. In video-3, all the users have detected moving objects, and eight of them have detected partial open space in frames 5–8. In video-4, all the users have detected moving objects in frames 1–8 and

partial open space in the frame 8. Eight users have detected partial open space in frames 6 & 7. In video-5, all the users have detected open space in the frame 1, 2 & 3, and partial open space in frames 4, 5 & 6.

VII. Conclusion

In this work an assistive wearable prototype-device for the visually impaired was presented. The device was using two cameras to capture images from the environment and a vest with an array of vibrators. A method that combines computer vision techniques was used to detect moving objects and their distance from the two cameras, by providing depth map-images. Then, these images were reduced into a low resolution ones by using pyramidal reduction techniques. Finally the reduced depth-map images were mapped onto the array of vibrators to generate a 3D vibrating sensation of the surrounding 3D space for assisting a visually impaired user to "sense" collision free navigation corridors.

In particular, the first part of this method (motion and obstacle representation) is very important for the efficient reduction of visual information and the preservation of the interest pieces of information in the second part. Thus, the first parts included a video stabilization approach, a spatio-temporal diffusion model that produces a motion activity measure and a Parzen kernel-based method to create a probability map for moving objects. The final area is detected by means of watershed segmentation. The second part (H-L representation) included a pyramidal scheme for reducing the image resolution from high (640×480 pixels or converted 256×256) to low resolution (32×32 or 4×4 pixels) based on various criteria of interest and vibration representation scheme. This scheme generates 3D patterns from the corresponding low-resolution images after the H-L conversion. The 3D vibration patterns are applied on the vibration vest to generate the 3D sensation patterns on the user's body.

Results obtained by using this prototype system on a group of visually impaired participants (10) have shown that this is a very useful assistive-device for the users. Note that at WSU there is a large population of people with disabilities (more than 500). Although the tests were performed on individuals without previous training, the results were satisfactory and the expectation after training is very high for much better results. Note here that the role of this device will be assistive but not a replacement of the cane, which the blind and visually impaired consider as an extension of their hand [23].

The proposed methodologies were developed to operate in an integrated environment (Tyflos navigator) as an assistive vision technology tool-device for motion detection, tracking and obstacle representation. The performance of the current of the Tyflos navigator prototype is 2–3 frames per sec using a portable computer with a 2.66 GHz Intel Core 2 Duo processor. Our future objectives are to improve the computational speed of the navigator by using FPGAs for certain algorithms, and conduct a six months training and perform the same experiments to evaluate the differences. We also plan to include an object recognition scheme to recognize simple objects and a face recognition scheme to identify humans and to investigate the real-time implementation of the proposed scheme using FPGAs.

Acknowledgment

The authors would like to thank the anonymous reviewers for their fruitful comments improving the quality of this paper.

References

[1] D. Dakopoulos and N. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[2] S. Makrogiannis and N. Bourbakis, "Motion analysis with application to assistive vision technology," in *Proc. 16th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 2004, pp. 344–352.

[3] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Efficient moving object segmentation algorithm using background registration techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 577–586, Jul. 2002.

[4] B. K. P. Horn, *Robot Vision*. Cambridge, MA, USA: MIT Press, 1986.

[5] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[7] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 1154–1160.

[8] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–597, Jun. 1991.

[9] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539–546, Sep. 1998.

[10] N. Bourbakis and D. Kavraki, "Intelligent Assistants for handicapped peo-ple's independence: Case study," in *Proc. IEEE Int. Joint Symp. Intell. Syst.*, Nov. 1996, pp. 337–344.

[11] N. Bourbakis and D. Kavraki, "An Intelligent agent for blind people navigation," in *Proc. IEEE Symp. Bioinformat. Bioeng.*, Rockville, MD, USA, Nov. 2001, pp. 230–235.

[12] N. Bourbakis, "A 3D dynamic space sensing using a 2D vibration pad for blinds navigation," in *Proc. IEEE Symp. Bioinformat. Bioeng.*, Oct. 2005, pp. 222–226.

[13] D. Dakopoulos, "A wearable navigation prototype-system for visu-ally impaired: Design, modeling and experimental results," Ph.D. dissertation, Dept. Comput. Sci. Eng., Wright State Univ., Dayton OH, USA, 2009.

[14] *National Federation of the Blind*. (2009, May 8) [Online]. Available: http://www.nfb.org/

[15] B. B. Blasch, W. R. Wiener, and R. L. Welsh, *Foundations of Orien-Tation and Mobility*, 2nd ed. New York, NY, USA: AFB Press, 1997.

[16] T. Ifukube, T. Sasaki, and C. Peng, "A blind mobility aid modeled after echolocation of bats," *IEEE Trans. Biomed., Eng.*, vol. 38, no. 5, pp. 461–465, May 1991.

[17] S. Shoval, J. Borenstein, and Y. Koren, "Mobile robot obstacle avoidance in a computerized travel aid for the blind," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Diego, CA, USA, May 1994, pp. 2023–2029.

[18] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 112–121, Feb. 1992.

[19] A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," in *Proc. 6th Int. ACM SIGACCESS Conf. Comput. Accessibility*, Jan. 2004, pp. 147–152.

