

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

---

2009

# Tyflos: A Wearable Navigation Prototype for Blind & Visually Impaired; Design, Modelling and Experimental Results

Dimitrios Dakopoulos  
*Wright State University*

Follow this and additional works at: [https://corescholar.libraries.wright.edu/etd\\_all](https://corescholar.libraries.wright.edu/etd_all)



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

---

## Repository Citation

Dakopoulos, Dimitrios, "Tyflos: A Wearable Navigation Prototype for Blind & Visually Impaired; Design, Modelling and Experimental Results" (2009). *Browse all Theses and Dissertations*. 288.  
[https://corescholar.libraries.wright.edu/etd\\_all/288](https://corescholar.libraries.wright.edu/etd_all/288)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact [corescholar@www.libraries.wright.edu](mailto:corescholar@www.libraries.wright.edu), [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).

TYFLOS: A WEARABLE NAVIGATION PROROTYPE  
FOR BLIND & VISUALLY IMPAIRED;  
DESIGN, MODELLING AND EXPERIMENTAL RESULTS

A dissertation submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

By

DIMITRIOS DAKOPOULOS  
B.S., National and Kapodistrian University of Athens, 2005

---

2009  
Wright State University

COPYRIGHT BY  
DIMITRIOS DAKOPOULOS  
2009

WRIGHT STATE UNIVERSITY  
SCHOOL OF GRADUATES STUDIES

May 12, 2009

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION  
BY Dimitrios Dakopoulos, ENTITLED Tyflos: a Wearable Navigation Prototype for Blind & Vi-  
sually Impaired: Design, Modeling and Experimental Results BE ACCEPTED IN PARTIAL FUL-  
FILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

---

Nikolaos G. Bourbakis, Ph.D.  
Dissertation Director

---

Arthur A. Goshtasby, Ph.D.  
Director, Computer Science & Engineering  
Ph.D. Program

---

Committee on Final Examination

---

Nikolaos G. Bourbakis, Ph.D.

---

Mateen M. Rizki, Ph.D.

---

John C. Gallagher, Ph.D.

---

David B. Reynolds, Ph.D.

---

Partha P. Banerjee, Ph.D.

## ABSTRACT

Dakopoulos, Dimitrios, Ph.D., Department of Computer Science & Engineering, Wright State University, 2009.

*Tyflos: a Wearable Navigation Prototype for Blind & Visually Impaired: Design, Modeling and Experimental Results.*

The need for assistive devices has, had and will have a large merit in many engineering research arenas. This dissertation deals with the design, modeling, implementation and experimentation of the navigation component of a wearable assistive system for blind and visually impaired people, called Tyflos (Τυφλός) which is the Greek word for “Blind”.

The current prototype consists of two mini cameras attached to a pair of conventional eye-glasses, a 2D tactile display (vibration array) which consists of 16 vibrating elements arranged in a  $4 \times 4$  manner, attached to an elastic vest worn on the user's abdomen, a portable computer, an ear-speaker and a microphone.

The Tyflos Navigator is an Electronic Travel Aid (ETA) with primary goal to help users towards their independent mobility in indoor environment. Its main sensor unit, the stereo vision system, captures environmental information from the user's field-of-view. 3D representations are created and moving objects are identified using stereoscopic vision and motion detection methodologies. The high resolution output of the methodologies is projected on the low resolution vibration array via a high-to-low methodology based on navigation criteria and modeled with a formal language called Vibration Array Language (VAL). The spatial distribution and temporal characteristics (varying frequencies) of the vibrating elements of the vibration array can inform the user for safe navigation paths and obstacles, giving distance and location information.

All parts of the system will be continuously adapted until the users' needs are fulfilled

or the technological constraints are reached. A step towards that goal will be shown at the last part of this work with the development of a tactile vocabulary and the experimentation with users where they provide feedback giving us directions for refinements, changes and future work.

## TABLE OF CONTENTS

<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
1.1 THE NEED FOR ASSISTIVE DEVICES FOR BLIND AND VISUALLY IMPAIRED.....	1
1.2 TYFLOS NAVIGATION SYSTEM OVERVIEW AND HISTORY .....	1
1.3 DISSERTATION GOALS AND OUTLINE .....	3
<b>CHAPTER 2: A SURVEY ON ELECTRONIC TRAVEL AIDS FOR BLIND .....</b>	<b>5</b>
2.1 ELECTRONIC TRAVEL AIDS.....	7
2.2 Maturity Analysis .....	31
2.3 DISCUSSION ON THE SURVEY .....	34
<b>CHAPTER 3: SYSTEM'S ARCHITECTURE &amp; OPERATION .....</b>	<b>37</b>
3.1 HARDWARE AND SOFTWARE ARCHITECTURE.....	37
3.2 OPERATION MODELING.....	39
<b>CHAPTER 4: IMPLEMENTATION .....</b>	<b>51</b>
4.1 THE CANE PARADIGM .....	51
4.2 HUMAN FACTORS.....	52
4.3 THE STEREO VISION SYSTEM .....	56
4.4 THE 2D VIBRATION ARRAY.....	60
4.5 INTERFACE.....	64
4.6 OPERATION .....	65
<b>CHAPTER 5: VIBRATION ARRAY LANGUAGE .....</b>	<b>69</b>
5.1 FORMAL MODEL.....	69
5.2 SIMULATION.....	71
<b>CHAPTER 6: VIDEO STABILIZATION &amp; MOTION DETECTION .....</b>	<b>75</b>
6.1 THE METHODOLOGY .....	76
6.2 SIMPLE EXPERIMENTAL RESULTS .....	86
<b>CHAPTER 7: HIGH-TO-LOW RESOLUTION REPRESENTATIONS.....</b>	<b>88</b>
7.1 THE RESOLUTION OF THE VIBRATION ARRAY .....	88
7.2 NAVIGATION ISSUES .....	88
7.3 STANDARD HIGH-TO-LOW ALGORITHMS .....	91
7.4 CRITERIA OF INTEREST .....	92
7.5 VISUAL REPRESENTATION .....	93
7.6 3D REPRESENTATION .....	97
7.7 INFORMATION MODELING.....	104
<b>CHAPTER 8: TACTILE VOCABULARY AND TESTING.....</b>	<b>114</b>
8.1 EXPERIMENTAL DESIGN.....	115
8.2 TESTING ON SIGHTED USERS.....	116
8.3 PATTERN MATCHING .....	126
8.4 TESTING NAVIGATION SCENARIOS.....	131
8.5 TESTING ON VISUALLY IMPAIRED USERS .....	133
8.6 EVALUATION .....	137
<b>CHAPTER 9: RESEARCH CONTRIBUTIONS &amp; NEW EXPERIMENTAL DESIGN.</b>	<b>140</b>
9.1 THE NEW SYSTEM/PROTOTYPE (PUB. 1, 2) .....	140
9.2 ANALYSIS OF THE NEW RESEARCH CONTRIBUTIONS .....	142

9.3 A NEW EXPERIMENTAL DESIGN.....	146
<b>REFERENCES.....</b>	<b>156</b>

## LIST OF FIGURES

FIG. 1-1. OUTLINE OF DISSERTATION.....	3
FIG. 2-1. THE REVIEWED ETAs WITH CATEGORIZATION BASED ON HOW THE INFORMATION IS SENT TO THE USER (THE LATEST PROTOTYPE'S YEAR IS SHOWN IN PARENTHESIS).....	7
FIG. 2-2. THE NAVBELT'S OPERATION: THE 8 ULTRASONIC SENSORS CREATE AN ANGLE MAP WHICH REPRESENTS THE DISTANCE OF OBSTACLES IN THOSE ANGLES.....	9
FIG. 2-3. AN IMPLEMENTATION OF THE VOICE – “SEEING WITH SOUND” SYSTEM (GLASSES WITH ATTACHED CAMERA, EAR SPEAKERS AND PORTABLE COMPUTER).....	10
FIG. 2-4. THE SENSOR OF THE PROJECT FROM UNIVERSITY OF STUTTGART.....	11
FIG. 2-5. THE FIU PROJECT PROTOTYPE (LEFT) AND AN EXAMPLE OF ITS OPERATION (RIGHT). .....	12
FIG. 2-6. VIRTUAL ACOUSTIC SPACE PROTOTYPE (CAMERAS AND HEADPHONES MOUNTED ON EYEGLASSES AND MICROPROCESSOR (RIGHT)).....	13
FIG. 2-7. THE NAVI AND ITS COMPONENTS.....	14
FIG. 2-8. PROTOTYPE FROM UNIVERSITY OF GUELPH AND THE HAND SPATIAL CORRESPONDENCE.....	15
FIG. 2-9. SCHEMATIC (LEFT) AND OPERATION (RIGHT) OF THE GUIDECAKE PROTOTYPE.....	17
FIG. 2-10. THE ENVS AND ITS COMPONENTS.....	18
FIG. 2-11. THE PROTOTYPE CYARM AND THE CONCEPT OF OPERATION.....	19
FIG. 2-12. DESCRIPTION OF VIBROTACTILE ACTUATORS POSITIONS ON THE TACTILE HANDLE WITH THREE NAVIGATION SCENARIOS.....	21
FIG. 2-13. THE TVS PROTOTYPE.....	22
FIG. 2-14. AN EXAMPLE OF TVS OPERATION: IMAGE FROM THE TWO CAMERAS, DISPARITY MAP AND THE CORRESPONDING SIGNALS SENT TO THE TACTOR BELT.....	22
FIG. 2-15. HARDWARE DETAILS OF THE EPFL PROTOTYPE.....	24
FIG. 2-16. OPERATION AND HIGH-LEVEL DESIGN OF THE EPFL PROTOTYPE.....	24
FIG. 2-17. USER WEARING THE 2 <sup>ND</sup> TYFLOS' PROTOTYPE. A) STEREO CAMERAS ATTACHED ON DARK EYEGLASSES AND VIBRATION ARRAY VEST ON THE USER'S ABDOMEN, B) PORTABLE COMPUTER, C) MICROCONTROLLER AND PCBs, D) ARRANGEMENT OF THE 4×4 VIBRATING ELEMENTS INSIDE THE VIBRATION ARRAY VEST.....	26
FIG. 2-18. OPERATION OF TYFLOS WITH TWO NAVIGATION SCENARIOS (ONE IN EACH ROW): LEFT COLUMN SHOWS THE IMAGES CAPTURED BY THE CAMERAS. THE MIDDLE THE DEPTH MAP AND THE RIGHT WHAT THE USERS SENSES VIA THE 4×4 VIBRATION ARRAY (LIGHT BLUE = NO VIBRATION = OBSTACLE FURTHER THAN 4M, YELLOW = VIBRATION LEVEL 1 = OBSTACLE IN RANGE [2,4]M AND RED = VIBRATION LEVEL 2 = OBSTACLE IN RANGE [1,2]M ). IN THE FIRST SCENARIO THE USERS FEELS THERE IS AN OPEN PATH (DOOR). IN THE SECOND THE USER FEELS THAT THERE IS NO OPEN PATH.....	27
FIG. 2-19. THE UCSC's HANDHELD DEVICE EQUIPPED WITH LASER RANGE SENSOR.....	29
FIG. 2-20. THE TIME PROFILE OF TWO STEPS ACQUIRED AS THE DEVICE WAS PIVOTED IN AN UPWARD MOTION.....	29
FIG. 2-21. COMMERCIAL PRODUCTS: A) K-SONAR CANE, B) MINI-RADAR, C) MINI-GUIDE AND D) LASERCANE.....	30
FIG. 2-22. MATURITY RANKING THAT SHOWS THE TOTAL SCORE FOR EACH SYSTEM.....	34
FIG. 3-1. HARDWARE ARCHITECTURE OF THE TYFLOS 2 <sup>ND</sup> PROTOTYPE, EMPHASIZING ON ITS MULTIMODAL CHARACTERISTICS. THE GREEN BOXES ARE INPUT MEDIA (STEREO CAMERAS: CAPTURE VISUAL INFORMATION, MICROPHONE: SPEECH INFORMATION FROM THE USER) AND THE YELLOW BOXES ARE OUTPUT MEDIA (EAR-SPEAKER: SPEECH MESSAGES TO THE USER, VIBRATION ARRAY VEST: TACTILE INFORMATION TO THE USER).....	38
FIG. 3-2. HARDWARE AND SOFTWARE ARCHITECTURE OF THE 2 <sup>ND</sup> TYFLOS PROTOTYPE.....	38
FIG. 3-3. EXTENDED ARCHITECTURE OF THE 2 <sup>ND</sup> TYFLOS PROTOTYPE, INCLUDING THE OCR- TYPE READER MODULE, RANGE SENSOR VALIDATION AND PLD CAPABILITIES USING GPS	

SENSORS.....	39
FIG. 3-4. TYFLOS NAVIGATOR GRAMMAR AS NETWORK.....	43
FIG. 3-5. LEVEL 1 SPN MODEL (USER-DEVICE).....	45
FIG. 3-6. LEVEL-2 SPN MODEL (DEVICE).....	45
FIG. 3-7. LEVEL-3 SPN MODEL (MODULES INTERACTION).....	46
FIG. 3-8. LEVEL-4 SPN MODEL (NAVIGATOR) (NOTE: THE STATES WITHOUT LABEL CORRESPOND TO PSEUDO-STATES). .....	47
FIG. 3-9. NAVIGATION SCENARIO 1: A,B) LEFT AND RIGHT CAMERA 384×288 IMAGE (AFTER RECTIFICATION), C) DISPARITIES MAP, D) DISTANCES MAP WITH SELECTED SQUARE AREA OF 256×256. ....	49
FIG. 3-10. NAVIGATION SCENARIO 1: FINAL 4×4 DISTANCES MAP, PROJECTED ON USER VIA THE VIBRATION ARRAY (THE BLUE GRID IS USED FOR VISUALIZE BETTER THE 4×4 VIBRATING CELLS OF THE ARRAY) A) CLASSIC PYRAMID, B) REVISED PYRAMID WITH NAVIGATIONAL CRITERIA. ....	49
FIG. 3-11. NAVIGATION SCENARIO 2: A) FUSED IMAGE FROM CAMERAS, WHILE THE SELECTED RED AREA CORRESPOND TO MOTION DETECTION AND THE GREEN SQUARE TO DETECTED PERSON, B) LOW RESOLUTION 4×4 DISTANCES MAP PROJECTED ON THE VIBRATION VEST. .....	50
FIG. 4-1. A) BRAILLE STAR 80, A BRAILLE KEYBOARD, B) SYNCBRAILLE: A BRAILLE DISPLAY ..	54
FIG. 4-2. A) THE CREATIVE LIVE! CAM VIDEO IM PRO WEB-CAMERA BY CREATIVE INC.[58] AND B) THE TWO WEB-CAMERAS WITHOUT THE PACKAGING, ATTACHED ON THE ALUMINUM PLATE WHICH IS SCREWED ON CONVENTIONAL EYE GLASSES. ....	57
FIG. 4-3. EXPERIMENTAL GRAPH FOR DISPARITY-DISTANCE CORRESPONDENCE. ....	59
FIG. 4-4. THE JINLONG MACHINERY C1030B028F COIN-TYPE VIBRATOR (1CM IN DIAMETER AND 3MM THICK) .....	60
FIG. 4-5. THE 555 TIMER IN UNSTABLE OPERATION. ....	61
FIG. 4-6. THE VIBRATION MODULE SCHEMATIC: THE DIGITAL POTENTIOMETER DS1803 CONTROLS TWO LM555CN TIMERS THAT CAN DRIVE ONE VIBRATOR EACH. ....	62
FIG. 4-7. THE VIBRATION ARRAY SCHEMATIC: THROUGH THE 2-WIRE INTERFACE, 8 VIBRATION MODULES CAN BE CONTROLLED WHICH CAN DRIVE UP TO 16 VIBRATORS.....	62
FIG. 4-8. A) THE PROTOTYPE VIBRATION ARRAY BOARDS (GREEN BOARDS) CONNECTED WITH THE 8051 MICROCONTROLLER DEVELOPMENT BOARD (TOP) AND B) THE INSIDE OF THE VIBRATION ARRAY VEST WITH THE 16 VIBRATORS ARRANGED IN A 4×4 MANNER.....	63
FIG. 4-9. THE HIGH-LEVEL DESIGN OF THE TYFLOS' VIBRATION ARRAY.....	64
FIG. 4-10. EXPERIMENTAL FREQUENCIES OF THE SQUARE PULSES PRODUCED BY THE TIMERS AND DRIVE THE VIBRATORS.....	66
FIG. 4-11. EXPERIMENTAL DUTY CYCLES OF THE SQUARE PULSES PRODUCED BY THE TIMERS AND DRIVE THE VIBRATORS.....	66
FIG. 4-12. TYPICAL CALIBRATION CURVE (HEX VALUE – OUTPUT FREQUENCY) FOR A VIBRATION MODULE. ....	68
FIG. 5-1. THE SEVEN SIMULATED SCENARIOS ILLUSTRATING THE USE OF THE VIBRATION ARRAY LANGUAGE (VAL). .....	74
FIG. 6-1. FIRST ROW: TWO CONSECUTIVE FRAMES IN THE 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.....	77
FIG. 6-2. THE ESTIMATED DENSITY OF MOTION ACTIVITY MAP USING PARZEN KERNELS.....	84
FIG. 6-3. THREE REAL INDOOR NAVIGATION SCENARIOS. FOR EVERY SCENARIO A, B AND C ARE THE 3 CONSECUTIVE FRAMES USED BY THE METHODOLOGY. IN D, WHITE AREAS ARE THE DETECTED PIXELS THAT CORRESPOND TO MOVING AREA. ....	87
FIG. 7-1. PYRAMIDAL REDUCTION USING A 2×2 KERNEL. EVERY 2×2 PIXELS KERNEL IS REDUCED TO 1 PIXEL AS SHOWN. .....	91

