AN INTELLIGENT ASSISTANT FOR NAVIGATION OF VISUALLY IMPAIRED PEOPLE

N.G. Bourbakis*# and D. Kavraki #
#AIIS Inc., Vestal, NY, email: aiis@stny.rr.com
**WSU, ITRI, Image-Video-Machine Vision Research Lab, Dayton OH*

Abstract
This paper presents the navigation methodology employed by an intelligent assistant (agent) for people with disabilities partial independence. In particular, the intelligent assistant, called Tyflos, would help a visually impaired user to be partially independent and able to walk and work in a 3-D dynamic environment. The Tyflos system carries two vision cameras and captures images from the surrounding 3-D environment, either by the user's command or in a continued mode (video), then it converts these images into verbal descriptions for each image into a verbal communication with the user. With other words the system plays the role of human assistant, who describes to the user the 3-D visual environment.

***KEYWORDS**: AI Systems, 3-D Visual Navigation, Knowledge Representation and Conversion, Multimedia, Intelligent Agents for People with disabilities*.

1. INTRODUCTION

Typical humans are gifted by the nature with multiple senses that allow them to develop a good 3-D perception of the surrounding world. Unfortunately, a number of people for different reasons have lost or deprived their visual sense. In particular, in USA there are more than 3 millions with low vision and millions more completely blind. This issue stands alone as a great challenge for the scientific community to develop a system able to assist visually impaired people to obtain verbal descriptions of the surrounding environment and increase their independence. Thus, the development of a robust robotic-like vision system, able to obtain 3-D representations of the surrounding world and provide verbal descriptions of it, will be a great contribution and relief for millions of blind people. Of particular importance is how a vision system will formulate and synthesize high resolution 3-D images taken from an unknown environment, analyze them by extracting the description of objects and events based on simple queries. In addition, how the vision system will convert visual descriptions into verbal communication.

Recently the modern computer engineering, the applied artificial intelligence and software technologies have advanced to the stage of successfully supporting large scale complex engineering and scientific projects, such as QBIC a multimedia based system [9], EUROTRA an automatic natural languages translation system [10], STARS for automatic software development at mega-programming level [11], MEL an autonomous intelligent vehicle [12], intelligent walking robots [13-15], 3-D space maps generation [14], autonomous vision systems [16], neuromorphic vision chips [17], the robot with feelings [18], etc.

These technological achievements have reached a level of maturity and could successfully be used to serve the humanity in very sensitive fields, such as visually impaired people. Recent challenges came to the scientific surface with the development of prototype artificial eyes [7], available after 2010. Also, implants for blind people [2,19,20] will be available at that time.

These technological breakthroughs need some time before to be available to the public. In the meantime, the advances of the state of the art in intelligent technologies could help these evolutionary efforts above by developing intelligent assistants to help handicapped people. In particular, an intelligent system to make visually impaired people to "verbally see" via a machine vision system equipped with speech and natural language understanding technologies. Moreover these technologies are now available and more than 70% functional, and there is no need of waiting 15 or more years to use them.

Based on these challenges and technological achievements above, this paper the navigation methodology of an intelligent assistant for visually impaired people's partial independence. The intelligent assistant, called Tyflos, would help a blind user to be independent and able to walk and work in a 3-D dynamic environment. The system captures images from the surrounding 3-D environment, either by the user's command or in a continued mode, and converts the visual description of each image into a verbal (low or high level) descriptions. With other words the system plays the role of human assistant, who describes to the user the 3-D visual environment.

2. THE TYFLOS SYSTEM [1]

2.1. System's Main Components

The global configuration of the Tyflos system shown in figure 1. It consists of seven main parts, such as the pair of glasses, the vision cameras, the laser range scanner, the microphone, the portable computer, the ear speaker, and the communication system.

. The Pair of Glasses

The pair of glasses is a important part of the Tyflos system, and several system's parts are mounted on it, see figure 2.