[20] D. Aguerrevere, M. Choudhury, and A. Barreto, "Portable 3D sound/sonar navigation system for blind individuals," in *Proc. 2nd Int. Latin Amer. Caribbean Conf. Eng. Technol.*, Miami, FL, USA, Jun. 2004, pp. 1–6.

[21] J. L. González-Mora, A. Rodríguez-Hernández, L. F. Rodríguez-Ramos, L. Díaz-Saco, and N. Sosa. (2009, May 8). *Development of a New Space Perception System for Blind People, Based on the Creation of a Virtual Acoustic Space* [Online]. Available: http://www.iac.es/proyect/eavi

[22] G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 257–264, Jan. 2007.

[23] R. Audette, J. Balthazaar, C. Dunk, and J. Zelek, "A stereo-vision system for the visually impaired," School of Engineering, Univ. Guelph, Guelph, ON, Canada, Tech. Rep. 2000-41x-1, 2000.

[24] I. Ulrich and J. Borenstein, "The guideCane-applying mobile robot technologies to assist the visually impaired people," *IEEE Trans. Syst., Man Cybern. A, Syst., Humans*, vol. 31, no. 2, pp. 131–136, Mar. 2001.

[25] S. Meers and K. Ward, "A substitute vision system for providing 3D perception and GPS navigation via electro-tactile stimulation," in *Proc. 1st Int. Conf. Sens. Technol.*, Nov. 2005, pp. 1–5.

[26] K. Ito, M. Okamoto, J. Akita, T. Ono, I. Gyobu, T. Tagaki, T. Hoshi, and Y. Mishima, "CyARM: An alternative aid device for blind per-sons," in *Proc. Extended Abstracts Human Factors Comput. Syst. CHI*, Portland, OR, USA, Apr. 2005, pp. 1483–1488.

[27] M. Bouzit, A. Chaibi, K. J. De Laurentis, and C. Mavroidis, "Tactile feedback navigation handle for the visually impaired," in *Proc. ASME Int. Mech. Eng. Congr. Expo.*, Anaheim, CA, USA, Nov. 2004, pp. 1–7.

[28] C. Shah, M. Bouzit, M. Youssef, and L. Vasquez, "Evaluation of RU-netra- tactile feedback navigation system for the visually impaired," in *Proc. Int. Workshop Virtual Rehabil.*, New York, NY, USA, 2006, pp. 71–77.

[29] L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, NY, USA, 2006, pp. 6298–6292.

[30] S. Cardin, D. Thalmann, and F. Vexo, "A wearable system for mobility improvement of visually impaired people," *Visual Comput.*, vol. 23, no. 2, pp. 109–118, Jan. 2007.

[31] M. Adjouadi, "A man-machine vision interface for sensing the environment," *J. Rehabil. Res. Develop.*, vol. 29, no. 2, pp. 57–56, 1992.

[32] D. Yuan and R. Manduchi, "A tool for range sensing and environment discovery for the blind," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, vol. 3. Jun. 2004, p. 39.

[33] T. C. Davies, C. M. Burns, and S. D. Pinder, "Mobility interfaces for the visually impaired: What's missing?" in *Proc. 8th Int. Conf. Comput.-Human Interaction: Design Centered HCI*, vol. 254. Jul. 2007, pp. 41–47.

[34] R. D. Easton, "Inherent problems of attempts to apply sonar and vibrotactile sensory aid technology to the perceptual needs of the blind," *Optom. Vis. Sci.*, vol. 69, no. 1, pp. 3–14, 1992

[35] R. Ravi, D. Pawluk, T. V. Ketchum, and M. Jessica, "Issues of using tactile mice by individuals who are blind and visually impaired," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 3, pp. 311–318, Jun. 2010.

Nikolaos G. Bourbakis (A'80-SM'88-F'96) received the Ph.D. degree in CE & informatics from the University of Patras, Patras, Greece.

He is currently an OBR Distinguished Professor of IT, the Director of the ATR center, and a J. A. Professor of Geriatrics with Wright State University, Dayton, OH, USA. He pursues research in applied AI, machine vision, bioengineering, assistive technologies, cybersecurity, and parallel/distributed processing. He has published in refereed International Journals and Conference Proceedings.

Dr. Bourbakis received the IEEE Computer Society Outstanding Award, the IEEE Technical Achievement Award, and the IEEE Russ Bioengineering Award.

Sokratis K. Makrogriannis (M'98) received the B.S. degree in physics, the M.S. degree in electronics, and the Ph.D. degree in image analysis from the University of Patras, Patras, Greece, in 1995, 1998, and 2002, respectively.

He is currently with the Translational Gerontology Group, National Institute on Aging of, NIH, Baltimore, MD, USA. He was with Visualization and CV Laboratory, General Electric's Global Research Center, University of Pennsylvania, Wright State University, Dayton, OH, USA, and Glaxo-SmithKline's R&D. His current research interests include image analysis and computer vision.

Dimitrios Dakopoulos received the B.S. degree in physics from the National University of Athens, Athens, Greece, and the Ph.D. degree in computer science and engineering from Wright State University, Dayton, OH, USA.

He is currently with a Greek Company. His current research interests include assistive technologies, human-computer interaction, computer vision, image processing, and time series analysis.