FIG. 7-2. FIGURE 7-2. A) FUSED 256×256 IMAGE, PRESENTING A LOW HEIGHT CLOSE DISTANCE OBSTACLE ON THE LEFT AND A WALKING PERSON IN MEDIUM DISTANCE ON THE RIGHT AND B) IMAGE-OF-INTERESTS REPRESENTING 2 AREAS OF INTEREST WITH WHITE COLOR (OBSTACLE AND PERSON).....	93
FIG. 7-3. EXAMPLE OF CLASSICAL AND MODIFIED PYRAMIDAL REDUCTIONS FROM AN 8×8 TO A 4×4 RESOLUTION FOR VISUAL REPRESENTATION. FOUR PIXELS CORRESPOND TO ONE PIXEL IN THE LOWER PYRAMIDAL LEVEL (SHOWN WITH THE SAME BORDER COLOR); NOTICE THAT THE UPPER LEFT OBJECT OF INTEREST (WHITE) IS PRESERVED WHEN USING THE MODIFIED PYRAMID.....	95
FIG. 7-4. THE LOW RESOLUTION RESULTS FROM THE IMAGES USING (A) LINEAR INTERPOLATION, (B) NEAREST NEIGHBOR INTERPOLATION, (C) GAUSSIAN PYRAMID, (D) PYRAMIDAL REDUCTION 2×2, (E) PYRAMIDAL REDUCTION 3×3.....	96
FIG. 7-5. RESULTS FOR IMAGES USING CRITERIA OF TYPE C2/C4. (A) PYRAMIDAL REDUCTION 2×2 (B) PYRAMIDAL REDUCTION 3×3.....	97
FIG. 7-6. A) FUSED IMAGE AND B) IMAGE OF DISTANCES.....	98
FIG. 7-7. 3D LOW RESOLUTION RESULTS FOR IMAGES OF DISTANCES: (A) LINEAR INTERPOLATION, (B) NEAREST NEIGHBOR INTERPOLATION, (C) GAUSSIAN PYRAMID, (D) PYRAMIDAL REDUCTION 2×2, (E) PYRAMIDAL REDUCTION 3×3.....	98
FIG. 7-8. HORIZONTAL NAVIGATION SPACE REPRESENTATION (C IS THE USER/CAMERA), A) DIFFERENT DISTANCE RANGES (0,1.5M), (1.5M TO 2.5M), (2.5M TO 4M) AND (OVER 4.0M) WITH DIRECT CORRESPONDENCE TO TABLE 3-2, B) DETAIL OF THE GEOMETRY FOR CALCULATING THE CRITICAL PATH WIDTH.....	99
FIG. 7-9. CROP FACTOR OF THE CAPTURED IMAGES. THE BLACK FRAME IS THE CAMERAS' 4:3 AND THE RED IS THE SQUARE 1:1 NEEDED FOR THE PYRAMIDAL REDUCTIONS. THE RESULT IS A SMALLER HORIZONTAL VIEW ANGLE (THE VERTICAL VIEW ANGLE IS NOT AFFECTED SIGNIFICANTLY).....	101
FIG. 7-10. VERTICAL NAVIGATION SPACE REPRESENTATION.....	102
FIG. 7-11. EXAMPLE OF CLASSICAL AND MODIFIED PYRAMIDAL REDUCTIONS FROM AN 8×8 TO A 4×4 RESOLUTION FOR 3D REPRESENTATION. FOUR PIXELS CORRESPOND TO ONE PIXEL IN THE LOWER PYRAMIDAL LEVEL (SHOWN WITH THE SAME BORDER COLOR); NOTICE THAT THE CLASSICAL PYRAMID DOESN'T KEEP IMPORTANT INFORMATION FOR PROXIMITY OBJECTS (MORE WHITE) OR OPEN PATHS (BLACK).....	103
FIG. 7-12. 3D LOW RESOLUTION RESULTS: A) CRITERION 1, PYRAMIDAL REDUCTION 2×2, B) CRITERION 1, PYRAMIDAL REDUCTION 3×3, C) CRITERION 2, PYRAMIDAL REDUCTION 2×2, D) CRITERION 2, PYRAMIDAL REDUCTION 3×3.....	104
FIG. 7-13. THE 4 TEST NAVIGATION SCENARIOS: FUSED-IMAGE, IMAGE-OF-DISTANCES (DARKER GRAY-LEVEL CORRESPONDS TO LARGER DISTANCE FROM THE USER) AND IMAGE-OF-INTERESTS (WHITE PIXELS BELONG TO REGIONS OF INTEREST).....	106
FIG. 7-14. 3D REPRESENTATION AND PYRAMIDAL LEVELS FOR SCENARIO 0: A) STANDARD PYRAMID, B) MODIFIED PYRAMID WITH CRITERION 1 AND C) MODIFIED PYRAMID WITH CRITERION 2.....	107
FIG. 7-15. SCENARIO 2 FOR DIFFERENT RESOLUTIONS FOR THE 2 METHODOLOGIES (A) AND (B) FOR THE FUSED AND THE IMAGE-OF-INTERESTS. WE NOTICE THAT THE MODIFIED PYRAMID (B) EMPHASIZES MORE ON THE REGIONS OF INTEREST (WHITE).....	108
FIG. 7-16. TEXTURAL ENTROPY FOR SCENARIO 1 AT 45° ORIENTATION (LEFT TO RIGHT: FUSED-IMAGE, IMAGE-OF-DISTANCES, IMAGE-OF-INTERESTS). THE BLUE LINES REPRESENT THE CLASSICAL APPROACH (A) AND (C) AND THE RED AND GREEN THE MODIFIED METHODOLOGIES (B), (D) AND (E).....	111
FIG. 7-17. TEXTURAL CORRELATION FOR SCENARIO 0 AT 90° ORIENTATION (LEFT TO RIGHT: FUSED-IMAGE, IMAGE-OF-DISTANCES, IMAGE-OF-INTERESTS).....	112
FIG. 7-18. ANGULAR SECOND MOMENT FOR SCENARIO 3 AT 135° ORIENTATION (LEFT TO RIGHT: FUSED-IMAGE, IMAGE-OF-DISTANCES, IMAGE-OF-INTERESTS).....	113
FIG. 8-1. VIBRATION LEVELS/FREQUENCIES IDENTIFICATION ACCURACY (SIGHTED USERS)...117	
FIG. 8-2. EXPERIMENT #7: USER'S ACCURACY IN IDENTIFYING THE CORRECT POSITION OF A	

RANDOM SYMBOL (SIGHTED USERS) .....	120
FIG. 8-3. EXPERIMENT #7 (SIGHTED USERS): USERS' ACCURACY IN IDENTIFYING THE DIFFERENT SYMBOLS IN RANDOM POSITIONS.....	121
FIG. 8-4. PATTERN MATCHING METHODOLOGY FLOW.....	129
FIG. 8-5. EXAMPLES OF SELECTED FRAMES. EVERY FRAME IS CONSISTING OF 4 IMAGES: 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 (4x4; 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).....	130
FIG. 8-6. FREQUENCIES/VIBRATION LEVELS IDENTIFICATION ACCURACY (VISUALLY IMPAIRED USERS), SIMILAR TO FIG. 8-1.....	134
FIG. 8-7. POSITION IDENTIFICATION ACCURACY FOR RANDOM SYMBOL AND POSITION (VISUALLY IMPAIRED USERS), SIMILAR TO FIG. 8-2.....	135
FIG. 8-8. SYMBOL IDENTIFICATION ACCURACY IN RANDOM POSITION (VISUALLY IMPAIRED USERS), SIMILAR TO FIG. 8-3 (SHOWN ON THE RIGHT). .....	136

## LIST OF TABLES

TABLE 2-1. STRUCTURAL AND OPERATIONAL FEATURES. FEATURES F1 TO F7 CORRESPOND TO USER'S NEEDS WHILE F8 TO F14 REFLECT THE DEVELOPER'S AND ENGINEER'S VIEWS. ....	32
TABLE 2-2. BINARY TABLE FOR CALCULATION OF WEIGHTS OF THE FEATURES. ....	33
TABLE 2-3. TABLE OF SCORES FOR ALL SYSTEMS AND FEATURES; A-G: AUDIO FEEDBACK, H-O: TACTILE FEEDBACK, P-Q: NO INTERFACE.....	33
TABLE 3-1. USER-SYSTEM COMMANDS .....	43
TABLE 3-2. VIBRATION LEVEL, VIBRATION FREQUENCY, DISTANCE AND COLOR REPRESENTATION CORRESPONDENCE. ....	50
TABLE 4-1. CALIBRATION TABLE FOR THE 16 POTENTIOMETERS (8 VIBRATION MODULES). E.G. TO ACHIEVE FREQUENCY OF 3HZ IN VIBRATING MOTOR #2 I NEED TO SEND 0X76 AT THE POTENTIOMETER #2.....	67
TABLE 5-1. SIMILAR WITH TABLE 3-2: VIBRATION LEVELS, VIBRATION FREQUENCIES AND OBJECT DISTANCES CORRESPONDENCE. ....	71
TABLE 5-2. COLOR REPRESENTATIONS OF THE 4 VIBRATION LEVELS. ....	72
TABLE 8-1. THE 6 VIBRATION SYMBOLS (ACTIVATED VIBRATING MOTORS ARE FILLED WITH BLACK). ....	118
TABLE 8-2. THE FOUR DIFFERENT POSITIONS THAT THE SYMBOLS CAN APPEAR ON THE 4x4 VA. ....	118
TABLE 8-3. EXPERIMENT #1 TO #5 FOR POSITION IDENTIFICATION ACCURACY FOR THE DIFFERENT SYMBOLS (B, C, D, E, F) .....	119
TABLE 8-4. EXPERIMENT #6: SYMBOL IDENTIFICATION ACCURACY .....	120
TABLE 8-5. EXPERIMENT #6 (SIGHTED USERS): AVERAGE SYMBOLS' CONFUSION MATRIX [%](SIGHTED USERS).....	122
TABLE 8-6. EXPERIMENT #7 (SIGHTED USERS): AVERAGE SYMBOLS' CONFUSION MATRIX [%]. .....	122
TABLE 8-7. HORIZONTAL RULES EXPERIMENTAL PATTERNS. AFTER APPLICATION OF THE RULES, WHITE MOTORS WITH X ARE ENABLED AND BLACK MOTORS WITH X ARE DISABLED.....	125
TABLE 8-8. EXAMPLE OF SYMBOL MATCHING USING THE STANDARD AND MODIFIED DISSIMILARITIES. THE BLACK CIRCLES CORRESPOND TO ACTIVE PIXELS I.E. CORRESPONDING MOTORS VIBRATE.....	128
TABLE 8-9. CORRECT IDENTIFICATION OF: PATTERNS WITHOUT VIBRATIONS; OPEN PATHS WITHIN PATTERNS; OPEN AREAS; OBSTACLES; DISTANCE OF OBSTACLES; NON-ZERO FREQUENCIES (SIGHTED USERS).....	131
TABLE 8-10. NAVIGATION SCENARIOS: AVERAGE CONFUSION MATRIX FOR THE DIFFERENT FREQUENCIES (SIGHTED USERS).....	132
TABLE 8-11. AVERAGE CONFUSION MATRIX FOR SYMBOL IDENTIFICATION (VISUALLY IMPAIRED), SIMILAR TO TABLE 8-6 (SHOWN ON THE RIGHT). .....	136
TABLE 8-12. CORRECT IDENTIFICATION OF: PATTERNS WITHOUT VIBRATIONS; OPEN PATHS WITHIN PATTERNS; NON-ZERO FREQUENCIES (VISUALLY IMPAIRED USERS), SIMILAR TO 0. ....	137
TABLE 8-13. NAVIGATION SCENARIOS: AVERAGE CONFUSION MATRIX FOR THE DIFFERENT FREQUENCIES (VISUALLY IMPAIRED USERS), SIMILAR TO TABLE 8-10 (SHOWN ON THE RIGHT).....	137
TABLE 9-1. EXPERIMENT 1: LAYOUT .....	151
TABLE 9-2. EXPERIMENT 2: LAYOUT .....	152
TABLE 9-3. EXPERIMENT 3: LAYOUT .....	154
TABLE 9-4. EXPERIMENT 4: LAYOUT .....	154

## **ACKNOWLEDGEMENT**

During all my time in Dayton and at Wright State University I was fortunate to meet some wonderful people who supported me in many different ways towards the completion of this dissertation.

The first and most important person that I would like to sincerely thank is my advisor Prof. Nikolaos Bourbakis for his continuous support, advising and trust. During all those years he gave me knowledge but the most important, a way of facing and dealing with challenges in academic and everyday life.

I would also like to thank all members of my dissertation committee: Dr. John Gallagher, Dr. Mateen Rizki, Dr. David Reynolds and Dr. Partha Banerjee for their fruitful comments and discussions we had, a necessary guidance that made this dissertation possible.

I would like thank Donetta Bantle and Leona Miller at ITRI and ATRC for their help wherever I needed and all my friends who participated in the experimental phase of this dissertation. I would never forget to deeply thank all my current and former colleagues: Raghu Kannavara, Alex Karargyris, Rob Keefer, Alekos Pantelopoulos, Allan Rwabutaza, Sanjay Boddhu, Dr. Ming Yang and Dr. Praveen Kakumanu.

Last but not least I would like to thank my family to whom I dedicate this dissertation, Prof. Gust and Mrs. Elli Bambakidis, Mrs. Despina Bourbaki and all my new friends that I met here in Dayton, as well as all my old friends from Greece.

*Στην οικογένειά μου,  
για όλη τους την αγάπη και υποστήριξη.*

# **Chapter 1: Introduction**

## **1.1 The Need for Assistive Devices for Blind and Visually Impaired**

According to NFB (National Federation for the Blind) [1] and AFB (American Foundation for the Blind) [2], the estimated number of legally blind people in the U.S. is 1.3 million and the total number of blind and visually impaired is approximately 10 million with around 100,000 to be students. Worldwide more than 161 million people are visually impaired with 37 million to be blind [3]. The need for assistive devices was and will be constant.

Devices can assist navigation (obstacle detection, positioning), reading as well as other every day activities; we will deal only with the former. White cane and guide dogs are the most popular navigation aids. Guide dogs are efficient but expensive. White cane is simple to use, inexpensive, extremely reliable and thus the most popular among the visual impaired. On the other hand, none provides all the necessary information normally perceived with vision [4]. Scientists and engineers strive to provide solutions for the limitations of those two popular navigation aids by developing a wide range of navigation systems and tools [5], [6].

## **1.2 Tyflos Navigation System Overview and History**

The main role of the Tyflos Navigator mobility assistant is to capture the environmental data from various sensors and map the extracted and

processed content onto available user interfaces in the most appropriate manner. It is a secondary mobility aid which means is complementary to the user's primary mobility aid (usually the white cane) so its purpose is to assist the user in situations that the primary mobility aid is limited.

As a final goal the Tyflos prototype will integrate a wireless handheld computer, cameras, range sensors, GPS sensors, microphones, natural language processor, text-to-speech device, and a digital audio recorder. The audio-visual input devices and the audio output devices can be worn (or carried) by the user. Data collected by the sensors is processed by the Tyflos' modules; each specialized in one or more tasks. In particular, it interfaces with external sensors (such as GPS, if applicable, range sensors, etc.) as well as the user, facilitating focused and personalized content delivery. The user communicates the task of interest to the mobility assistant using a speech-recognition interface.

The preliminary design and development of the Tyflos prototype has already been carried out. The first prototype consists of two cameras, a range scanner, an ear speaker, a microphone, a speech synthesizer, a two-dimensional tactile display and a portable computer. This device has been evaluated by students with visual impairments and their feedback has been used in the design requirements. The mobility prototype is based on the integration of several software components that reflect to some of the methodologies presented here. This dissertation deals with the design, modeling and experimentation of the second prototype.

### 1.3 Dissertation goals and outline

The goal of this dissertation is to develop a complete working prototype and not the optimization of its components. Fig. 1-1 shows the steps followed towards that goal. The first step is a careful study of the literature. The next step is the operation modeling using stochastic Petri-Nets and a hidden Markov model speech recognition tool. The hardware implementation includes the stereo vision system, the PCBs and the microcontroller interface with the computer and finally the fabrication the 2D vibration array vest. Next, a formal tactile language called Vibration Array Language (VAL) models the information displayed on the vibration array and represents robustly and in a compact mathematical form, every possible combination of environmental scenes that can appear in the navigation space of the user.

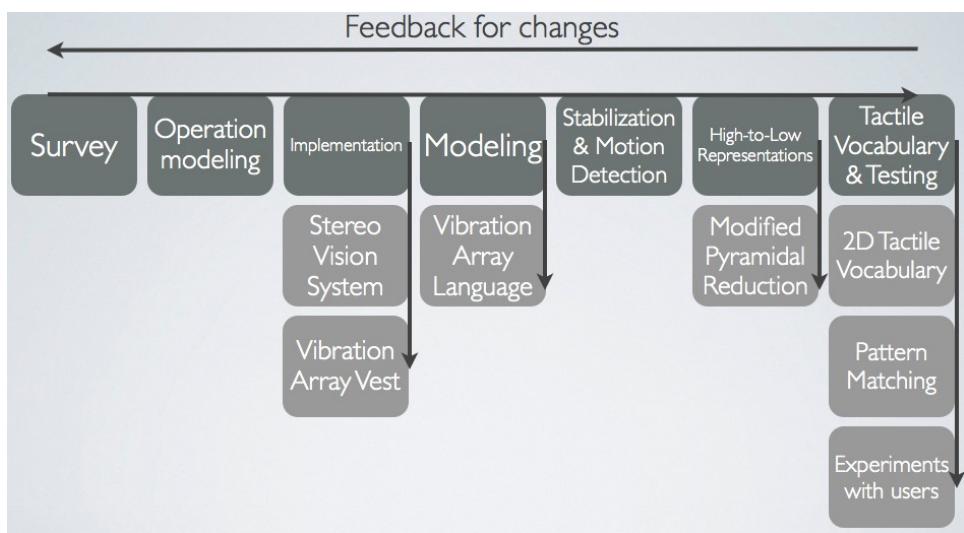


Fig. 1-1. Outline of dissertation.

The next step is the development of a camera stabilization and motion detection methodology: moving objects in the immediate environment of the user are crucial for the user's understanding of his/ her navigation space while

the camera stabilization is an important preprocessing step for dealing with the natural movement of the cameras because of the user's gait. The sixth step discusses the high-to-low resolution methodologies used for reducing the high resolution camera output to the low resolution of the vibration array. Here navigational criteria were used.

In the final step we discuss about the development of a tactile vocabulary (VAL vocabulary) and the use of a new pattern matching methodology. The methods were tested on subjects and the feedback and statistical results collected will be used as a starting point for modifications throughout the design and modeling of the prototype.

## **Chapter 2: A Survey on Electronic Travel Aids for Blind**

Since 1960's, evolving technology helped many researchers built electronic devices for navigation. A first level categorization is the following: (i) vision enhancement, ii) vision replacement, and iii) vision substitution. The function of any sensory aid, as described in [4], is "*...to detect and locate objects and provide information that allows user to determine (within acceptable tolerances) range, direction, and dimension and height of objects. It makes non-contact trailing and tracking possible, enabling the traveler to receive directional indications from physical structures that have strategic locations in the environment*".

Vision enhancement involves input from a camera, process of the information, and output on a visual display. In its simplest form it may be a miniature head-mounted camera with the output on a head-mounted visual display (as used in some virtual reality systems). Vision replacement involves displaying the information directly to the visual cortex of the human brain or via the optic nerve. We will not deal with this category since they deal with scientific, technological and medical issues whose study is beyond the purpose of this survey. Vision substitution is similar to vision enhancement but with the output being non-visual, typically tactful or auditory or some combination of the two and since the senses of touch and hearing have a much lower information capacity than vision, it is essential to process the information to a level that can be handled by the user. The category that we will focus in this work is the

“vision substitution”. Here someone can find these subcategories:

- **Electronic Travel Aids (ETA):** they transform information about the environment that would normally be relayed through vision into a form that can be conveyed through another sensory modality.
- **Electronic Orientation Aids (EOA):** they provide orientation prior to, or during the travel. They can be external to the user and/ or can be carried by the user (e.g. infrared light transmitters and handheld receivers).
- **Position Locator Devices (PLD):** they provide location information and incorporate technologies as GPS, EGNOS etc.
- Combination of the above.

We are mostly interested in ETAs and more specifically in obstacle detection systems, not emphasizing in any PLD characteristics. ETAs can also be categorized depending on how the information is gathered from the environment and depending on how this information is given to the user. Information can be gathered with sonars, laser scanners or cameras and the user can be informed through the auditory and/ or tactile sense. Sounds or synthetic voice are the options for the first case and electrotactile or vibrotactile stimulators for the second. Tactile feedback has some great advantage because it doesn't block the auditory sense (free-ears), which is the most important perceptual input source (the others are touch, wind, odors and temperature) for a visually impaired user.

Additionally, some ETAs offer to the user free-hands since they are

wearable but some others don't since the user is required to hold them; it's up to the user to select which is more appropriate to his/ her habits.

## 2.1 Electronic Travel Aids

Twenty-two ETAs (five are products already in the market) are briefly described and reviewed and finally a maturity analysis for all the aids is provided. We will study these systems taking the above guidelines into consideration and then give some comparative results to answer the questions of how advance, useful and desirable each system is. The systems are presented based on how the feedback is sent to the user (Fig. 2-1). The first eight, use audio feedback, the next seven, tactile and the last two do not offer any interface in that stage of their development.

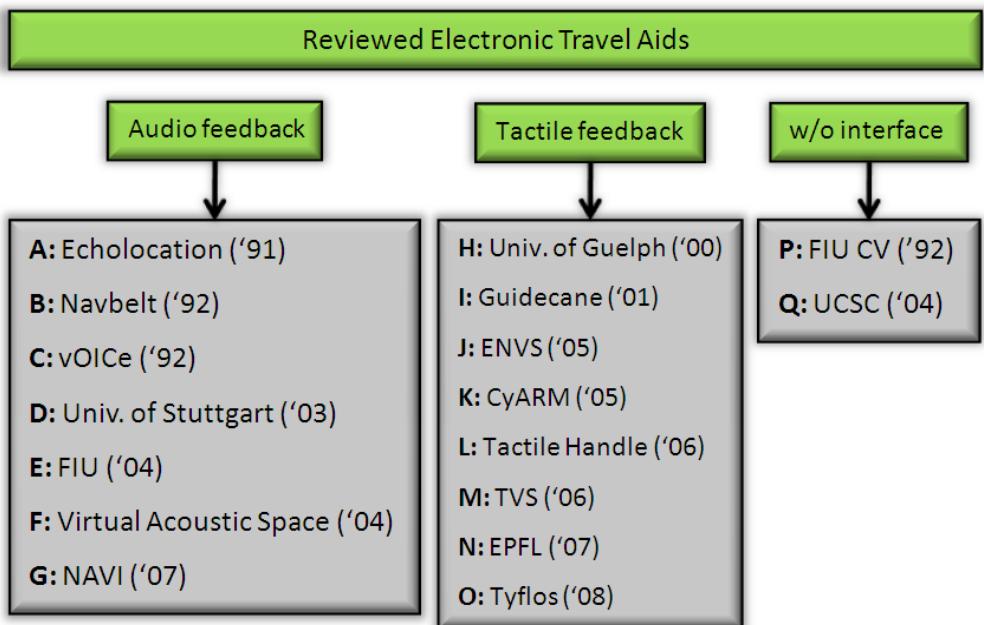


Fig. 2-1. The reviewed ETAs with categorization based on how the information is sent to the user (the latest prototype's year is shown in parenthesis).