. The Vision Cameras

The vision camera is mounted on the top of the glasses in order to provide a better view of the targeted scene. The vision camera is associated with a system of adjustable lenses that provide to the system the ability to capture images from long and short distance according to the user wishes. The camera is high resolution, color, small in size and weight.

. The Laser Range Scanner

The laser range scanner is attached to the vision camera, with the ability to "scan" the same view with the camera. It is used for the determination of the distances of the existing objects and for the 3-D view of a captured image. The size of the laser scanner has also to be small and with resolution at least 1cm radius in a distance of 15 meters.

. The Ear Speaker

This part is very important for the user. In particular, the Tyflos system has only one ear speaker through it the user receives the detailed description of the surrounding environment, and gets explanation about the sounds that the user's free ear receives. The ear speaker has to be a high fidelity component.

. The Microphone

The microphone is the device through it the user communicates with computer and the appropriate Databases. It has a filter to reduce or if it possible to eliminate the noise from the user's environment.

. The Portable Computer

The portable powerful computer is the most valuable component of the Tyflos system. It hosts all the important software tools that make the Tyflos system functional. In particular, it hosts all the DBs (Image, Speech, Natural Language DBs), Software Interfaces, Knowledge Conversion Tools, Communication Tools, etc.

. The Communication System

This component is important when the user needs to establish communication (email, software transfer) with other resources, such as portable or mobile information DBs, emergency centers and etc.

3. HUMAN TYFLOS INTERACTION

The Tyflos system main role is to assist a visually impaired person by describing in natural language words the scene that surrounds the human user. Some important cases that it is able to assist the user are:

. where am I?

. describe the surrounding scene.

. guide me to a particular point

. where is a particular object?

With other words it will play the role of the human assistant by describing to the user the 3-D visual environment making him/her more independent. In particular, when the user wishes to find an "object" in the 3-D environment that he/she is working, he/she speaks though the microphone asking for that particular "object". The voice is converted into a natural language text (command), which accesses that portable DB. The system then requires from the user to slowly rotate his/her head scanning the surrounding space. Each image from a sequence of images captured by the vision camera is analyzed and recognized by the portable computer. If the "object" asked by the user is detected and recognized, then the system informs the user about it though the ear speaker. With a similar way the system explains and describes to the user via the ear speaker what object made a certain sound and where that object is from the current position of the user.

. Visual to Audio Communication

The conversion of information plays one of the most critical roles in the Tyflos system operation. In particular, visual information extracted from the surrounding environment is converted into speech. There are different cases:

A. Where am I?

There is available in the market a system that uses GPS coordinates to define the position of its user on a particular city map available in a DB. The human user presses a certain button of the device activating the system's receiver for receiving the appropriate GPS coordinates of his/her position. These coordinates activate a map Database from which a predetermined spoken message verbally informs the user about his/her location. The system uses a speech synthesizer. This a useful device for city navigation and mainly provides to the user the names of the streets and/or the names of certain buildings. The Tyflos system is coming to complement the abilities of such devices by providing current information related with the location of the user not only inside cities but also at any place with or without changes. To achieve this goal, Tyflos provides to the human user a verbal description of the surrounding 3-D scene. Thus, for the question *Where am I?* may use the GPS capabilities or provide additional information about the surrounding scene as given below.

B. Describe the Surrounding Scene

In this case a part of the surrounding scene is described. In particular, the description of each main recognizable object (landmark) and its relationships with the other objects in the same scene are extracted and described by SPN graphs. Then these SPN graphs are transformed into NL sentences. The sentences are appropriately converted into speech patterns, which are played by the ear speakers. More specifically, the description of the surrounding scene follows the next steps.

. Generation and Description of a 3-D Surrounding Unknown Environment

An additional step of improving the performance of the Tyflos system is the development of a 3-D images via stereo speaker [22] and compare them with the ones generated by the vision camera and fused with the laser range data from the same scene [5,22]. In particular, Dr. Albus and Dr. Bourbakis have an interest to further investigate and develop 3-D images via audio signals generated by two speakers during the motion of a sound generation equipment. In particular, a 3-D image can be created by using the sounds generated by the equipment and define the highs and lows (pitch), the width between left and right (face), and the depth (density). Thus, if a human brain can be trained to understand these sounds then the human user may have a 3-D understanding of the surrounding space. The 3-D space model generated by the stereo speaker method is not precise and it does not provide information about the surrounding objects. Thus, the 3-D image model generated by the sound approach will be fused with the 3-D image generated by a laser scanner and 2-D imager to finally produce a more accurate model of the 3-D space.