### **2.1.1 *Echolocation***

The main goal of this project, which started in the early 90s in Japan, was to design a new mobility aid modeled after the bat's echolocation system [7]. The two ultrasonic sensors are attached on conventional eyeglasses and their data, using a microprocessor and A/ D converter, are down-converted to a stereo audible sound, sent to the user via head-phones. The different intensities and time differences of the reflected ultrasound waves transmitted by the sensors indicate the different directions and sizes of obstacles, creating a form of localized sound images.

Some preliminary experiments were performed to evaluate the user's capability to discriminate between objects in front of the user's head, using different ultra-sound frequencies. The results provided show that the users can identify and discriminate objects in some limited cases but more experiments and statistical results are required to support the viability of the project. The simplicity and portability of the prototype are also major advantages.

### **2.1.2 *Navbelt***

Navbelt is developed by Borenstein et al. in University of Michigan [8] as a guidance system using a mobile robot obstacle avoidance system. The prototype as implemented in 1992 and it is consisted of: ultrasonic range sensors, a computer and earphones. The computer receives information from the 8 ultrasonic sensors (Fig. 2-2) and creates a map of the angles (each for every sensor) and the distance of any object at this angle. Then the obstacle avoidance algorithm (including noise reduction algorithm EERUF) produce sounds appropri-

ate for each mode.

Navbelt has 2 modes; the guidance mode and the image mode. During the guidance mode, the computer knows the user's destination and with a single recurring beep guides him/ her in the generated optimal direction of travel. But in practice, a realistic (non-simulation) implementation would require more sensors. In the image mode, 8 tones of different amplitudes are played in quick succession from 8 different virtual directions (similar to a radar sweep). The computer translates (depending on the mode) these maps to sounds that the user can listen from his earphones. The disadvantages of the systems are the use of audio feedback (exclusively), the bulky prototype and that the users are required extensive training periods.

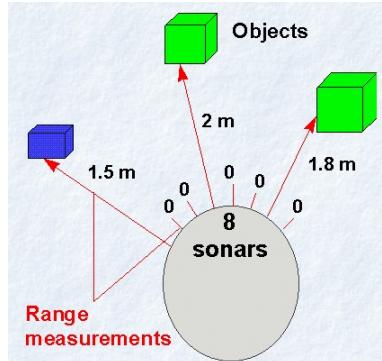


Fig. 2-2. The Navbelt's operation: the 8 ultrasonic sensors create an angle map which represents the distance of obstacles in those angles.

### 2.1.3 vOICe

P. Meijer [9] started a project with the basic argument that human hearing system is quite capable of learning to process and interpret extremely complicated and rapidly changing sound patterns. The prototype shown in Fig. 2-3 consists of a digital camera attached to conventional eye-glasses, headphones and a portable computer with the necessary software.

The camera captures images and the computer uses a direct, unfiltered, invertible 1-to-1 image-to-sound mapping. The sound is then sent to the headphones. No filters were used to reduce the risk of filtering important information since the main argument is that human brain is powerful enough to process complex sound information. The system is very simple, small, lightweight and cheap. Lately the software was embedded on a cell-phone and thus the user can use the cell-phone's camera and earphones. Additionally sonar extension is available for better representation of the environment and increased safety. Many individuals tried the system returning very promising feedback but they required extensive training because of the complicated sound patterns.



Fig. 2-3. An implementation of the vOICe – “Seeing with sound” system (glasses with attached camera, ear speakers and portable computer).

#### 2.1.4 *University of Stuttgart project*

A portable-wearable system that assists blind people orienting themselves in indoor environments was developed by researchers in University of Stuttgart in Germany [10]. The prototype is consisted of a sensor module with a detachable cane and a portable computer. The sensor (Fig. 2-4) is equipped with 2 cameras, a keyboard (similar to those in cell phones), a digital compass, a 3D inclinometer and a loudspeaker. It can be handled like a flashlight and “By

*pressing designated keys, different sequence and loudness options can be chosen and inquiries concerning an object's features can be sent to the portable computer. After successful evaluation these inquiries are acoustically answered over a text-to-speech engine and the loudspeaker”.*

The computer contains software for detection of color detection distance and size of objects and WLAN capabilities. The device works almost in real time. In order to improve the performance of the system, a virtual 3D model of the environment was built, so the information from the sensor can be matched with the data stored in the 3D model. A matching algorithm for sensor information and 3D model's data and embedding the system to Nexus framework (a platform that allows a general description of arbitrary physical real-world and virtual objects) are the future work proposals.

Concluding, the system's positives are the robustness of the sensor, the near real-time operation and the friendliness to the user. The negatives are that the hold-and-scan operation and the, until this moment, limited, simulated testing.



Fig. 2-4. The sensor of the project from University of Stuttgart.

### 2.1.5 FIU project

This project from researchers in Florida International University [11] is an obstacle detection system that uses 3D spatialized sounds based on readings from a multidirectional sonar system. The prototype (Fig. 5) is consisted of 2 subsystems; The Sonar and Compass Control Unit which is consisted of six ultrasonic range sensors pointing in the six radial directions around the user and a microcontroller; and the 3D Sound Rendering Engine which is consisted of headphones and a PDA equipped with software capable of processing information from the Sonar and Compass Control.

The algorithm, using Head Related Transfer Functions (HRTF), creates a 3D dimensional sound environment that represents the obstacles detected by the sensors. The user in that way creates a mental map of the layout of his/ her surroundings so that obstacles can be avoided and open passages can be considered for path planning and navigation. The system was tested on 4 blindfolded individuals who were asked to navigate in a building. The results were promising but the navigation speed was slow. As seen in Fig. 2-5, the design of the ranging unit is not ergonomic but the system is small and wearable.

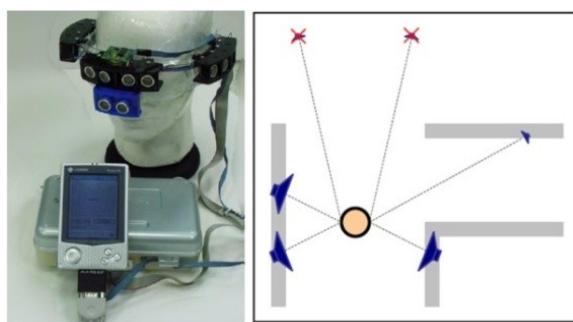


Fig. 2-5. The FIU project prototype (left) and an example of its operation (right).

### 2.1.6 Virtual Acoustic Space

Virtual Acoustic Space was developed by researchers in Instituto de Astrofísica de Canarias (IAC) [12]. A sound map of the environment is created and so the users can orientate by building a perception of space itself at neuronal level.

The prototype (Fig. 2-6) is consisted of two color micro cameras attached to the frame of some conventional eye glasses, a processor and headphones. The cameras, using stereoscopic vision, capture information of the surroundings. The processor, using HRTF, creates a depth map with attributes like distance, color or texture and then generates sounds corresponding to the situation in which sonorous sources exist in the surroundings. The experimental results on visually impaired people showed that in most cases ( $>75\%$ ), individuals could detect objects and their distances and in small simple experimental rooms, it was possible for them to move freely and extract information for objects like walls, table, window and opened door.

The major advantage of this system is that the eye glasses are convenient and the size of the processor is small (like a portable CD-player). The major disadvantage is that is not tested in real environments.



Fig. 2-6. Virtual Acoustic Space prototype (cameras and headphones mounted on eyeglasses and microprocessor (right)).

### 2.1.7 NAVI

Sainarayanan et al. from University Malaysia Sabah [13] developed an ETA (sound-based) to assist blind people for obstacle identification during navigation, by identifying objects that are in front of them. The prototype NAVI (Navigation Assistance for Visually Impaired) (Fig. 2-7) is consisted of a digital video camera, headgear (holds camera), stereo headphones, the Single Board Processing System (SBPS), rechargeable batteries and a vest (that holds SBPS and batteries).

The idea is that people focus in objects that are in front of the center of vision and so it's important to distinguish between background and obstacles. The video camera captures grayscale video which is re-sampled to 32x32 resolution. Then using a Fuzzy LVQ neural network the pixels are classified to either background or objects using different gray level features. Then the object pixels are enhanced and the background suppressed. The final stage cut the processed image into left and right parts, transform to (stereo) sound that is send to the user through the headphones.

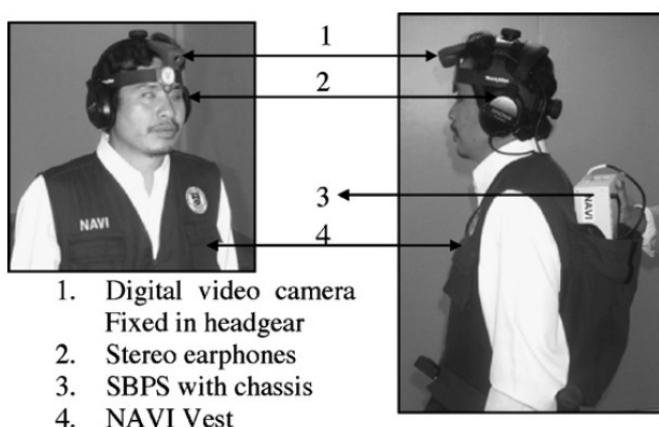


Fig. 2-7. The NAVI and its components.

Blind persons were trained with simulated experiments and then asked to identify obstacles of indoor environment and they were able to identify slowly moving objects. Although the distance of objects was not (and is not) aimed to be identified, it is possible to be done by the change of an object's shape e.g. when the user approaches an object, its size will become bigger. The advantage of this system is that the prototype is developed and it is operational and real-time. The disadvantages are the use of audio feedback and that no information about the distances of objects is given.

#### *2.1.8 University of Guelph project*

J. Zelek with students from University of Guelph [14], in Canada developed an inexpensive, built with off-the-shelf components, wearable and low power device that will transform depth information (output of stereo cameras) into tactile or auditory information for use by visually impaired people while navigation.

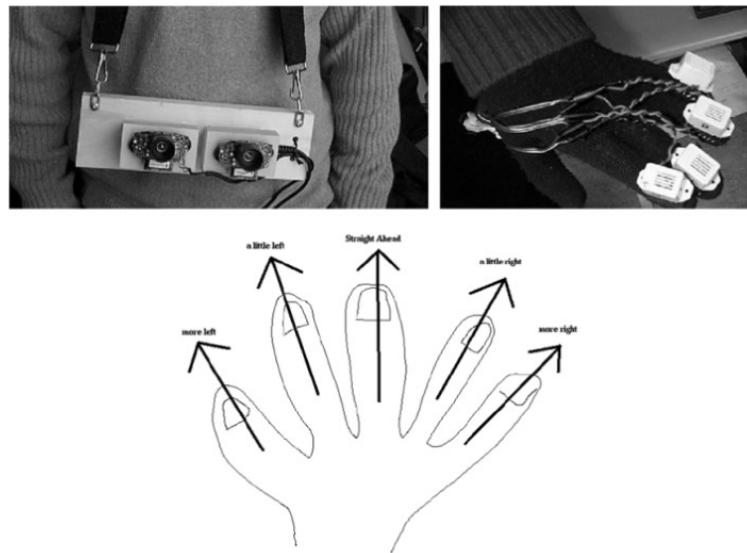


Fig. 2-8. Prototype from University of Guelph and the hand spatial correspondence.

The prototype, shown in Fig. 2-8 (top) is consisted of two stereo cameras, a tactile unit (glove with 5 piezoelectric buzzers on each fingertip) and a portable computer. Each finger corresponds to a spatial direction (Fig. 8 bottom). For example the middle finger corresponds to straight ahead. Using a standard stereo vision algorithm, the depth map is created and then divided into 5 vertical sections each one corresponding to a vibration element. If a pixel in an area corresponds to a threshold distance (here 3 feet) then the corresponding vibration element is activated, informing the user about a close obstacle in that direction. The low power/ cost are the pros but the lack of sophisticated methodologies (e.g. the stereo-vision algorithm needs improvement) does not offer interesting results.

### 2.1.9 *Guidecane*

Guidecane [15] is the second project by Borenstein and it serves as an update for Navbelt. It is a device that the user can hold like a white cane and that guides the user by changing its direction when an obstacle is detected (Fig. 9).

The sketch of the prototype is shown in Fig. 2-9 (left). A handle (cane) is connected to the main device. The main device has wheels, a steering mechanism, ultrasonic sensors and a computer. The operation is simple: the user moves the Guidecane and when an obstacle is detected the obstacle avoidance algorithm chooses an alternate direction until the obstacle is cleared and route is resumed (either in a parallel to the initial direction or in the same). There is

also a thumb operated joystick at the handle so that the user can change the direction of the cane (left or right). The sensors can detect small obstacles at the ground and sideways obstacles like walls.

Compared to the competitive ETAs, the Guidedcane does not block the users hearing with audio feedback and since the computer automatically analyzes the situation and guides the user without requiring him/ her to manually scan the area, there is no need for extensive training. The drawbacks are the limited scanning area since, small or overhanging objects like pavements or tables cannot be detected and that the prototype is bulky difficult to hold or carry when needed.

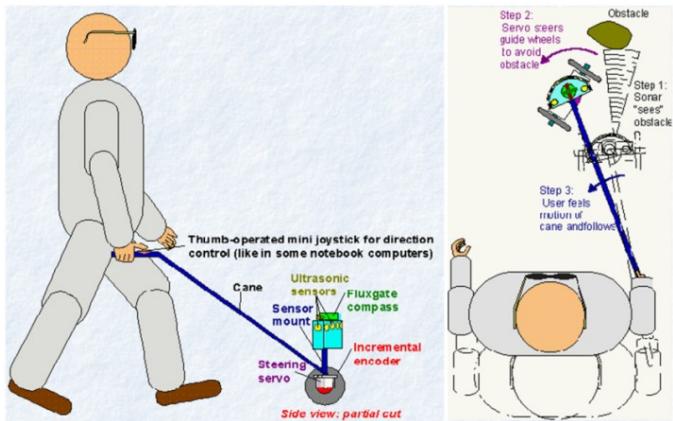


Fig. 2-9. Schematic (left) and operation (right) of the Guidedcane prototype.

### 2.1.10 ENVS

The ENVS (Electron-Neural Vision System) by Meers and Ward from University of Wollongong in Australia [16] aims to achieve obstacle avoidance and navigation in outdoor environments with the aid of visual sensors, GPS and electro-tactile simulation. The prototype (Fig. 2-10) is consisted of a headset with 2 stereo cameras and digital compass, a portable computer with GPS ca-

pabilities and database of landmarks, the TENS unit (microcontroller) and the TENS gloves.

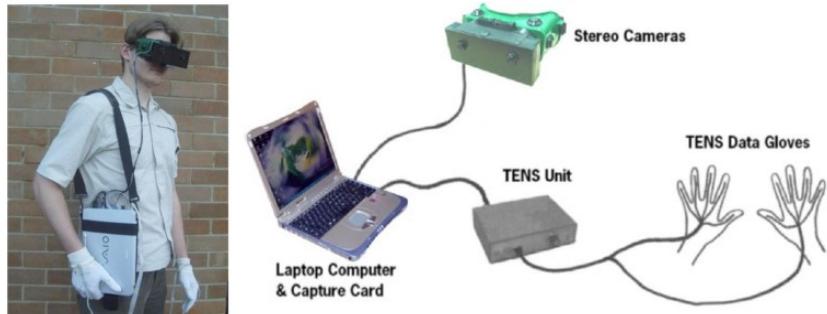


Fig. 2-10. The ENVS and its components.

The basic concept behind the ENVS prototype is: the stereo cameras, using stereoscopic vision, create a depth map of the environment and using the portable computer, information regarding the obstacles (from the depth map) or landmarks (from GPS) is transformed via TENS to electrical pulses that stimulate the nerves in the skin via electrodes located in the TENS data gloves. The user perceives the information if imagines that his/ her hands are positioned in front of abdomen with fingers extended. The amount of stimulation is directly proportional to the distance of the objects in the direction pointed by each finger.

The prototype was tested with blindfolded users in outdoor campus environment, working in real-time (video of 15frames/ sec). With a minimum training (1hour) the users were able to report the location of obstacles, avoid them and arrive at a predefined destination. The system is one of the most complete in this survey because it is portable, real-time, it has GPS capabilities, it does not block user's hearing and the first experimental results are very promising. Some of the drawbacks are that the ground or over-hanging objects

are not detected, that a flat path is required (i.e. no stairs or drop-offs) and that the user is required to wear the TENS gloves.

### 2.1.11 CyARM

CyARM is developed by researchers in Japan (Future University-Hakodate, Kanazawa University, Ochanomizu University and Fuji Xerox Co. Ltd) [17]. It is an aid for use in guiding orientation and locomotion, using a non-standard interface: ultrasonic sensors detect obstacles and calculate their distance from the user. The user is informed about the distance via the tension of a wire that is attached on him (e.g. his belt): high tension indicates close distance (the user can reach the obstacle by extending his/ her hand), while a lower tension indicates longer distance.

The prototype is a hand-held device weighting 500gr. It contains a microcontroller that processes the information from the sensors and operates a geared motor/ reel which controls the tension of the wire (Fig. 2-11).

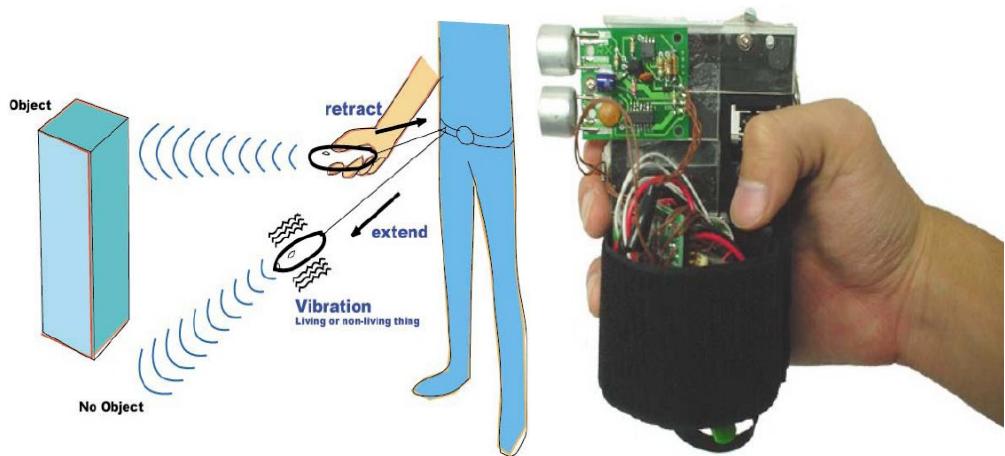


Fig. 2-11. The prototype CyARM and the concept of operation.

Small scale experiments were performed to evaluate CyARM's efficiency.

cy in detecting obstacles, navigation-through paths and target-tracking. The results for the obstacle detection and navigation-trough tasks were promising since more than 90% of the times the subjects were able to detect the large obstacles placed in front of them or to judge if it's possible to navigate through two of them. On the contrary, the moving target tracking results were not so encouraging.

The system's major advantage is its easy-to-learn (as the authors claim) alternative interface. The main disadvantages are that the user needs to hold it and scan the environment continuously and the lack of many experimental results with visually impaired users.

### ***2.1.12 Tactile Handle***

Bouzit et al. from State University of New Jersey developed the Tactile Handle [18], [19], a device that will help visually impaired people navigate in familiar and non-familiar environments without any assistance. The prototype is a compact (5cm x 5cm x 20cm), lightweight, ergonomic, low power (80 hours autonomy) handheld device. It embeds a microcontroller, a 4x4 tactile array where each actuator matches one finger phalanx, and 4 sonar sensors which detect obstacles in the front, left, right and bottom.

Information about the obstacles is given in an encoded form through the actuators. The location of the feedback represents different direction of the obstacle (Fig. 2-12). The intensity represents different distance and the timing of the feedback makes the user feel more comfortable and helps him/ her understand dynamic aspects of the environment such as speed. Simple experiments

with blind-folded users were performed in controllable indoor environments. The results show that training is necessary and the device can perform as an obstacle detection system.

The contributions of this project are mostly the development of low power ergonomic and compact prototype actuators which don't block the user's hearing. On the other hand it requires from the user to constantly scan and use one of his/ her hand. Furthermore the results show that excessive training is necessary.

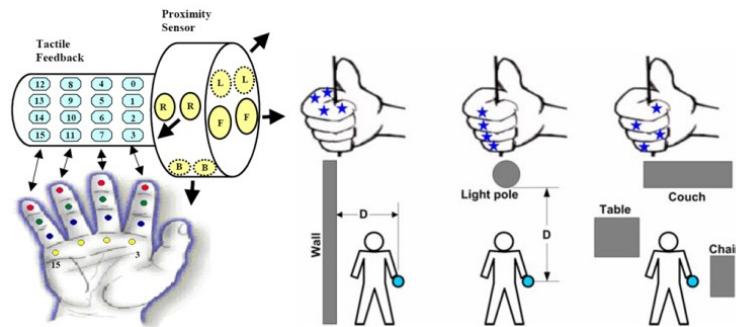


Fig. 2-12. Description of vibrotactile actuators positions on the Tactile Handle with three navigation scenarios.

### 2.1.13 TVS

The objective of Johnson and Higgins from University of Arizona [20] was to create a wearable device that converts visual information into tactile signal to help visually impaired people self-navigate through obstacle avoidance. The prototype is named TVS (Tactile Vision System) (Fig. 2-13) and is consisted of a tacter belt with 14 vibrator motors spaced laterally, a camera belt with 2 web cameras attached and a portable computer carried in a backpack.

A 2D depth map is created using the images from the two cameras. Then it is sliced in 14 vertical regions. Each vibrator motor is assigned one re-

gion and the value of the closest object in each region is transformed to vibration (Fig. 2-14). Vibration frequency and distance of object are non-linear (increases dramatically for closer objects) and very far or very close objects are ignored. Information given by the tacter belt is applied on the skin of the abdomen (flat, large, easily accessible, no interference with other navigation functions of user). Video is captured with rate up to 10frames/ sec which makes the system real-time for normal walking speeds.



Fig. 2-13. The TVS prototype.

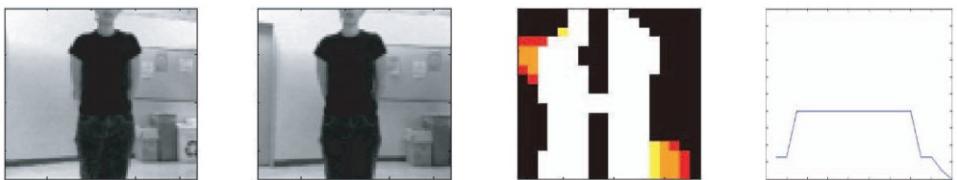


Fig. 2-14. An example of TVS operation: image from the two cameras, disparity map and the corresponding signals sent to the tacter belt.

The major advantages of TVS are that it is wearable, it gives user free hands without blocking hearing and it operates in real-time. The disadvantages are that it cannot differentiate between overhanging and ground obstacles and that no real experiments with visually impaired people have been performed.

Future works consists of using different stereo vision algorithms, different configuration of the tacter array and a possible VLSI implementation. Additionally studies will be performed on what type and what quantity is minimally necessary for navigation and what is the point of saturation beyond which perceptual improvements are minimal.

#### **2.1.14 EPFL project**

Cardin, Thalmann and Vexo from École Polytechnique Fédérale de Lausanne (EPFL) [21] developed a wearable system that detects obstacles on shoulder height via a stereoscopic sonar system and sends back a vibrotactile feedback to inform the user about its localization. The prototype consists of sonar sensors, a microcontroller, 8 vibrators and a calibration console (PDA).

The microcontroller (Fig. 2-15) gathers information from the sonars (Fig. 2-16) proportional to the distance of the obstacle detected. It calculates the approximate distance of the obstacle and then converts the distance to a PWM signal that is redirected to the vibrators (different vibration speeds), so that the user can be informed for the detection. The sonars and the vibrators are mounted on the clothes of the user, starting from one shoulder and ending to the other. Finally the calibration console communicates with the microcontroller via Bluetooth and allows dynamical modification of the calibration curve (real distance between object and sensor).

Experimental results were obtained by testing the device in a controlled indoor environment (corridor with people walking and doors opening and closing) on 5 users. The results were encouraging since the users managed after

a small training to walk through the corridor, distinguish obstacles (which are on the left or on the right side) and localize themselves in the corridor.

The pros of this project are that it is a wearable light, low power consumption and low-cost system. The cons are that is not tested on visually impaired people and that 4 sonars cannot represent adequately 3D space (different heights). Another practical problem mentioned by the authors is the interference of hands and their detection as obstacles.

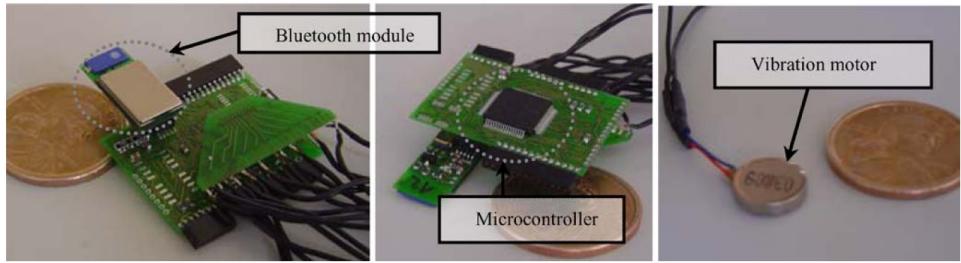


Fig. 2-15. Hardware details of the EPFL prototype.

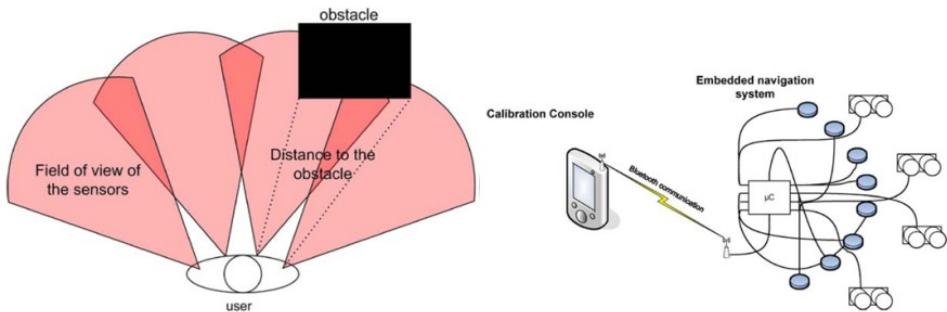


Fig. 2-16. Operation and high-level design of the EPFL prototype.

### 2.1.15 Tyflos

Tyflos navigation system was conceived by Bourbakis in the mid 90s and various prototypes have been developed [22]-[26]. The Tyflos navigation system is consisted of 2 basic modules: the Reader and the Navigator which is an ETA. The main goal for the Tyflos system is to integrate different navigation assistive technologies such as: a wireless handheld computer, cameras, range

sensors, GPS sensors, microphone, natural language processor, text-to-speech device, and a digital audio recorder etc and methodologies such as region based segmentation, range data conversion, fusion etc. in order to offer to the blind more independence during navigation and reading. The audio-visual input devices and the audio-tactile output devices can be worn (or carried) by the user. Data collected by the sensors is processed by the Tyflos' modules each specialized in one or more tasks. In particular, it interfaces with external sensors (such as GPS, range sensors, etc.) as well as the user, facilitating focused and personalized content delivery. The user communicates the task of interest to the mobility assistant using a multimodal interaction scheme.

The role of the Navigator is to capture environmental data from various sensors and map the extracted and processed content onto available user interfaces in the most appropriate manner. Previous Tyflos prototypes are designed using many of the technologies mentioned above and tested yielding promising results. The latest Tyflos Navigator system prototype developed in Wright State University is shown in Fig. 2-17. It consists of two cameras, an ear speaker, a microphone, a 2D vibration array vest (attached on the user's abdomen) controlled by a microprocessor and a portable computer and it integrates various software and hardware components.

The stereo cameras create a depth map of the environment (which can be verified by the range sensor's output). A high-to-low resolution algorithm drops the resolution of the depth map into a low resolution keeping necessary information for navigation such as safe navigation paths and objects of interest (moving objects and people; using motion detection and face-detection meth-

dologies). This final “image” is a representation of the 3D space and it is converted into vibration sensing on a 2D vibration array/ vest that is attached on the user’s abdomen or chest. The element of the array that vibrates represents the direction where an object is detected and the different vibration levels represent the distance of the object (Fig. 2-18). Optional audio feedback can inform the user for objects of interest.

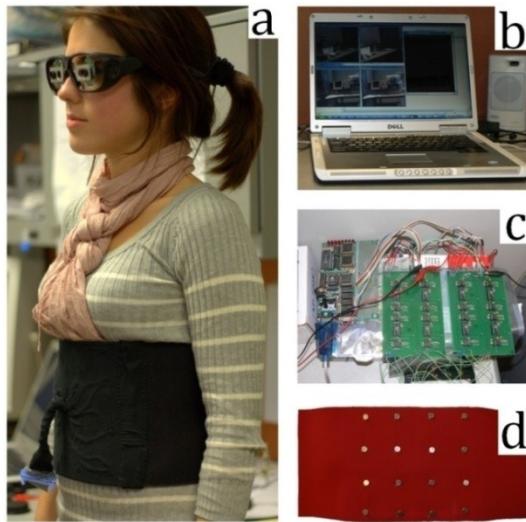


Fig. 2-17. User wearing the 2<sup>nd</sup> Tyflos’ prototype. a) stereo cameras attached on dark eyeglasses and vibration array vest on the user’s abdomen, b) portable computer, c) microcontroller and PCBs, d) arrangement of the 4×4 vibrating elements inside the vibration array vest.

The main advantages of the Tyflos are that it is free-ears and that the use of the 2D vibration array with the variable vibration frequencies offers the user a more accurate representation of the 3D environment (including ground and head height obstacles) giving also information for distances. The disadvantages are that the system is not yet tested on blind users, which is an important step for receiving feedback for future hardware and software changes.

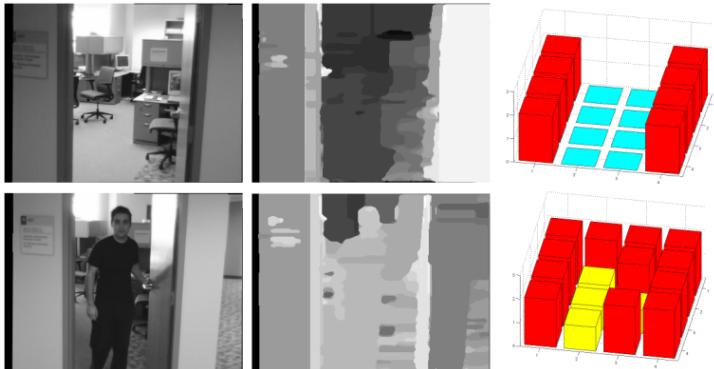


Fig. 2-18. Operation of Tyflos with two navigation scenarios (one in each row): Left column shows the images captured by the cameras. The middle the depth map and the right what the users senses via the 4x4 vibration array (light blue = no vibration = obstacle further than 4m, yellow = vibration level 1 = obstacle in range [2,4)m and red = vibration level 2 = obstacle in range [1,2)m ). In the first scenario the users feels there is an open path (door). In the second the user feels that there is no open path.

### 2.1.16 FIU Computer Vision project

M. Adjouadi from Florida International University [27] worked on a computer vision project in order to exploit, in an optimal fashion, the information acquired by cameras to yield useful descriptions of the viewed environment. Then, efficient and reliable cane cues can be sought in order to improve the mobility needs of individuals with visual impairments.

The system is consisted of digital cameras and a microcomputer which is equipped with software for detection of depression or drop-offs, discrimination of upright objects from flat objects, identification of shadows, identification of special objects (staircase, crosswalk, doorway etc), planning of safety path / direction. Although, this project is not yet to be considered as an operational ETA since issues, as how the user will be informed during navigation are still open, the algorithms are specially designed and implemented for navigation of blind and visually impaired. The author proposed audio verbal messages or tactile devices. As far as the software part, the strong points are that the algo-

rithms were tested with good results since many special cases are considered (staircases, vertical edges, depressions etc) with the limitation that there are good-lightning conditions.

#### *2.1.17 UCSC project*

R. Manduchi et al. from University of California Santa Cruz [28] developed a non-contact hand-held tool for range sensing and environment discovery for the visually impaired. The basic argument is that a perception through exploratory movements (similar to those using a white cane), appears to be a natural procedure for environment discovery. Thus, the tool is hand-held and as the user swings it around (vertical or horizontal) he/ she will receive information by means of tactile devices. The system deals only with one-dimensional data which is computationally cheaper than computer vision or spatial sound techniques. The prototype is consisted of a laser range sensor (point laser matched with a matrix CCD), as seen in Fig. 2-19, and a computer. The rage sensor is based on active triangulation. Additionally, the time profile of the range is analyzed by the computer to detect environmental features that are critical for mobility, such as curbs, steps and drop-offs (Fig. 2-20), by means of an extended Kalman filter tracker. The detection technique used works for detecting planar structures.

The system is reliable for local range measurements and gives promising environmental features detection. Additionally, although it is hand-held, it is small and easy to carry. The disadvantages is that it is not tested with visually impaired people, there is no interface between device and user and that it is

constraint in the detection of only planar structures and objects near the ground. Some of the future improvements that are proposed by the authors are: improvement of feature detection algorithms; replace of point laser with laser stiper; built in processor in the device will replace computer; tactile devices that will inform user for features detected.



Fig. 2-19. The UCSC's handheld device equipped with laser range sensor.

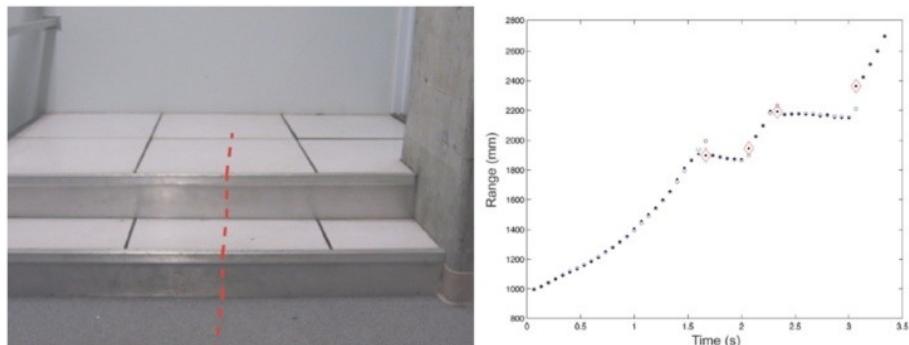


Fig. 2-20. The time profile of two steps acquired as the device was pivoted in an upward motion.

### 2.1.18 Commercial products

There are various commercial products available in the market. Their functionalities are limited and they have small scientific and technological value. Additionally their cost is relatively high and they are not widely accepted by the users. Therefore we will present some of them with small descriptions, without going into deeper analysis.

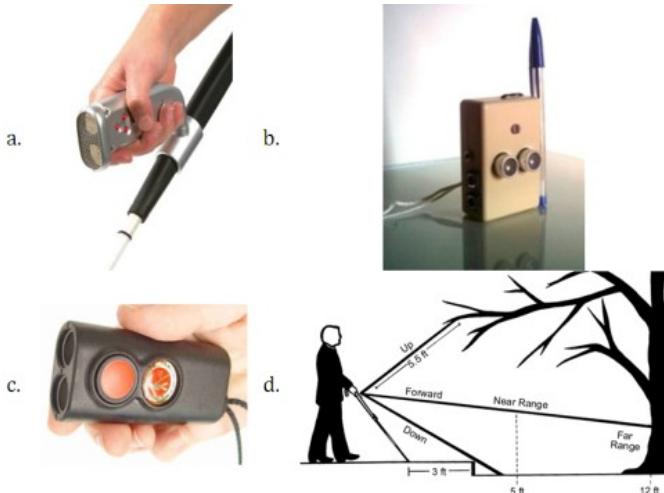


Fig. 2-21. Commercial products: a) K-Sonar Cane, b) Mini-radar, c) Mini-guide and d) LaserCane.

K-Sonar Cane [29] is a device that is attached in traditional white canes (Fig. 2-21-a). It is consisted of an ultrasonic range sensor (sonar) and a micro-processor that converts the distances to sound that the user can listen through earphones. Distant objects are related to high-pitch sounds and near objects to low-pitch. Its price is approximately \$700. Mini-Radar [30] is a device (Fig. 2-21-b) that uses sonar to detect frontal obstacles. It produces audio language messages when an object is detected. It can also provide information about the distance of the object. Another function is the “Directional Stability” that helps user to walk straight without changing his direction. It is priced approximately \$600. Miniguide [31] is a small device like a hand-light (Fig. 2-21-c) that indicates the distance to the closest object, via its vibration rate. It has multiple modes and ranges (up to 8m). The faster the vibration rate, the closer the object. The aid has an earphone socket which can provide audio feedback. It is priced at approximately \$330. LaserCane [32] is a cane with 3 laser range sensors: for head-height, straight-ahead and drop-offs obstacles (Fig. 2-21-d) and an audio

system that produces warning sounds (or corresponding to the obstacles distance) and vibration stimulators for warnings. The user can select between sound, vibration or both. It is priced at approximately \$3000. Ultracane [33] is also a cane with embedded laser range scanners. If an obstacle is detected then certain vibration buttons warn the user. There are different vibrations for different directions and different vibration rates depending on the distance of the obstacle. Its price is approximately \$900.

## 2.2 Maturity Analysis

### 2.2.1 *Structural and operational features*

After discussion with several groups of visually impaired users, software developers and engineers we came up with a set of features that better represent their views about an ETA. Those features will be used for the maturity analysis of each ETA. Table 2-1 describes those features.

### 2.2.2 *Maturity tables*

At this point we attempt to quantitatively evaluate the systems' progress/ maturity in order to offer some reference work, rather than a comparison. For every feature we assign a weight ( $w_i$ ) which reflects its importance from the user's view. The weights are calculated using a win-or-lose one-by-one comparison described below.

Every feature is compared with every other feature. A binary table is created (Table 2-2) following this procedure: if the feature from the row i is more important than feature from column j then we assign element (i,j) of the table as 1 (win). If it is less important, we assign else 0 (lose). The weight for

every feature is calculated by summing the number of the 1s and normalizing to 1.

Table 2-1. Structural and operational features. Features F1 to F7 correspond to user's needs while F8 to F14 reflect the developer's and engineer's views.

#	Feature	Description
F1	<b>Real time</b>	The system operates fast enough so that the information exchange with the user is useful e.g. if an obstacle detection system needs 10 seconds to detect an obstacle that is 6 feet in front of the walking-user, then the device is not real time.
F2	<b>Wearable</b>	The device is worn on the user's body or as a piece of his clothing. Wearable devices are useful for applications that require computational support while the user's hands, voice, eyes, ears or attention are actively engaged with the physical environment. The interaction between the user and the device is constant. Another feature is the ability to multi-task: it is not necessary to stop what you are doing to use the device; it is augmented into all other actions.
F3	<b>Portable</b>	The device is light and small with an ergonomic shape so that the user can carry it without effort, for long distances and time.
F4	<b>Reliable</b>	The system functions correctly in routine but also in different hostile or/ and unexpected circumstances.
F5	<b>Low-cost</b>	The device is (or it will be, when it comes to the massive production stage) affordable for most users.
F6	<b>Friendly</b>	The device is easy to learn, easy to use and encourages the user to regard the system as a positive help in getting the job done.
F7	<b>Functionalities</b>	The number and the importance of the system's functionalities.
F8	<b>Simple</b>	The complexity of both hardware and software is small. The hardware parts are few and simple to use (from the user's part) and simple to build (from the designer's part).
F9	<b>Robust</b>	The device is well constructed so it can resist in difficult environmental conditions or in hard use. Its functionality varies minimally despite of disturbing factor influences. It can still function in the presence of partial failures.
F10	<b>Wireless connectivity</b>	The device is connected wireless to a computer (server/ database) in order to continuously exchange information. Additionally, part of the processing needed for its operation can be done on the remote computer.
F11	<b>Performance</b>	Overall performance
F12	<b>Originality</b>	The idea and the methodology are original promoting scientific and technological knowledge.
F13	<b>Availability</b>	The system is implemented. A device that is ready to use and real-time experiments can be performed e.g. a system that is only in the software stage is not available.
F14	<b>Future</b>	Future improvements or enhancements

Table 2-2. Binary table for calculation of weights of the features.

	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>	<b>F7</b>	<b>F8</b>	<b>F9</b>	<b>F10</b>	<b>F11</b>	<b>F12</b>	<b>F13</b>	<b>F14</b>
<b>F1</b>		0	0	0	0	0	0	0	0	0	1	0	0	0
<b>F2</b>	1		0	0	0	0	0	0	0	0	1	0	0	0
<b>F3</b>	1	1		0	0	0	0	0	0	0	1	1	1	1
<b>F4</b>	1	1	1		0	0	0	0	0	0	1	0	0	0
<b>F5</b>	1	1	1	1		0	0	0	0	0	1	0	1	1
<b>F6</b>	1	1	1	1	1		1	0	0	0	1	0	0	1
<b>F7</b>	1	1	1	1	1	0		0	0	0	1	0	0	0
<b>F8</b>	1	1	1	1	1	1	1		0	0	1	0	1	1
<b>F9</b>	1	1	1	1	1	1	1	1		0	1	0	1	1
<b>F10</b>	1	1	1	1	1	1	1	1	1		1	0	1	1
<b>F11</b>	0	0	0	0	0	0	0	0	0	0		0	0	0
<b>F12</b>	1	1	0	1	1	1	1	1	1	1			1	1
<b>F13</b>	1	1	0	1	0	1	1	0	0	0	1	0		1
<b>F14</b>	1	1	0	1	0	0	1	0	0	0	1	0	0	
<b>w<sub>i</sub></b>	0.93	0.86	0.57	0.71	0.50	0.43	0.57	0.29	0.21	0.14	1.00	0.14	0.50	0.64

Table 2-3. Table of scores for all systems and features; A-G: audio feedback, H-O: tactile feedback, P-Q: no interface.