. Fusion of Images & range Data for 3-D Modeling of the Free Space

The fusion methodology will be based on the integration of different sensory data generated by a vision system and a laser scanner under noisy conditions for the 3-D representation of the surrounding free space.

Figure 1 graphically shows the generation the steps used by the fusion methodology. In particular, the range matrix generated by the laser scanner is expanded by using the ratio produced by comparison with the real color image. Then the color image is segmented by using fuzzy reasoning. Thus, the expanded range matrix and the segmented image are integrated into a 3-D one. The noise due to range uncertain estimations is eliminated by using the color.

. Cameras Stabilization Issues

One important problem in this project is the appropriate stabilization of the vision camera to acquire images in real-time. This a very significant feature required by the synthesis of images.

. Landmarks Detection, Recognition and Interrelation

. Generation & Representation of the Navigation Space

The determination of the free navigation space, Free Navigation Space (FNS), requires the calculation of some parameters, such as the orientation-angle, of the current position of Tyflos with respect to magnetic north, the current position, the previous position, the inclination-angle of Tyflos direction from the direction of the destination point [4]. At this point, the Tyflos is able to generate the space FNS by using the echo-distances from the surrounding obstacles. The Tyflos system measures the length of the echo-distances and connect their ends, generating the shape SH of the space FNS, as illustrated in an example bellow. The shape SH can be expressed as a relationship among straight line-segments and curve-segments respectively.

A hierarchical approach is also used to express and manipulate these relationships among the segments in an abstracted form. The determination of the navigation space, especially its shape, allows us to define the "open-navigation corridors", OC, and their widths, WC, for the final determination of the navigation path.

C. Where is a Particular object?

In this case the particular object is converted into a hierarchical SPN graph [14]. The SPN graph is compared to the system's KB in order the object to be recognized or appropriately classified. Thus, the recognizable object activates the appropriate speech patterns, from the Speech DB, to be played by the ear speakers.

E. Guide me in a Particular Place

. Collision Free Path Planning

The path planning strategy used for a collision free real-time navigation in a dynamic, or unknown environment is based on a human like approach (trial and error) [6]. In particular, the system attempts to select the First Best Choice (FBC) open navigation corridors.

. Detection of Other Moving Objects

The structural form of the shape SH of the space FNS plays a very significant role in the detection of other moving objects in the same navigation space. In particular, an object is moving in its own free navigation space with velocity VRa, which is defined by the distances of the "open navigation corridor and the time Dt, which Tyflos computing system needs to process the information received by its sensors. However, if more than one moving object exist in the same navigation space, then the velocity of each moving object is restricted by a new factor, "the motion (or trajectory) of other moving objects". By "motion" we mean the direction and the velocity of each object. At this point, a number of questions has arisen related to the existence of other moving objects in the same space. In particular:

. "How Tyflos detects other moving objects in the same navigation space"

For the detection of other moving objects in the same navigation space we combine two methodologies: the space shape changing methodology [4] and the differences detected from the comparison of two consecutive 2-D images. If both method agree that some motion has been detected in the same geographic region of the free space, then this motion (the moving object) is tracked by using a window around the motion. If however, there is a disagreement between these two methods, then probably a 2-D motion occurs in a 2-D plane (surface) of the space shape, like a motion on a movie screen, a projection of a motion on a wall, etc., which there is no real moving objects in the navigation space, or the Tyflos vision cameras select images through a closed window where the glass keeps the shape of the space unchanged.