Features		<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>	<b>F7</b>	<b>F8</b>	<b>F9</b>	<b>F10</b>	<b>F11</b>	<b>F12</b>	<b>F13</b>	<b>F14</b>
<b>Norm. Weights</b>		9.3	8.6	5.7	7.1	5.0	4.3	2.7	2.9	2.1	1.4	10.0	1.4	5.0	6.4
<b>System Scores</b>															
<b>A</b>	<b>Echolocation</b>	9	9	9	4		3	3	9	5		2	5	10	3
<b>B</b>	<b>Navbelt</b>	7	5	7	4	6	3	4	6	6		5	5	10	1
<b>C</b>	<b>vOICe</b>	6	9	9	3	9	3	3	9	5		6	4	10	6
<b>D</b>	<b>University of Stuttgart</b>	8	0	9	6	7	8	6	6	7	10	6	8	10	8
<b>E</b>	<b>FIU</b>	7	8	9	5	7	4	5	7	5		6	5	10	8
<b>F</b>	<b>Virtual Acoustic Space</b>	6	9	9	6	8	4	5	6	5		6	7	10	7
<b>G</b>	<b>NAVI</b>	9	8	8	6	8	4	5	6	5		6	7	10	7
<b>H</b>	<b>University of Guelph</b>	9	8	8	5	6	7	8	5	5		3	6	10	9
<b>I</b>	<b>GuideCane</b>	9	0	5	6	6	7	6	5	7		6	7	10	5
<b>J</b>	<b>ENVS</b>	9	8	8	8	6	7	9	5	7	10	7	8	10	7
<b>K</b>	<b>CyARM</b>	8	0	9	5	9	8	5	7	5		6	9	10	7
<b>L</b>	<b>Tactile Handle</b>	9	0	8	7	7	7	6	7	8		4	8	10	6
<b>M</b>	<b>TVS</b>	9	8	8	7	6	8	8	5	6		5	7	10	8
<b>N</b>	<b>EPFL</b>	9	9	9	6	9	9	5	9	6		6	8	10	8
<b>O</b>	<b>Tyflos</b>	6	8	8	7	5	9	8	5	5		6	9		9
<b>P</b>	<b>FIU cv project</b>	5	0	7	6	9	5	8	4	7			8		3
<b>Q</b>	<b>UCSC</b>	9	0	8	4	6	4	3	5	3		5	6	10	6

## 2.3 Discussion on the Survey

### 2.3.1 Scores

The table of scores (Table 2-3) reveals that there is no system incorporating all the features in a satisfactory degree. Features reflect mostly the user's perspective but also the designer's perspective. Every system offers something special over the others but it cannot meet all the needs, since an ideal system should have all the features (e.g. reliability, wireless capabilities, low price etc). The most important finding here is that there is no system yet that the visually impaired users are confident about its reliability, its robustness and its overall performance. This is because most of the systems are, in the best case, at the prototype stage and real time, long time experiments with visually impaired people have not been performed.

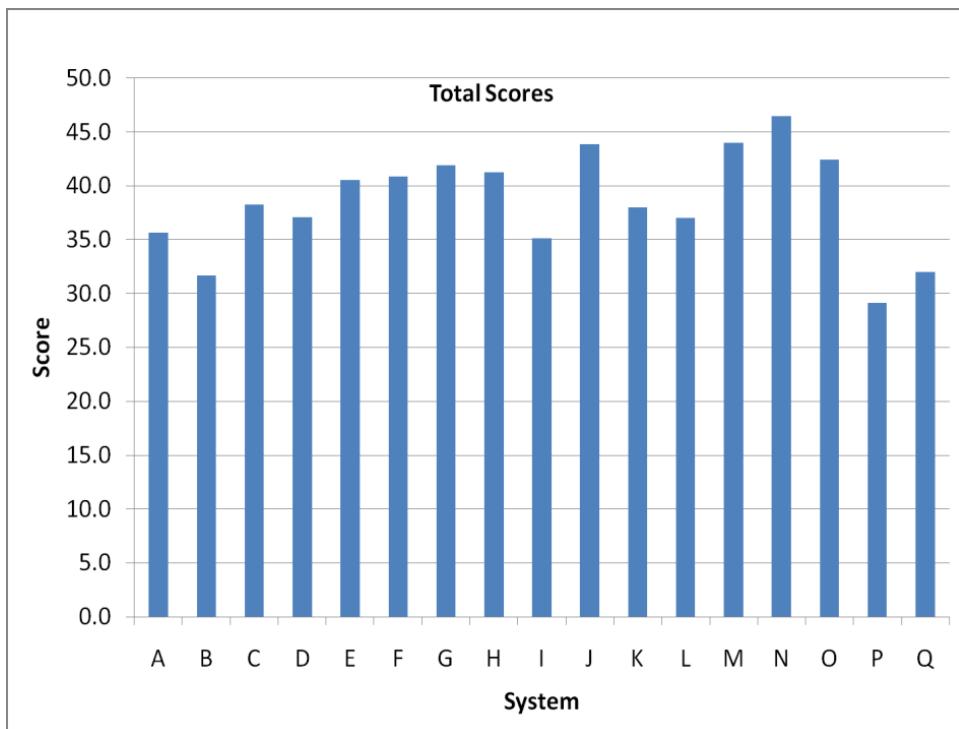


Fig. 2-22. Maturity ranking that shows the total score for each system.

The maturity ranking (Fig. 2-22) gives us a big picture for all the reviewed ETAs; a measure of the system's progress/ maturity. The ones with higher scores show better progress and/ or more features. The systems that got lower scores do not have less technological or usage value, but they are still in the early stage of their progress and they haven't reached their maximum of their performance. Finally we want to mention that the commercial products have limited functionalities, small scientific/ technological value and high cost, so they were excluded from the analysis.

### 2.3.2 Guidelines

After carefully studying the above systems and taking into account previous works [4], [34], [35] and our personal experience we gained from the research, design and development of our system, we can summarize some guidelines for the development of electronic travel aids. Some of them are already mentioned in the introduction of this survey.

According to Loomis and Golledge [36] there are three needs for people with visual impairments when navigating. The first is *accessing* information with the remaining senses, in other words the aids should provide information that the user cannot have by himself. The second is accessibility, meaning to be able to navigate safely in different environments. The last is *independence*. This is the most important and means that the ultimate goal for designers and engineers are to make users more independent human beings.

In the development of ETAs, the most challenging is to define the proper interface between the system and the user; how and what information is sent

to the user; define a robust human-computer interaction scheme. For example, early ETAs “...present as much information about the environment as possible... an excessive, confusing and unnecessary amount of information which only confused blind users” [37]. We believe that with appropriate modeling and processing is required so that only the necessary useful information should be presented to the user.

Concluding, we would like to emphasize in the following characteristics:

- **Free-hands**: not requiring from the user to hold them. Remember that the users will still hold the white cane, the most undisputable travel aid.
- **Free-ears**: despite the advantages of echolocation, spatial sound and similar techniques, the user’s ability to listen environmental should not be interfered.
- **Wearable**: it offers flexibility to the users and utilizes the advantages of wearable technologies.
- **Simple**: easy to use (operation and interface not loaded with unnecessary features) and without the need of extensive training period.

# **Chapter 3: System's Architecture & Operation**

The Tyflos prototype integrates various software and hardware components and each of them will be presented in detail in the following chapters. Here, an overview of the architecture and the operation of the prototype will be presented.

## **3.1 Hardware and software architecture**

Fig. 3-1 presents the high-level hardware architecture of the current prototype. The interaction with the user involves different modalities:

- **Visual:** the stereo cameras are the main input medium and they capture visual environmental information.
- **Speech commands:** using the microphone the user can give speech commands and operate the system.
- **Vibration Array vest:** the user perceives information about the environment (obstacles, objects, paths etc) in a two dimensional tactile manner.
- **System commands:** via the ear-speaker the system informs the user with speech messages, about the environment (Navigator module) or about texts currently being read e.g. newspaper (Reader module).

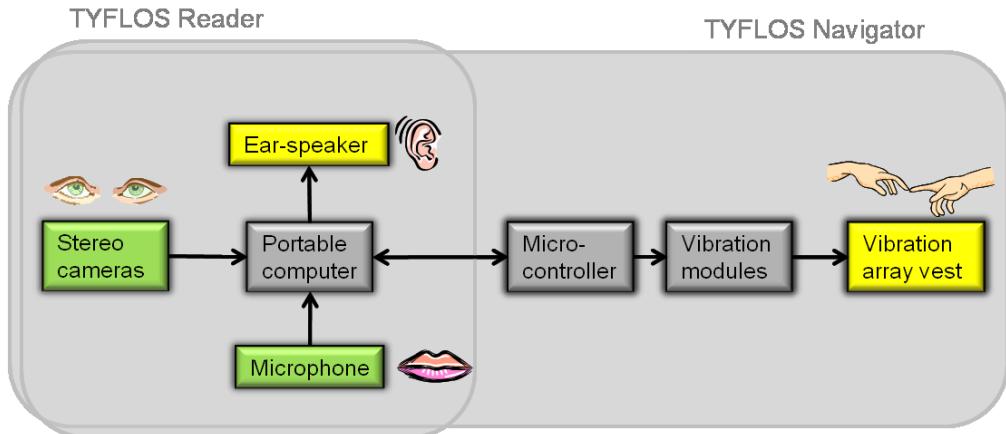


Fig. 3-1. Hardware architecture of the Tyflos 2<sup>nd</sup> prototype, emphasizing on its multimodal characteristics. The green boxes are input media (stereo cameras: capture visual information, microphone: speech information from the user) and the yellow boxes are output media (ear-speaker: speech messages to the user, vibration array vest: tactile information to the user).

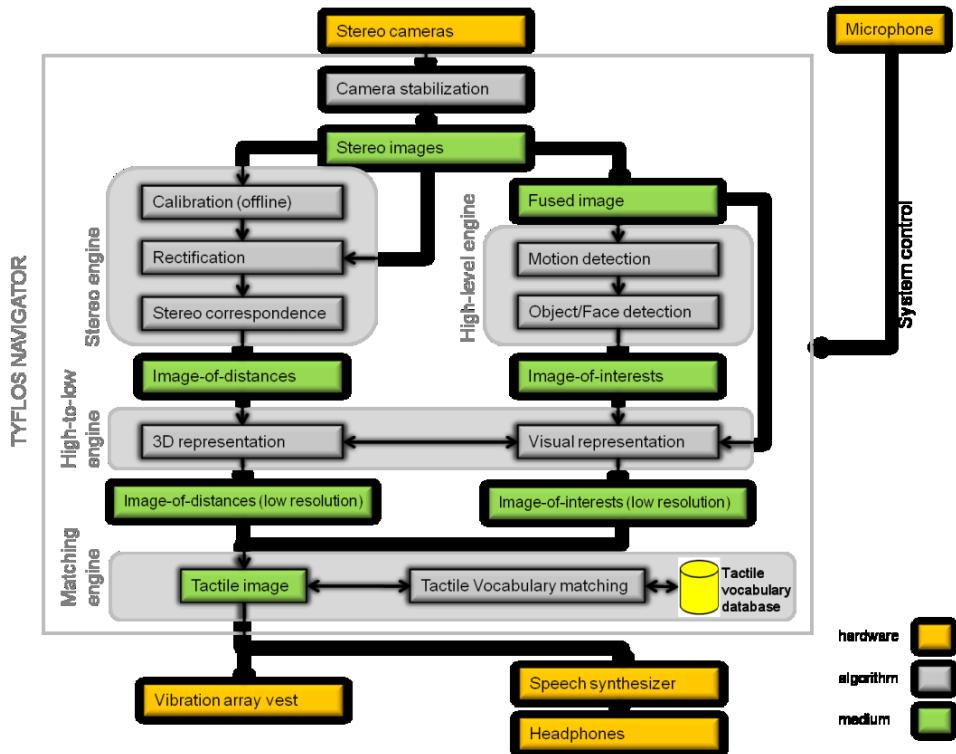


Fig. 3-2. Hardware and software architecture of the 2<sup>nd</sup> Tyflos prototype.

Fig. 3-2 presents the software architecture along with the involved hardware. An extended version, including the Tyflos Reader module, as well as the PLD (Position Locator Device) capabilities and the range verification com-

ponent, is shown in Fig. 3-3. It is important to mention that despite the fact that the system follows a one-way flow, an iterative feedback-based design of the software and hardware is followed so the system will fit the user's needs.

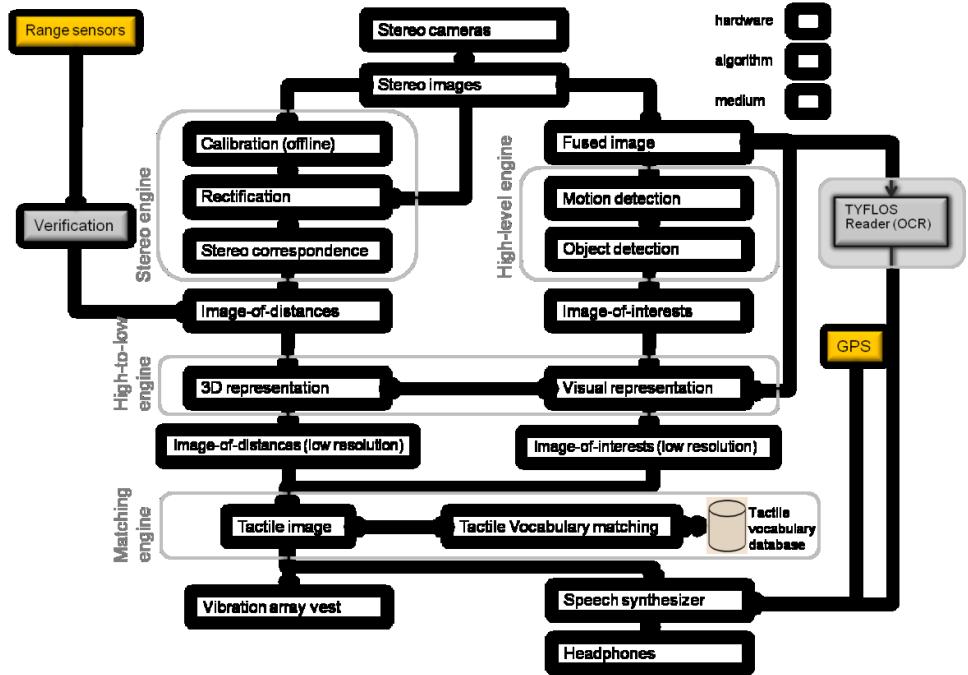


Fig. 3-3. Extended architecture of the 2<sup>nd</sup> Tyflos prototype, including the OCR-type Reader module, range sensor validation and PLD capabilities using GPS sensors.

## 3.2 Operation modeling

### 3.2.1 Stochastic Petri-Nets

There is a variety of methodologies used for operation modeling, such as formal languages, directed graphs, classical mathematical models, queuing models, Petri-Nets etc. In the next paragraphs of this chapter, a modified version of a Stochastic Petri-Net (SPN) will be used to model the framework of the functionality of Tyflos wearable prototype [38]. The major reasons for using the Stochastic Petri-Nets model rather than a Hidden Markov Model (HMM) are:

- SPN is an efficient modeling tool for the functional description and analysis of complex systems;
- SPN is able to simultaneously describe concurrency, parallelism, and synchronization of events that take place in a complex system, especially when other methodologies lack adequate results;
- SPN can be used as a modeling tool for hierarchical and abstracted (top-down or bottom-up) processes;
- SPN provides timing during the execution of various events;
- SPN presents compatibility with neural networks and
- SPN is an efficient interface for control and communication.

**Definition:** A generalized Petri-Net [39] model can be defined as

$$(Eq. 3-1) \quad SPNG = \{P, T, A, I, O, M, X, C, L, D, S\}$$

where:

- **P:** a finite set of places  $\{P_i, i \in Z\}$  that represent a particular state of a physical component. Here, for convenience, we will use more letters: P for system states (or places) or pseudo-states, U for system level commands, R for Reader's commands, N for Navigator's commands and D for system responses);
- **t:** a finite set of transitions,  $\{t_j, j \in Z\}$  that represent a process performed between two states;
- **A:** a finite set of arcs  $\{a_{rj}, r, i, j \in Z\}$  that represent relationships among places ( $P_i, P_j$ );
- **I<sub>i</sub>** ⊂ (PxT), represents the input function;

- $\mathbf{O}_i \subset (TxP)$ , represents the output function;
- $\mathbf{M}_i$ : a vector of marking (tokens  $\mathbf{T}$ ) ( $m_{ij}$ ,  $i,j \in Z$ ) that represent the status of the places;
- $\mathbf{X}$ : a vector of time values ( $x_i$ ,  $i \in Z$ ) related with the time required by a process to be performed;
- $\mathbf{C}$ : the alphabet  $\{c_i, i \in Z\}$  of communication;
- $\mathbf{L}$ : a finite set of possibly marking-dependent firing rates  $\{l_i, i \in Z\}$  associated with the transitions;
- $\mathbf{D}$ : a finite set  $\{d_i, i \in Z\}$  of delays associated with the transitions;
- $\mathbf{S}$ : a finite set of structural properties  $\{s_i, i \in Z\}$  associated with places.

In the following paragraphs we will present some of the operational interactive modes of the Tyflos prototype modeled by the SPN. These modes are presented in a hierarchical way. We start with a small set of commands in order to show the synergistic interaction among different modalities used by the Tyflos prototype.

### 3.2.2 Commands

Statistical speech recognition enables the recognition of a word, phrase, or sentence pronounced when matched with a finite set of possibilities. This technique is typically used in command-and-control settings, such as those found in issuing commands to the Tyflos prototype.

The speech recognition system used by Tyflos is called HTK and it is developed at University of Cambridge [40]. HTK is regarded as a state of the art research level Hidden Markov Model toolkit with speech recognition libra-

ries built on top of it. Thus it provides a good framework to prototype the speech recognition used in Tyflos.

HTK relies on a predefined set of interactions with the user. Grammars define words, phrases, and sentences for use during an interaction with a user. When a user speaks a word or phrase, HTK uses a probabilistic model to attempt to map the input to a grammar. Generally, the performance of HTK is inversely related to the size of the grammar vocabulary. However, for the Tyflos system, we have resolved this problem by keeping the grammar small.

HTK uses standardized models to represent each phoneme in the English language. Combinations of sequential phonemes were trained on acoustic data. Words in a knowledge base were thus constructed as interconnected networks of trained phonemes, using probabilistic transitions to describe the likelihood of two sound units being heard consecutively.

Grammars specify the constraints of the expected utterance. Grammars can be very simple or complex. A simple grammar may enable the user to simply respond with ‘yes’ or ‘no’ to a series of questions. Complex grammars may provide the ability for a user to give a lengthy command such as, “Find open paths and number of obstacles.” When a system is completely dependent on voice recognition, grammars are required to limit the vocabulary used in the interactions. Thus, in the Tyflos system we have limited the interaction commands, many of which are listed in Table 3-1.

Tyflos must be able to recognize various commands to control the device. Examples of commands include: Find paths, Find objects, Find people.

HTK provides a grammar definition language for specifying grammars for phrases like these. The language consists of a set of variable definitions followed by a regular expression which describes the phrases to recognize. The grammar could be depicted as a network, as shown in Fig. 3-4. A small portion of the grammar for the Tyflos Navigator module follows:

```
$command = NAVIGATE | STOP | PAUSE;
          (SENT-START( <$command> ) SENT-END )
$elements= PATHS | OBJECTS | PEOPLE;
          (SENT-START( FIND | (FIND <$elements>) ) SENT-END)
```

Table 3-1. User-System Commands

User voice-commands		
Name (System)	Action	Speech
<b>U0</b>	Puts system in IDLE state	“System on”
<b>U1</b>	Turns system off	“System off”
<b>U2</b>	Pauses current execution	“System pause”
<b>U3</b>	Resumes current execution	“System resume”
<b>U4</b>	Switch to Navigator mode	“System navigate”
<b>U5</b>	Switch to Reader mode	“System read”
(Navigator)		
<b>N0</b>	Find open paths / object avoidance	“Find paths”
<b>N1</b>	Recognize objects	“Find objects”
<b>N2</b>	Recognize faces	“Find people”

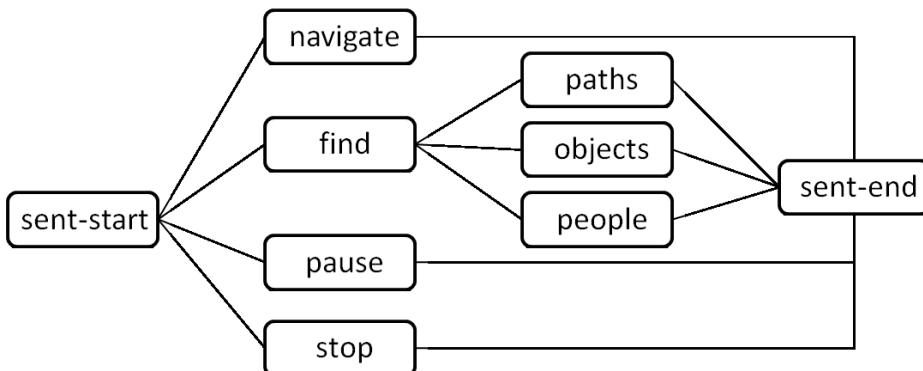


Fig. 3-4. Tyflos Navigator Grammar as Network

Another important aspect of using HTK is constructing a dictionary. The dictionary maps words to their phonetics. During training, a recording of a person speaking the words will be parsed and mapped to the phonetics. In this way the dictionary serves as a bridge between the actual sound and the grammar defined for an application.