In the shape changing method, the Tyflos system is able to calculate the echo-distances at the time t(i), and generate the shape SH. In addition, the Tyflos system is able to know the velocity (under regular conditions), which it has to travel for a specific interval of time Dt. It also knows the direction on which it is moving. Thus, it can predict and calculate the distance do(Dt) = VRa*Dt, and the location of its next projected point P'. Now, by using the new point P', Tyflos can generate a new "possible" predicted shape SH' of the free navigation space FNS', as Tyflos can "perceived" it from the point P. The generation of the new shape SH' is based on new echo-distances calculated from the point P', [4].

The Tyflos system saves the shape SH' and when it measures the real distance dk from the actual point P, it generates the real shape SH. At this point, Tyflos is able to regenerate another shape SH" similar or the same with the shape SH' on the space FNS, if possible. Now,

IF SH" = SH' THEN there is no other moving object in the same navigation space, with the assumption that the

velocity VR' < S / Dt, where S is the maximum distance between two different points on the shape SH.

ELSE there is at least one moving object in the same free navigation space.

The detection of "motion" in the free navigation space requires the use of more than two successive shapes of which at least one change must have happened.

In the image changing method, we select changes that occur on a selected view of the navigation space by comparing at least two consecutive images taken from that view.

. "What is the perceived size of the moving object"

When Tyflos, detects another moving object, Rb, in the same navigation area, then it attempts to define the size (dimensions) "perceived" from its current position P. Initially, it applies a matching process between the shapes SH' and SH" and if SH'= SH" then their difference SRb = O. The difference represents the shape of the object Rb perceived by Tyflos from the position P'. In addition, the rest of the perceived dimensions of the object Rb are provided by the difference of the echo-distances |dk'(t(i+1)) - dk(t(i+1))|. It is also obvious that the use of a more accurate estimate of the size of the object Rb requires the process of more than two consecutive shapes. In addition to the perceived shape, we use the corresponding window from the 2-D image (with colors and 2-D shapes) to identify the moving object and KB assists the classification of the object.

. "What is the direction and the velocity of the moving object"

When the location of the SRb shape has been detected, the robot Ra has to check that particular region for at least three consecutive shapes in order to define the current direction of the moving object Rb. In addition, the robot Ra, knowing at least two consecutive positions (P', P") for the Rb moving object, can calculate the velocity of the Rb robot at a specified time Dt.

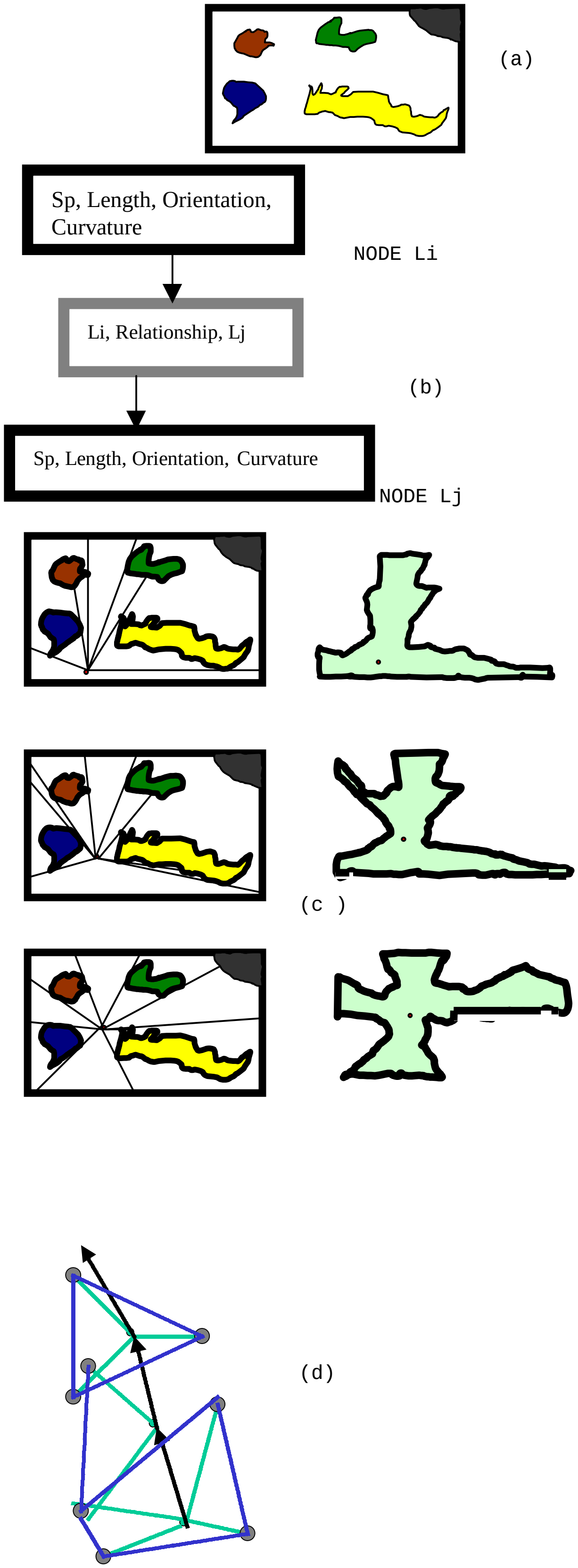
. "How many different moving objects are in the same free space"

When the Ra robot extracts and calculates the shape SH, then it is going to detect the "abnormalities" existed at the shape SH by using SH'. With the term "abnormality" we mean the difference SRb produced by the matching process M. If differences SRb discovered then the Ra robot processes the SRb shapes and attempts to separate them into smaller independent shapes SRm', if possible.

The number of different independent abnormalities SRk is the same with the number of different moving objects in the same free navigation space, with the condition that two different abnormalities don't have any connection to each other on the same shape SH. In case that these shape abnormalities overlap each other, the 2-D image corresponding to that area will provide additional information regarding color, 2-D shape, etc. to separate the moving objects if many.

G. Navigation with L-G Landmarks

After a successful or non-successful navigation effort, Tyflos saves the main characteristics from that effort. In particular, it saves, in the KB, the traveled navigation path in abstracted form of a L-G graph (see example below)and interrelates it with the consecutive free navigation spaces and their shapes. More specifically, it combines the shapes of the free navigation spaces into a global shape, and saves it in abstract form in the knowledge-base (KB). This abstracted form describes the shape in a generic way associated with the acquisition and synthesis which can regenerate the navigation space [6].



A navigation example: a) the FNS, b) the generation of relationships at the graph level, c) consecutive steps in FNS, d) the navigation path using L-G graph landmarks.

. Knowledge Base (KB)

One very important part for the successful and flexible generation of Space Maps is the development of a knowledge base able to store, process and access the Space Maps' forms. For this purpose, a number of operations (union, intersection, synthesis, subtraction, decomposition, save, search, match, insert, delete, extract, display, generate-map, find-path, update, analyze, etc.), [14].

3.4. Communication and Mobile DBs Issues

Today's technological limitations don't allow a portable computer to host many different databases, such as sports, travel, market, medical, etc. Thus, for the Tyflos system there are two possible alternatives, either the user to carry CDs with different DBs, or its computer to be connected with a multimedia network in order to download the appropriate DB at each particular time. The use of different databases is a necessity for the Tyflos system, since the user needs to access different forms of information according to the circumstances.

4. IMPLEMENTATION OF THE TYFLOS PROTOTYPE

4.1 . Integration of AI Methodologies

The actual implementation of the Tyflos system combines the state-of-the-art in AI methodologies, such as image understanding, speech understanding, natural language understanding, knowledge representation and conversion, etc., with an efficient interfacing of software tools (DBs, OS and Communication interfaces).

4.2. The Prototype

The hardware prototype of the Tyflos system has been designed and its implementation is in progress at AIIS. At this stage the Tyflos system consists of the pair of glasses, a Hitachi high resolution, color vision camera, a portable computer Mc book-note, an ear speaker, a voice synthesizer and a microphone.