Examples of words in the dictionary follow:

FIND	f ay n d
PATHS	p ae th s
OBJECTS	ax b jh eh k t s
PEOPLE	p iy p l
STOP	s t oh p
PAUSE	p ao z
NAVIGATE	n ae v ih g ey t

For a user to interact with Tyflos, we need more than a grammar definition. The speech recognition for Tyflos is created by assembling a series of prompts and responses into a usable interaction. This form of interaction lends itself to being modeled with Stochastic Petri Nets. This is due to the nature of the set of serial interactions with the user in which the user is presented with a series of questions and responds to the system accordingly. These responses are used as triggers to change the state of the system.

### 3.2.3 *The Tyflos hierarchical SPN model*

In this section we present the synergistic operation of the Tyflos modules in a hierarchical scheme modeled using SPNs. At the highest level, the graphical representation of the SPN model represents the interaction between the user and the assistive device. In particular the user commands the device to

operate in one of the two modes (Reader or Navigator) and the device responds to that command. This particular interaction is graphically shown in Fig. 3-5.

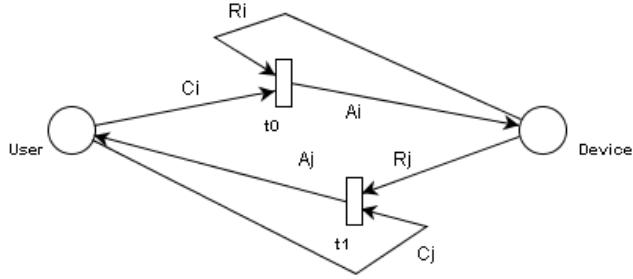


Fig. 3-5. Level 1 SPN model (user-device).

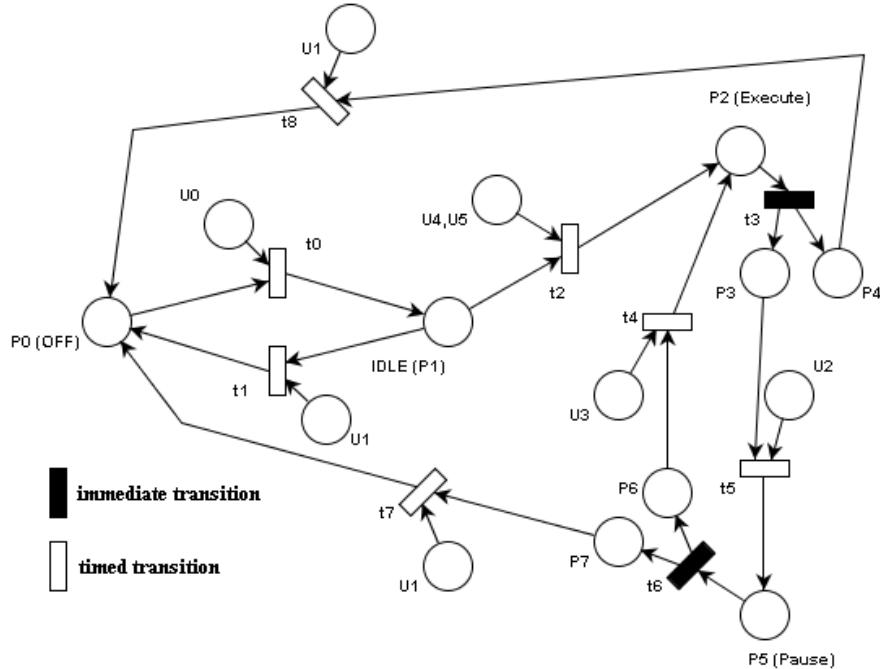


Fig. 3-6. Level-2 SPN model (device).

The next level in the model's hierarchy shows the various states into which the assistive device can be transferred based on the voice commands issued by the user and the current state of the device. In particular, at this opera-

tional level we present 4 states (off, idle, execute, pause) of the device to demonstrate the SPN model, as shown in Fig. 3-6. The third level in our hierarchy graphically presents the device's transition from the Reader mode to the Navigator mode following a user's command, see Fig. 3-7.

The following sections present each of the two modes (Reader, Navigator), and their corresponding SPN graphical model.

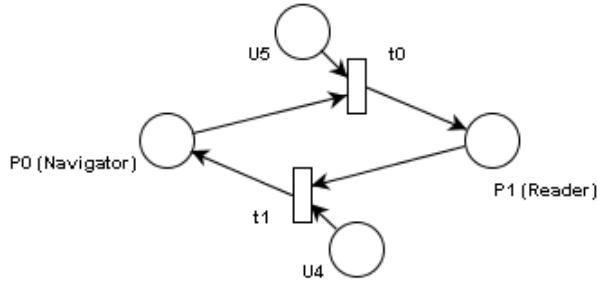


Fig. 3-7. Level-3 SPN model (modules interaction)

### 3.2.3.2 The Navigator's SPN module

Following the hierarchical scheme discussed in section 4.2, the SPN model for the Navigator module is shown in Fig. 3-8. In this model there are three primary states for the Navigator: the *Paths*, the *Objects*, and the *Faces*. The current state of the device proceeds from one state to another following user voice-commands N.

The *Paths* place represents the operational state for the obstacle detection system. The primary purpose of this state is to inform the user (via the VA) of obstacles and open paths. When in the *Objects* or *Faces* state, the device informs the user of objects or people respectively that may be of-interest, via the ear-speaker. It is important to notice that the model supports concurrent opera-

tional states of the system. Thus, the ability to provide information about people while safely navigating known open paths is supported.

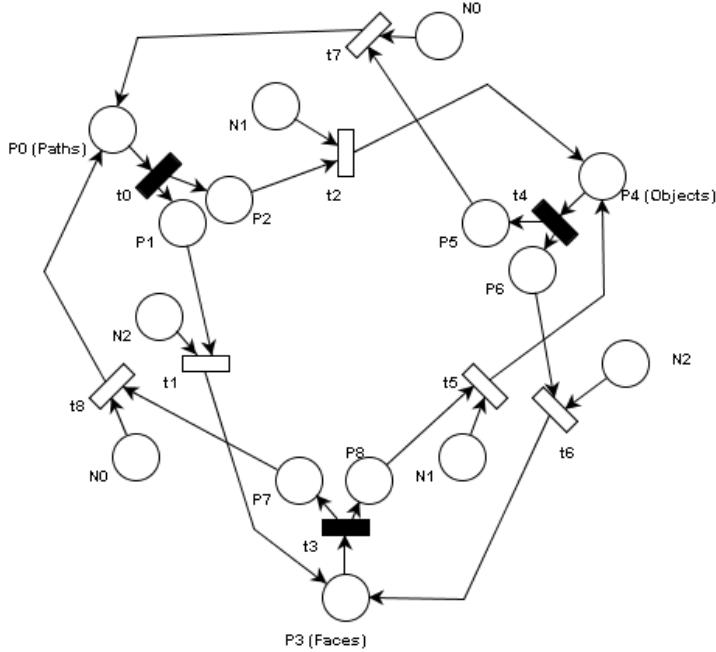


Fig. 3-8. Level-4 SPN model (Navigator) (note: the states without label correspond to pseudo-states).

### 3.2.4 Illustrative Examples

In this section we provide illustrative examples of the interaction between the various Tyflos modalities (images, audio and text).

#### 3.2.4.1 A case study

George was born blind. As he makes his way around his university campus, his tools for navigation are his long cane and senses of hearing, smell, and touch. By using the navigation module of the Tyflos system he is able to receive tactile feedback through the vibration mesh on his abdomen. As George walks through the cafeteria he feels the vibration grow in intensity on his left side, but his long cane does not touch anything. He realizes that this is due to

the shelf that he can set his tray on at the cashier counter. George is thankful for the feedback from Tyflos because many times in the past he has run into this shelf since the overhang is too high for his cane to touch.

#### 3.2.4.2 *Navigator example*

As discussed already the primary input for the Navigator module is the stereo vision system: the two cameras and the supporting software. Two navigation scenarios are presented here to illustrate the system's operation. Both scenarios require the system to be in the Navigator mode. The user issues the command “*System navigate*” to ensure that system is in the proper state. A second command “*Show paths*” is issued to enable *Paths* mode of the system.

The first scenario illustrates the user walking toward an open door. As the user progresses, the cameras capture stereo images as shown in Fig. 3-9a and Fig. 3-9b, and Tyflos produces depth and distance maps (Fig. 3-9c and Fig. 3-9d). After applying a high-to-low algorithm on the center square area of the maps, and using the correspondences presented in Table 3-2, the final low-resolution projected 3D environment is presented on user's vibration array vest (Fig. 3-10a and Fig. 3-10b). Notice that the two middle vibration columns of the vest do not vibrate (black color) and the two side columns vibrate in vibration level 1 (see Table 3-2). This information presents the user with a safe yet narrow path in the center of his/ her navigation route.

A second scenario illustrates a person exiting a room, blocking the navigation path (Fig. 3-10a). This causes all 16 elements of the vibration array cells to vibrate, indicating that no open path can be identified (Fig. 3-10b). Additionally, if the system is in the *Face* mode the person's face would be detected

(Fig. 3-11a) and recognized. The user would be informed with an audible message such as “*Person in front of you*” after successful face detection or “*Alex in front of you*” after successful detection and recognition.

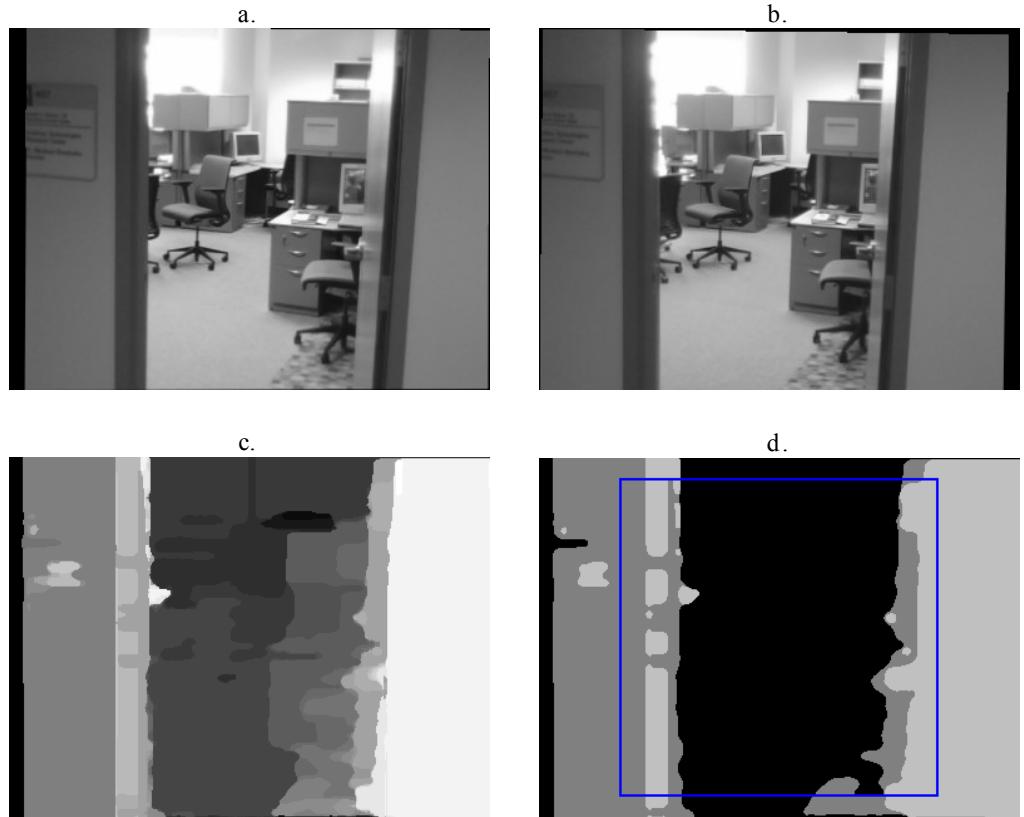


Fig. 3-9. Navigation scenario 1: a,b) Left and right camera 384×288 image (after rectification), c) Disparities map, d) Distances map with selected square area of 256×256.

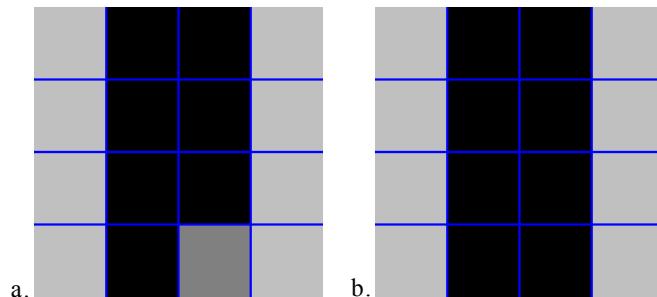


Fig. 3-10. Navigation scenario 1: Final 4×4 distances map, projected on user via the vibration array (the blue grid is used for visualize better the 4×4 vibrating cells of the array) a) classic pyramid, b) revised pyramid with navigational criteria.

Table 3-2. Vibration level, vibration frequency, distance and color representation correspondence.

Vibration level	Vibration frequency [Hz]	Distance range [m]	Distance characterization	Grayscale representation
3	8	(0,1)	Very close	White
2	4	[1,2)	Close	Light gray
1	2	[2,3)	Far	Dark gray
0	0	[3, $\infty$ )	Very far	Black

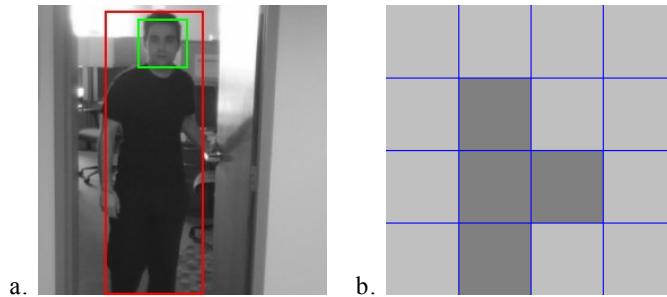


Fig. 3-11. Navigation scenario 2: a) Fused image from cameras, while the selected red area correspond to motion detection and the green square to detected person, b) Low resolution 4×4 distances map projected on the vibration vest.

# Chapter 4: Implementation

## 4.1 The cane paradigm

Tyflos, as well as most of the aids studied in the survey, are secondary navigation aids, supplementary for the first mobility aid which is usually the white cane. White cane is the most successful mobility aid and is being taught by Orientation and Mobility (O&M) instructors to the majority of the blind or visually impaired people.

Blasch, Wiener and Welsh [4] summarize its advantages as the high maneuverability, allowing investigation of the environment without actual hand contact. Additionally it is reliable, long-lasting, not affected by weather and temperature, cheap, almost no maintenance, can be accommodated to most users' physical specifications. The major limitations are that it offers no protection from **obstacles to the upper part of the body**, it requires the user to **continuously scan the area**, it cannot detect **obstacles further than the its length**, the difficulty to achieve a good foot placement preview, the interference of wind and finally the danger of tripping pedestrians in congested areas. We strongly believe that the white cane is an intuitive aid that is very difficult to be replaced, thus, studying its characteristics is a baseline for any designer of travel aids for blind or visually impaired, in order to understand the user's needs. For our design we will focus on resolving the first three (bold letters) limitation, which are also the most important according to users' opinions.

Before discussing the human factors of the prototype and the imple-

mentation details we will present the National Research Council's guidelines for ETAs [4]:

- Detection of obstacles in the travel path from **ground level to head height** for the full body width.
- Travel surface information including **textures** and **discontinuities**.
- Detection of **objects bordering the travel path** for shoreline and projection.
- **Distant object and cardinal direction information** for projection of a straight line.
- **Landmark** location and identification information.
- Information enabling **self-familiarization and mental mapping** of an environment.
- Additionally: **ergonomic**, operate with **minimal interface with natural sensory channels**, single unit, **reliable**, user choice of auditory or tactile modalities, **durable**, **easily repairable**, **robust**, low power and **cosmetically accepted**.

## 4.2 Human factors

In Human-Computer Interaction, the designers and engineers try to find efficient ways to facilitate the inter-communication between the user and the system. Tactation is a form of conveying messages to the users which has been studied not as extensively as the visual and sound feedback. The use of sound/ audio feedback is often unwanted because it blocks environmental sounds which are critical for navigation and perception of the environment,

especially for visually impaired users.

In the rest of the paragraph we will discuss the human factors of the design and implementation of our navigation/ mobility prototype with more emphasis on the vibration array, its main component.

#### **4.2.1 Tactile modalities**

The ways to exploit the modalities of the skin's sensors are [41] via:

- **Mechanical energy:** static pressure or vibration.
- Electric field
- Temperature difference: thermal flow.

The most commonly used mechanical energy devices are the pin-based displays and the vibrating motors. In pin-based displays, the tactile pattern is formed by an array of pins controlled usually with piezoelectric actuators. Popular examples are the Braille displays, already in the market such as the Braille Star 80 (Fig. 4-1-a), a Braille keyboard for taking notes, developed by Handy Tech Elektronik GmbH [42] or the SyncBraille (Fig. 4-1-b), a Braille display for reading, developed by GW Micro [43]. Pin-based arrays have a very high refresh rate but they are not suited well for wearable devices, as the vibrating motors do. Electrotactile devices apply small current or electric field that activate the nerves generating the sensation of pressure or vibration but they are more difficult to fabricate and integrate in systems [41].

Taking the above into consideration we decided that vibrating motors are best suited for our project because they are cheap, easy-to-use, easy to integrate into a wearable system and the most widely accepted.

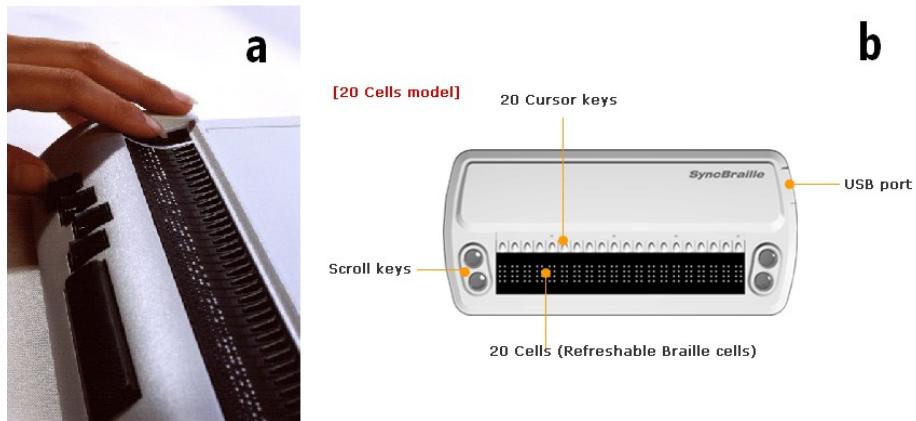


Fig. 4-1. a) Braille Star 80, a Braille keyboard, b) SyncBraille: a Braille display

#### 4.2.2 *Information coding*

The Vibration Array of the Tyflos system is a dynamic tactile display [41] because the states of its tactile elements (vibrating motors) change over time, reflecting environmental changes. The fundamental idea was to create an interface that will be able to display environmental information normally perceived by the eyes; in the field of view of the user. This is information regarding objects and open navigation paths, mostly focusing on their spatial characteristics (location in xyz-space, size and shape). The 2 dimensional arrangement of the vibrating motors corresponds to the two dimensions of the navigation space (width-x and height--y), while the depth-z will be represented from the different frequencies of the motors.