The software tools available in the Tyflos system are: an image processing library with 150 algorithms (EIKONES languages), an activity recognition language (EIRMA) for scene description (simple cases), an image understanding tool (55% functional), a speech to natural language conversion tool ("Speaking Dynamically" functional 60%), a learning tool, an IBM speech recognition tool, a rule based expert system, and a small scale Knowledge Base.

4.3. Synchronization of Multiple Modalities

A very important issue for both systems is the synchronization and differentiation of different modalities (images, voices, sounds, etc.) occurred in the same "working" environment.

. Differentiation [4]

This is a very difficult task in particularly in the speech and sounds domains. Thus, the distinction of single modalities at the speech and sounds domains can be done by using VS patterns, so that the actual problem is transferred from the S&S domain into the image domain.

. Synchronization [4,21]

The synchronization issue can be handled efficiently by using the stochastic Petri-net (SPN) model. This means that when an image or images are associated with certain voices and sounds, then these modalities are connected via an SPN model, where their activation and acquisition can be easily performed.

5. CONCLUSIONS

The benefits of the system presented here are significant for the visually impaired people. It will provide the ability to the user to understand better the surrounding environment, and give more independence in his/her motion. It will not be as good as the one proposed in IEEE Spectrum [7,19,20], since the user will not gain back his/her lost senses (eyes), but on the other hand it is an possible, available alternative for an independent life until the medical science is able to repair the human visual system.

The status of the system is not 100% functional yet. Several important problems (vision, speech) have to be solved, before such a system to be available to the users. However, a significant progress has been done, and new methods in both vision and speech developed by other researchers will be appropriately used.

References

[1] N.G.Bourbakis and D.Kavraki, Intelligent assistants for handicapped people's indepedence: case study, IEEE Symposium on Intelligence and Systems, pp. 337-344, Nov. 1996, MD

[2] The NC eye-chip project, 1999

[3] N.Bourbakis, 2-D space maps generation in unknown environments, IJPRAI,vol.11,1997

[4] N.Bourbakis and D.Kavraki, Visual representation and recognition of speech patterns, TR-1995

[5] N.Bourbakis and R.Andel, Fusing laser and image data for 3-D representations, IEEE Symp. on ATP, 1997

[6] D.Kavraki and N.Bourbakis, Improving spoken word sounds by matching them with spoken NL patterns, TR-1993, 11 pages

[7] G.Dagnelie and R.Massof, Toward an artificial eye, IEEE Spectrum,May 1996, pp.20-29

[8] ELMO microcameras, Image processing past and present, 1995

[9] M.Flickner et.al., Query by images and video content: IBM-QBIC system, IEEE Computer,28,9,23-32,1995

[10] The EUROTRA project: Machine languages translation, ESPRIT, 1986-96

[11] STARS , Newsletter, 12, March 1995

[12] Integration testbeds for mobility: the unmanned autonomous vehicle, NIST, Intelligent Systems Div., 1996

[13] M.Raibert, Legged robots that balance, MIT Press Cambridge MA 1986

[14] N.Bourbakis Generation of 3-D space maps and landmark navigation from unknown environments using SPNGs, IEEE T-SMC

[15] N.Bourbakis, A survey on 3-D navigation of autonomous mobile robots, TR-1996

[16]N.Bourbakis, S.Mertoguno, Kydon-an autonomous, multi-layers vision system, IJEAAI, 9,1,43-52,1996

[17] C.Koch & B.Mathur, Neuromorphic Vision Chips,IEEE Spectrum,May 1996,pp.38-46

[18] The Nanorobots Research Project in Japan, 1990

[19] J.Wyatt and J.Rizzo, Ocular implants for the blind,IEEE Spectrum,May 1996,pp.47-53

[20] R.Normann, E.Maynard, K.Guillory and D.Warren, Cortical implants for the blind, IEEE Spectrum, May 1996, pp. 54-59

[21] N.Bourbakis, SPN representation on spoken natural language, AIIS TR-1997-05

[22] Communication with J.Albus, 3-D Images via stereo audio signals, NIST,July 25,1996