According to Jan van Erp [44], the ways that a tactile stimulus can carry information are:

- By **subjective magnitude**: change in the magnitude/ strength of the stimulus e.g. for electro-tactile, increase the voltage.
- By **frequency**: e.g. different rotating frequencies for a vibrating motor

- By **temporal parameters**: temporal changes e.g. change of frequency over time.
- By **location interface**: spatial distribution of applied stimuli e.g. in the different parts of the user's torso.

Our system utilizes all of the above besides the subjective amplitude; the frequency of the motor is invariable but the motors are driven by variable frequency square pulses, which can be considered as a combination of the frequency and temporal characteristics while the 2 dimensions of the array exploit the coding information by location.

Experiments from psychophysics and human factors studies [44], [45], [46], [47], [48] give some guidelines for the design of vibrotactile systems:

- The (rotating) frequency should be between 50Hz and 600Hz for optimal results [49]. The motors used rotate in  $1200\pm300$ Hz. We consider that minor problem because we don't vary that frequency for delivering information.
- The time between signal must be at least 10ms (10ms pulses or gaps) [50], [51]. The vibrator motors are driven by square pulses with maximum frequency of 10.5Hz, corresponding to  $\sim 90$ ms pulses.
- A spatial acuity of 4cm will suffice for most parts of the upper body [52], [53]. The motors are placed in a 5.5cm distance between each other.

#### *4.2.3 Torso display*

Different studies show that the torso is the most appropriate part of the body to display directional and spatial information. Mapping stimuli directly

*“...to the body coordinates, tactile displays can present spatial information in an intuitive way”* [54]. The spatial and temporal resolution is not as good as other body parts, such as the fingertips or the tongue [41], but “*vibrotactile stimulus on the torso immediately leads to a percept of external direction, an effect called the “tap on the shoulder” principle*” [55]. In addition, torso tactile displays can easily be incorporated in a wearable device and not interfere with other activities of the user (talking, listening, holding).

## 4.3 The stereo vision system

### 4.3.1 Overview

One major part of the Tyflos navigation system is the stereo vision system. As mentioned it is composed of two micro-cameras attached to a pair of conventional dark eye-glasses. The two cameras are connected through USB to the portable computer of the system and they are capable of capturing images and video in real-time from the space in front of the user. Using stereoscopic vision techniques the outputs of the two cameras can produce a depth map of the environment. Depth map is a 2D image but instead the pixels don't represent colors but a distance of the corresponding point in 3D space.

All the steps discussed later are implemented using C++ programming language and the Intel's open source computer vision library OpenCV [56], [57]. The various processing steps slow down the frame rate of the video but still the speed can be considered as real time (approximately 3.5fps), for this project's purpose.

#### 4.3.2 The cameras

For this prototype two inexpensive web cameras Creative Live! Cam Video IM Pro (Fig. 4-2a) where selected sold by Creative Labs Inc. [58]. They are equipped with a 6mm lens with 62° visual angle. Their CCD sensor has a resolution of 300K pixels and it can capture color images of 352×288 pixels and color video with frame rate up-to 30 fps. The focusing is manual and it can be set in the range from 3cm to infinity and they connect to the portable computer through USB. They are attached on a custom thin aluminum plate which is attached behind the glasses' lenses (Fig. 4-2b). The optical axes of the cameras are parallel in a distance of 63mm from each other.

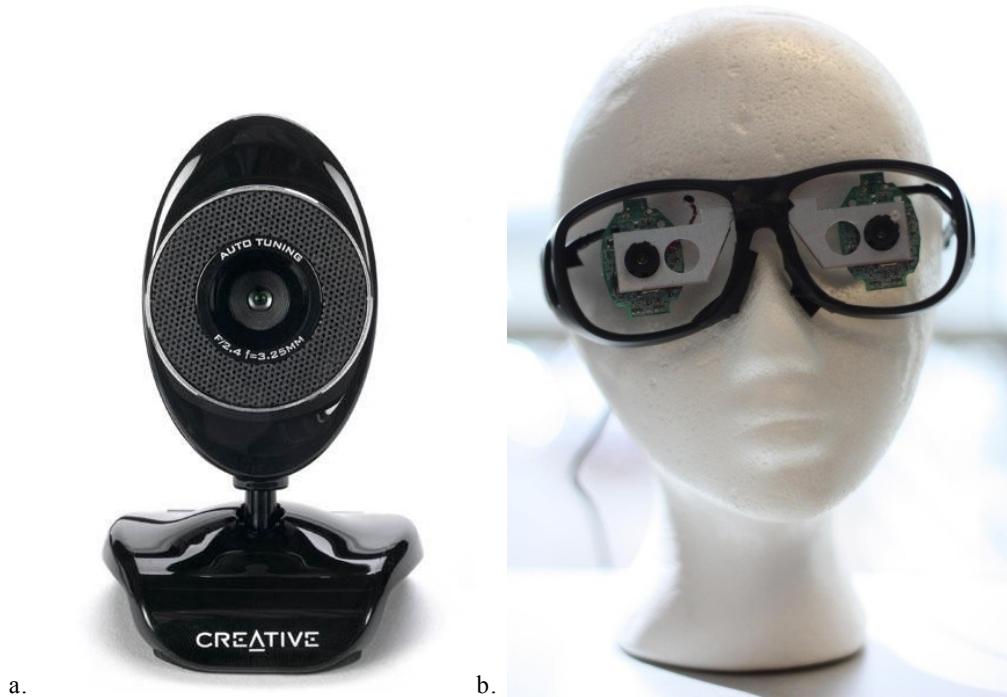


Fig. 4-2. a) The Creative Live! Cam Video IM Pro web-camera by Creative Inc.[58] and b) the two web-cameras without the packaging, attached on the aluminum plate which is screwed on conventional eye glasses.

The reasons why we chose these cameras are:

- **Small and light** so that they can easily be attached on the glasses without disturbing the user and being aesthetically accepted.
- They capture images and videos in relatively **high resolution**.
- Their **low price** (approx. \$50 each)

#### 4.3.3 *Methodologies overview*

Before any stereo vision algorithm is applied to the outputs of the cameras, the cameras need to be calibrated. This means, to find their intrinsic (focal length, principal point, skew coefficient and distortions) and extrinsic parameters (relative rotation and translation of the two cameras) [59].

Camera calibration is an off-line procedure performed only once when the hardware stereo vision system is built. Now that all the calibration parameters are known, the procedure is as follows:

- **Capture** images from cameras
- **Undistort** images: remove any geometric distortions e.g. barrel distortion)
- **Rectify** images: rotate images so that their epipolar lines are aligned horizontally.
- Apply **stereo correspondence** algorithm.

#### 4.3.4 *Depth perception*

The stereo correspondence algorithm is the stereo vision algorithm that will find the pixel's correspondence of one image to another (left camera image to right camera image) and practically produce the depth map. The algorithm

used for creating the depth map is from Birchfield and Tomasi [60], [61] The output of the stereo correspondence algorithm is an image called disparity map. A point in space is represented in different horizontal coordinates in the two stereo images; this difference in number of pixels is called disparity.

The correspondence of a disparity value and the distance in space can be calculated if we know the geometry of the system and the characteristics of the camera. Experimentally this is rather difficult and many manufacturers don't give important characteristics of the camera such as the pixel size in [ $\mu\text{m}$ ]. Thus, we decided to experimentally match disparities with distances by calculating the disparities for objects that were placed in predefined distances.

Fig. 4-3 is the experimental graph for disparity-distance correspondence. We fitted the points with a power function:

$$(Eq. 4-1) \quad D = 45.8 \cdot d^{-1.009}$$

where  $D$  is the distance in [m] and  $d \in \mathbb{N}$  is the disparity in [pixels].

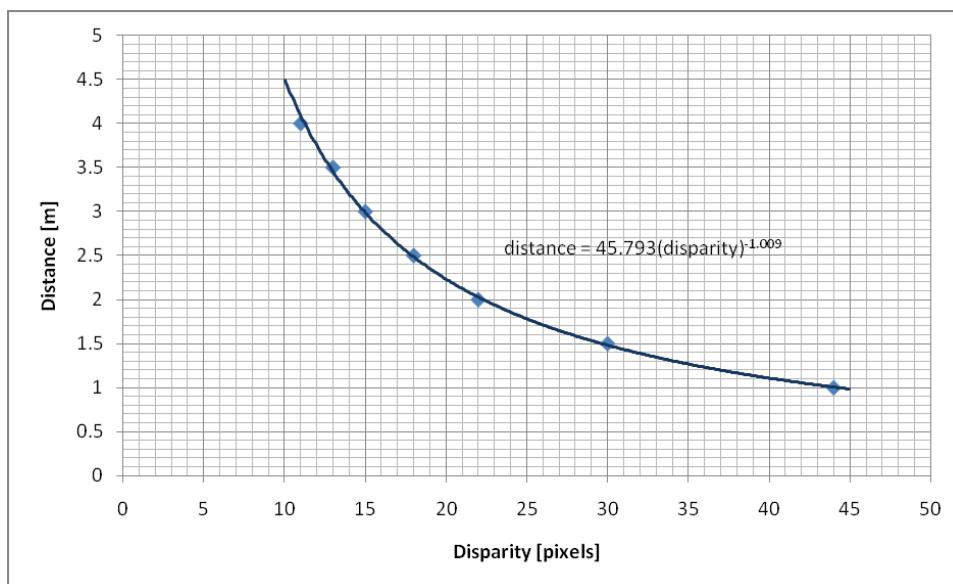


Fig. 4-3. Experimental graph for disparity-distance correspondence.

## 4.4 The 2D Vibration Array

### 4.4.1 Vibrating motors

The market doesn't offer a big variety of small size vibration devices. The most popular are the vibrators used by cell-phone manufacturers and they can be either coin-type or cylindrical. For this design, the vibrator motors (Fig. 4-4) used (C1030B028F) are of-the-shelf, manufactured by Jinlong Machinery [62]. They are coin-type and our market research showed that they are the smallest; 3mm thick, 10mm diameter, which means that they can be easily attach on the clothes of the user. Additionally their vibration level is small enough, so that will not harm the user, but on the other hand it will give the desired information. The power consumption is small which makes it appropriate for a wearable device and also safe for the user.



Fig. 4-4. The JinLong Machinery C1030B028F coin-type vibrator (1cm in diameter and 3mm thick)

### 4.4.2 Vibration module

The easiest way to perceive information from a vibrotactile device (i.e. vibrator motor) is by different frequency pulses because a different rotation speed will not be so easily perceived and is also not possible for most micro-vibrators like these. The circuit chosen to produce these square pulses is a 555

timer in unstable operation (Fig. 4-5). The advantages are that 555 can produce very low frequencies (in the order of Hz or even lower) which are necessary for the application. Additionally the duty cycle of the square pulse is increasing with the frequency. The frequency is controlled by a resistor value so this brings the need to use the digital potentiometer DS1803-100 by Dallas Semiconductors [63].

The DS1803-100 is a dual digital potentiometer with two 256 position potentiometers and maximum resistance at  $100\text{k}\Omega$ . It is controlled by a 2-wire serial interface [64], named SDA and SCL. The 3 address pins allow up to 8 potentiometers to share the same 2-wire interface. The vibration module is consisted of one DS1803 and two 555 timers (Fig. 4-6). That means that each module can drive 2 vibrators.

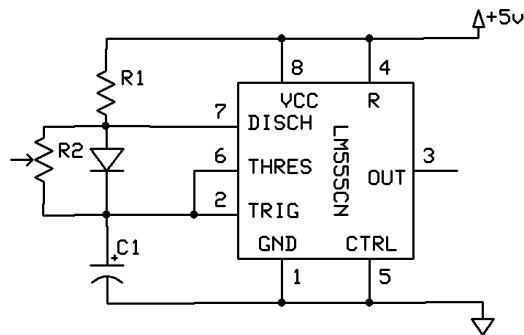


Fig. 4-5. The 555 timer in unstable operation.

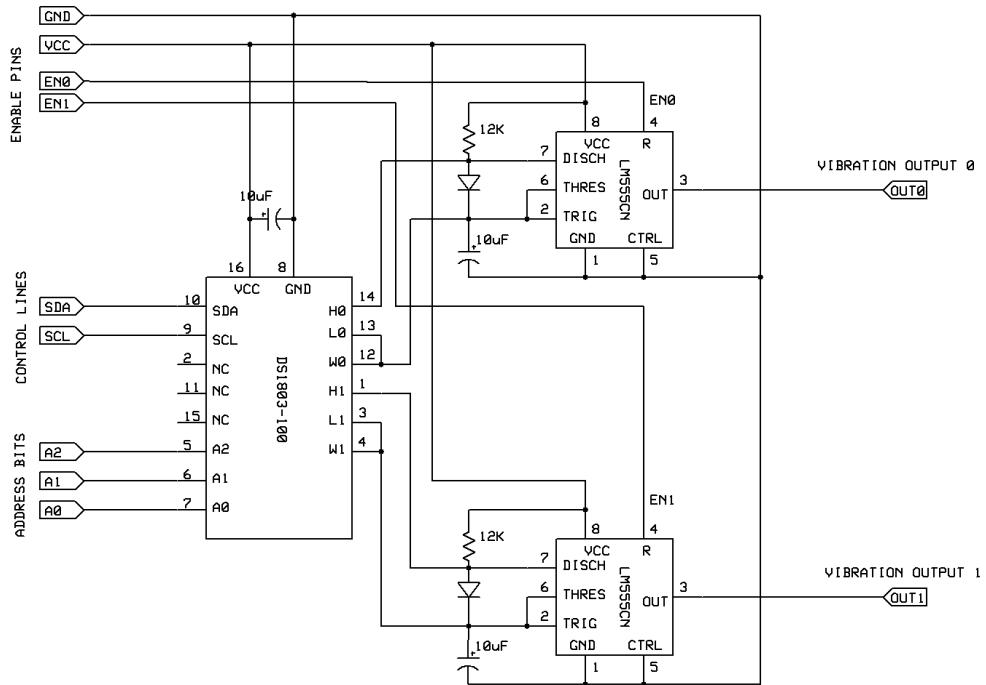


Fig. 4-6. The vibration module schematic: the digital potentiometer DS1803 controls two LM555CN timers that can drive one vibrator each.

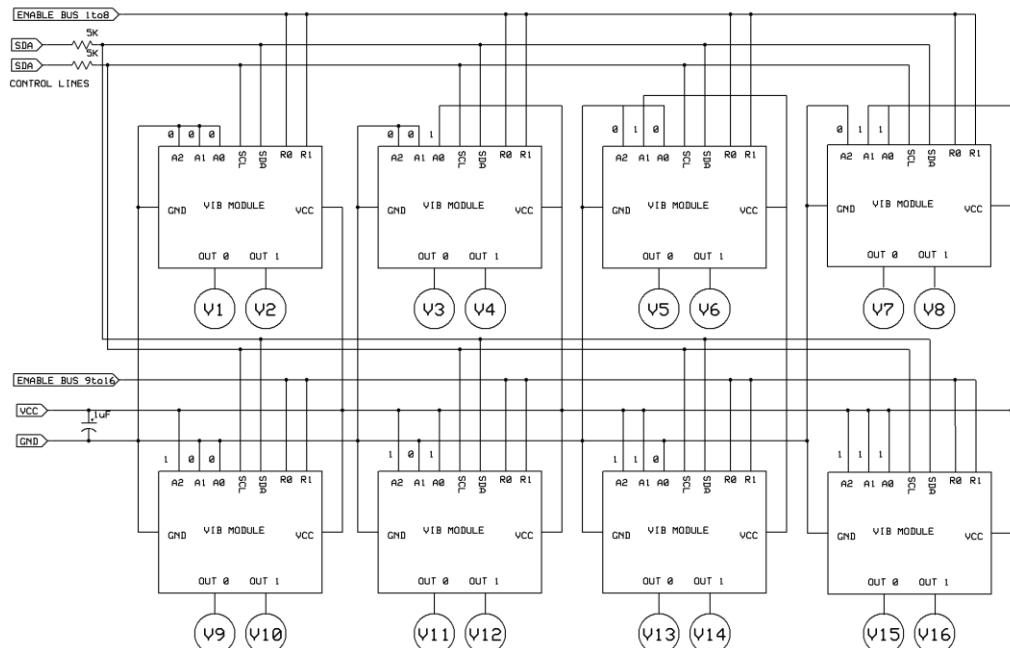


Fig. 4-7. The vibration array schematic: through the 2-wire interface, 8 vibration modules can be controlled which can drive up to 16 vibrators.

#### 4.4.3 The Vibration Array

As mentioned before there are 3 address bits for the DS1803 so the maximum number is 8. For this design the maximum number of vibration modules (8) is selected and they can be controlled under the same 2 wire bus and data transmission protocol. The 8 modules can be interpreted as 16 vibrators, defining a 16 elements array as see in figure 3-4. The size of the array can easily be extended by multiplexing. This possible expansion will be re-evaluated after the first experimental results and feedback from the users. Additionally, the size cannot increase a lot because as discussed earlier in the chapter, there are limiting factors on the minimum distance the vibrating elements can be in order for the user to perceive distinct vibrations.

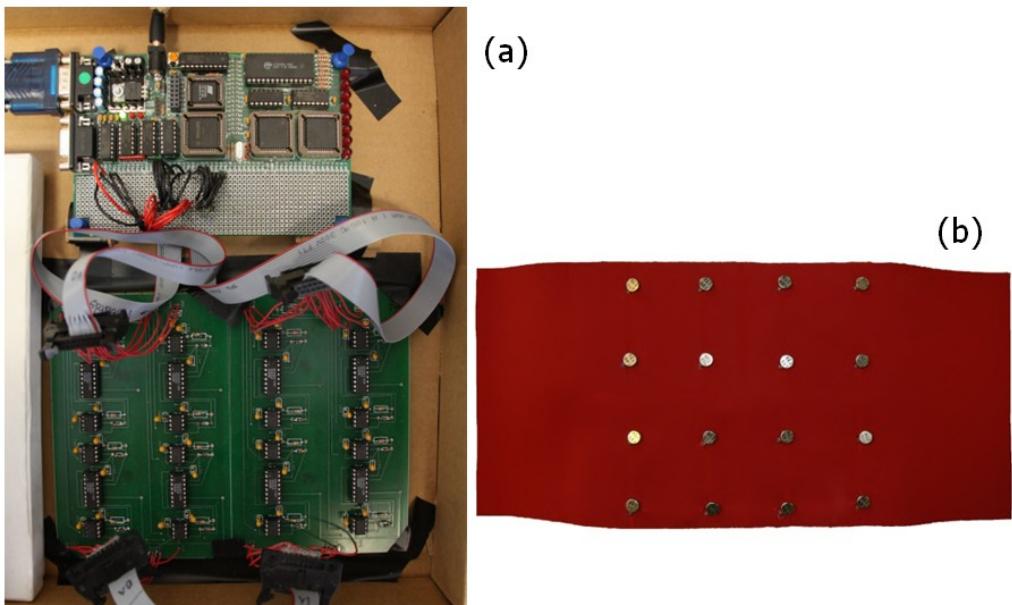


Fig. 4-8. a) The prototype vibration array boards (green boards) connected with the 8051 microcontroller development board (top) and b) the inside of the vibration array vest with the 16 vibrators arranged in a 4×4 manner.

#### 4.4.4 Hierarchical design

The hierarchical approach is a basic characteristic of this design. The vibration module is the basic unit. Combing the vibration modules we have vibration array (practically the circuit that controls the vibration array). The highest level is the operation through the portable computer and the microcontroller (Fig. 4-9).

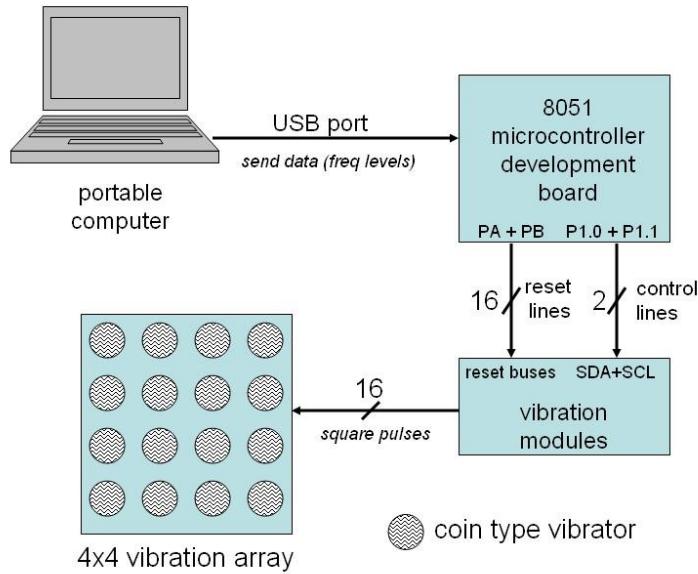


Fig. 4-9. The high-level design of the Tyflos' vibration array.

## 4.5 Interface

A development board equipped with a 8051 microprocessor [65] was used to support the 2-wire interface [64]. The interface was implemented in Assembly code and the communication with the portable computer is done through a USB port with a C/ C++ serial interface. The assembly code is loaded in the flash ROM (which is embedded on the microprocessor's development board) as a start-up program. Vibration frequencies for every vibrator (or in

other words the values of each potentiometer) are set by sending the values from the portable computer to the microcontroller in a serial manner (vibrator-1, vibrator-2,... vibrator-16, vibrator-1... etc). The bandwidth is very fast so there is no issue of delay since all vibrators are set almost simultaneously and thus real-time operation is guaranteed.

## 4.6 Operation

### 4.6.1 Frequency and duty cycle

The vibrators can produce frequencies between 1.25Hz and 10.5Hz (Fig. 4-10). The duty cycle changes also with the frequency (Fig. 4-11), giving a better perception of the vibrations from the user (lower frequencies have smaller duty cycle). For our application the range of frequencies is the desired. This was carefully selected by using the appropriate potentiometer, resistor ( $R_1$ ) and capacitor ( $C_1$ ), having in mind the equations for the 555's unstable operation:

$$(Eq. 4-2) \quad frequency = \frac{1.4}{(R_1 + 2R_2)C_1}$$

$$(Eq. 4-3) \quad duty\ cycle = \frac{R_1}{R_1 + R_2}$$

where  $R_2$  is the potentiometer's resistance.

The duty cycle is very important factor. For higher frequencies we want big duty cycle or else the positive voltage time will not be enough to set the motors in motion. For lower frequencies we want small duty cycles (spike like square pulses) because a big one will result long continuous vibrations, giving the impression to the user of a continuous vibration which can be interpreted as a close obstacle.

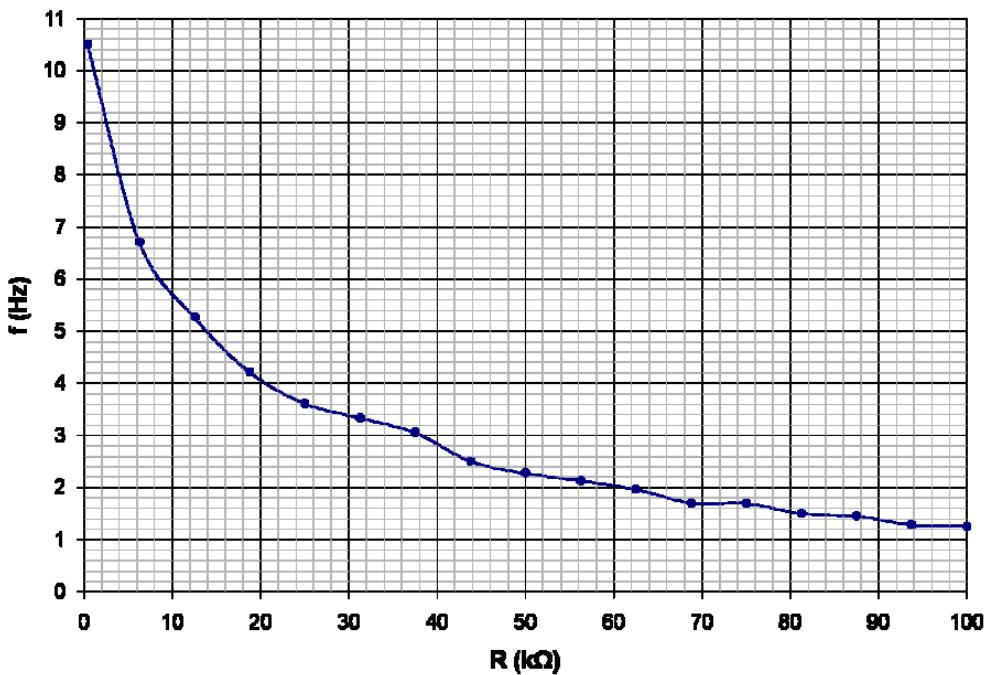


Fig. 4-10. Experimental frequencies of the square pulses produced by the timers and drive the vibrators.

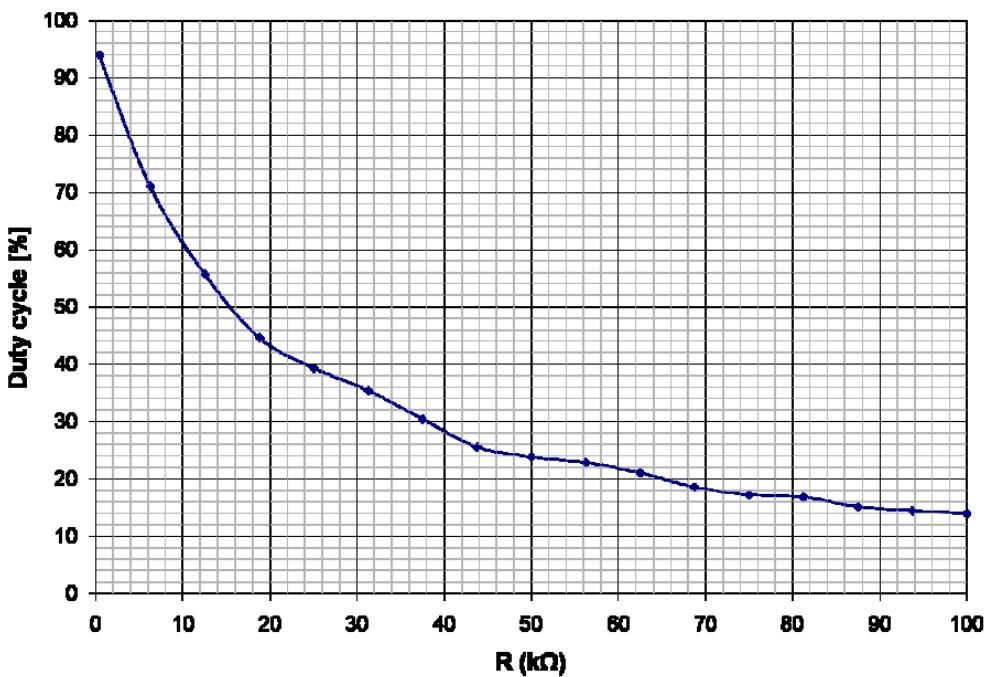


Fig. 4-11. Experimental duty cycles of the square pulses produced by the timers and drive the vibrators.

#### 4.6.2 Calibration

In practice, due to the analog characteristics of the design, the programming for every potentiometer/ timer module with a certain hexadecimal value will not result in the same output frequency, so a calibration of the vibration modules was necessary. Table 4-1 is the calibration table and presents the characteristic calibration curve for one potentiometer.

Table 4-1. Calibration table for the 16 potentiometers (8 vibration modules). e.g. to achieve frequency of 3Hz in vibrating motor #2 I need to send 0x76 at the potentiometer #2

pot#	Frequencies [Hz] and corresponding hex values									
	1.5	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
<b>0</b>	29	69	A8	C8	DB	E8	F0	F7	FA	FD
<b>1</b>	3E	79	B4	D0	E2	ED	F5	F4	FD	FF
<b>2</b>	3B	76	B1	CF	E0	EC	F4	F9	FC	FE
<b>3</b>	2E	6C	AB	CA	DD	E9	F2	F8	FB	FE
<b>4</b>	3A	75	B0	CE	E0	EC	F4	F9	FC	FE
<b>5</b>	33	71	AE	CC	DD	EB	F3	F9	FC	FE
<b>6</b>	37	73	AF	CD	DF	EB	F3	F9	FC	FE
<b>7</b>	38	75	B1	CD	E0	EC	F4	F9	FC	FE
<b>8</b>	37	73	AF	CD	DF	EB	F3	F9	FC	FE
<b>9</b>	38	75	B1	CE	E0	EC	F4	F9	FC	FE
<b>10</b>	3B	77	B1	CF	E0	EC	F4	F9	FC	FE
<b>11</b>	34	71	AE	CD	DF	EB	F3	F9	FC	FE
<b>12</b>	39	75	B0	CE	E0	EC	F3	F9	FC	FE
<b>13</b>	30	6D	AC	CB	DD	EA	F2	F8	FC	FE
<b>14</b>	3B	76	B1	CE	E0	EC	F4	F9	FC	FE
<b>15</b>	3E	7A	B4	D0	E0	EE	F5	F4	FD	FF

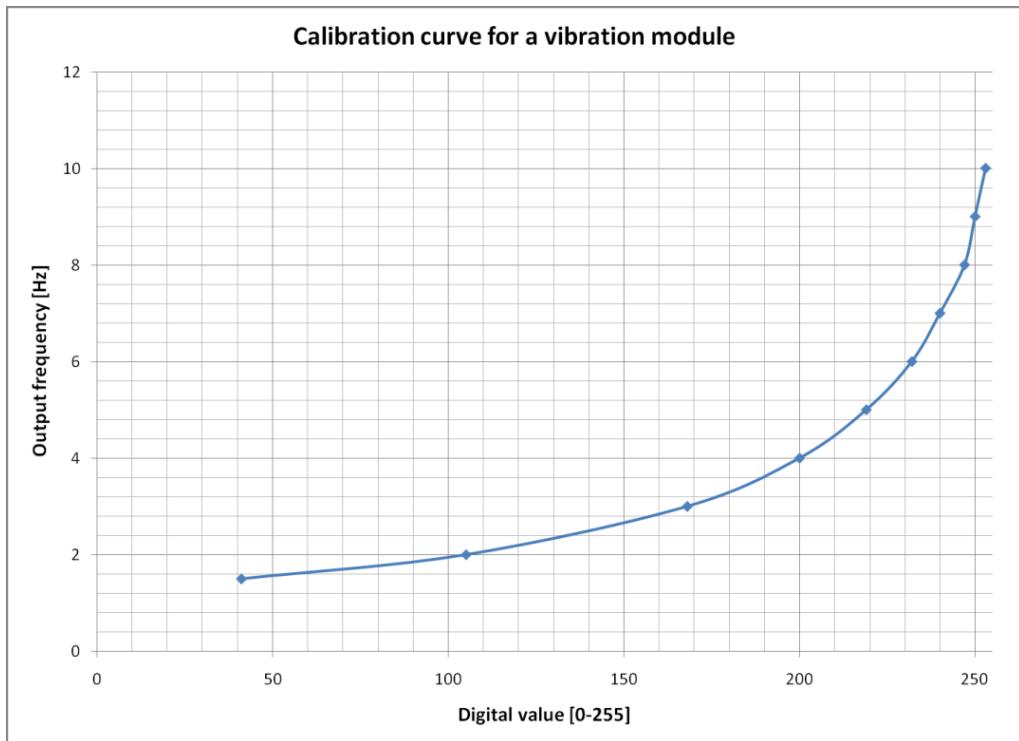


Fig. 4-12. Typical calibration curve (hex value – output frequency) for a vibration module.

# Chapter 5: Vibration Array Language

As mentioned earlier, the 2D vibration array is one of the important components of the Tyflos navigation system. The VAL language will assist us to associate high resolution visual patterns frequently appeared in the corridors of buildings, such as chairs, tables, books, humans, boxes, etc, with their low resolution representations. This association will be used as a training set for the visually impaired individuals and later as the working database during their navigation inside of buildings. Thus, in the following section we briefly provide the formal model of the language and a few simple examples explaining its 3D representation of patterns. Thus, the following sections are dedicated to the modeling, implementation and simulation of the Vibration Array.

## 5.1 Formal model

The modeling of the Vibration Array is a two-dimensional representation of a one-dimensional formal language [24], [66], [67], with two dimensional characteristics.

### 5.1.1 Symbols

Definitions: An array  $P$  is defined as:

$$(Eq. 5-1) \quad P = \{p_{ij} / \in [1, m] \times [1, n]\}$$

with  $m, n \in Z^+$ . We will define a context-free formal language consisting of one type of letters (symbols). We also define special symbols  $A$  with length  $l$ , which consists of  $l$  consecutive elements of the array in the same row or column e.g.

for  $l$ ,  $p_{23}, p_{33}$ . Thus,  $A$  can be defined as:

$$(Eq. 5-2) \quad A_l^{(x_0, y_0)} = \{p_{iy_0} / i \in [x_0, x_0 + l - 1]\}$$

where  $(x_0, y_0) \in [1, n] \times [1, m]$  are coordinates of the starting element,  $l \in [1, n - x_0 + 1]$  the length of the symbol (number of elements),  $p_{iy_0} = V_{i-x_0+1}^l$ ,  $i \in [x_0, x_0 + l - 1]$  and  $V^l$  is a  $1 \times l$  column array where  $v_i \in [1, V - 1]$ ,  $1 < V \in Z^+$ .

Propositions: Assume two symbols  $A = A_l^{(x_0, y_0)}$  and  $A' = A'_l^{(x'_0, y'_0)}$ . Then:  $A = A'$  if  $x_0 = x'_0$ ,  $y = y'_0$  and  $l = l'$ .  $a$  overlaps with  $a'$  if they have at least one element of the array in common. We can obtain the set of the special symbols:

$$(Eq. 5-3) \quad \Sigma = \{a_i / i \in [1, r]\}$$

where  $r = nm(m + 1)/2$  is the maximum number of symbols that can be represented in the array.

### 5.1.2 Grammar

Definition: A grammar is defined as:

$$(Eq. 5-4) \quad G = (V_N, V_T, P, S)$$

where:

$$(Eq. 5-5) \quad V_N = \{S, T, A\}$$

is the set of non-terminal symbols

$$(Eq. 5-6) \quad V_T = \Sigma \cup \{i / i \in Z^+\} \cup \{\#\}$$

is the set of terminal symbols. And the symbol  $\#$  represents the unique operator of the language and

$$(Eq. 5-7) \quad P = \{S \rightarrow T, S \rightarrow S\#T, T \rightarrow A, A \rightarrow a_1 | a_2 | \dots | a_m\}$$

is the set of production rules and  $S$  is the start symbol of grammar  $G$ .

The grammar is context-free [67] because all the production rules follow the rule  $X \rightarrow Y$  where  $X$  is a non-terminal symbol and  $Y$  is either a terminal or a non-terminal symbol. We name this grammar VAG (Vibration Array Grammar).

### 5.1.3 The VAL Formal Model

The words that VAG can produce are of the form:  $a_1 \# a_2 \# \dots \# a_t$ , where  $a_i \in \Sigma$ ,  $i \in [1, t]$  and  $t \in Z^+$  is the length of the word. Thus, the VAL language can be defined:

$$(Eq. 5-8) \quad L(G) = \left\{ \begin{array}{c} a_1 \# a_2 \# \dots \# a_t, \\ \text{with } a_i \in \Sigma, i \in [1, t], t \in Z^+ \end{array} \right\}$$

## 5.2 Simulation

A computer simulation of the language is presented. We will use a  $32 \times 32$  vibration array. The vibration frequency will be represented in the z-axis. There are 4 vibration levels from 0 to 3 shown in Table 5-1 and a color is assigned for visualization purposes (Table 5-2).

Table 5-1. Similar with Table 3-2: Vibration levels, vibration frequencies and object distances correspondence.

Vibration level	Frequency [Hz]	Distance range [m]	Vibration strength
0	0	$[3, \infty)$	None
1	2	$[2, 3)$	Low
2	4	$[1, 2)$	Medium
3	8	$(0, 1)$	High

Table 5-2. Color representations of the 4 vibration levels.

Vibration level	RGB representation	Grayscale representation
0	Cyan	Black
1	Yellow	Dark gray
2	Red	Light gray
3	Burgundy	White

Seven simulated cases will be shown covering different possible scenarios during navigation. For the object oriented software implementation in C++, letters (as well as words) are objects of the form:  $a = (x, y, length, V[ ])$  where  $x, y$  are the coordinates of the first element in the array (assume element  $a_{xy}$  being in the top left corner which will represent distances of obstacles being on the left top corner of the user's,  $length$  is the length of the letter and  $V[ ]$  is an array of length  $length$  and holds the vibration level for every element. The vibration array can hold a word  $W$  which is a combination of symbols. For better visualization the different vibration levels have different color (see table 4-1).

- Vertical obstacle, e.g. standing/ walking person, (fig. 4-1a).

$$w_1 = a_1 \# a_2 \# a_3, \quad \text{where} \quad a_1 = (5, 12, 25, V[ ]), \quad A_2 = (5, 13, 25, V[ ]), \\ a_1 = (5, 14, 25, V[ ]) \text{ and } V = (\underbrace{3 \ 3 \dots 3}_{\#25})$$

- Two vertical obstacles, e.g. 2 persons, (fig. 4-1b).

$$w_2 = w_1 \# a_4 \# a, \quad \text{where} \quad a_4 = (10, 25, 17, V[ ]), \quad a_4 = (10, 26, 17, V[ ]) \text{ and} \\ V = (\underbrace{1 \ 1 \dots 1}_{\#17})$$

- Side obstacles e.g. corridor or open door, (fig. 4-1c).

$$w_1 = a_1 \# a_2 \# a_3 \# a_4, \quad \text{where} \quad a_1 = (0, 0, 32, V[ ]), \quad a_2 = (0, 1, 32, V[ ]), \\ a_3 = (0, 30, 32, V[ ]), \quad a_4 = (0, 31, 32, V[ ]) \text{ and } V = (\underbrace{2 \ 2 \dots 2}_{\#32})$$

- Side and vertical obstacles, e.g. person in a corridor, (fig. 4-1d).

$w_4 = w_3 \# a_5 \# a_6$ , where  $a_5 = (10, 25, 17, V[ ])$ ,  $a_6 = (10, 26, 17, V[ ])$  and

$$V = (\underbrace{1 1 \dots 1}_{\#17})$$

- Overhanging obstacle on the right, (fig. 4-1e).

$w_5 = a_1 \# a_2 \# \dots \# a_{10}$ , where  $a_i = (5, 21 + i, 5, V[ ])$  where  $i = 1, 2, \dots, 10$

$$\text{and } V = (\underbrace{2 2 \dots 2}_{\#5})$$

- Ground obstacle on the right e.g. box on the floor, (fig. 4-1f).

$w_5 = a_1 \# a_2 \# \dots \# a_8$ , where  $a_i = (27, 19 + i, 5, V[ ])$ , where  $i = 1, 2, \dots, 8$  and

$$V = (\underbrace{2 2 \dots 2}_{\#5})$$

- Complex obstacle e.g. a workstation in an office, (fig. 4-1g).

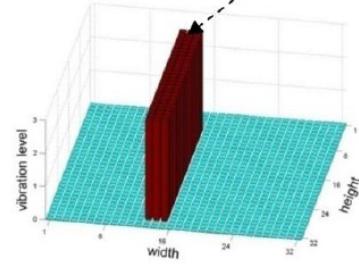
$w_7 = a_1 \# a_2 \# \dots \# a_{13}$ , where  $a_i = (6, 5 + i, 20, V[ ])$ , where  $i = 1, 2, \dots, 13$

$$\text{and } V = (\underbrace{3 3 \dots 3}_{\#6} \underbrace{1 1 \dots 1}_{\#6} \underbrace{2 2 \dots 2}_{\#8})$$

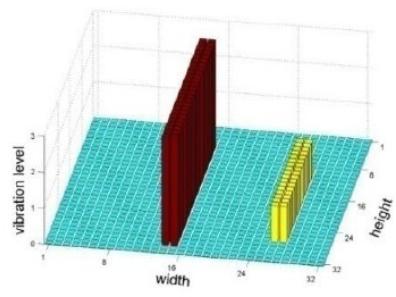
The characteristic of the VAL is that it can represent any possible obstacle (or combination of obstacles) in various distances. The array holds a word that is combination of symbols (row-type). The important characteristic is that every symbol can hold more than one vibration level through the array  $V//$  which is more evident in case f.

Perceived view of the 3D space

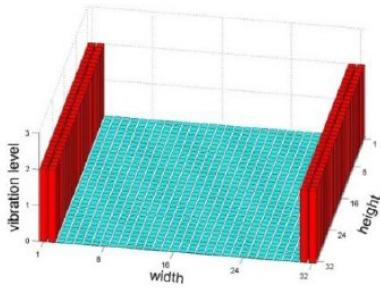
a.



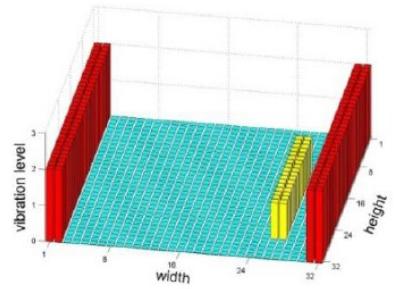
b.



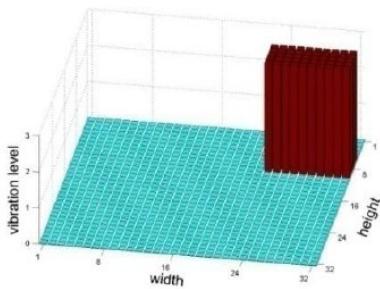
c.



d.



e.



f.

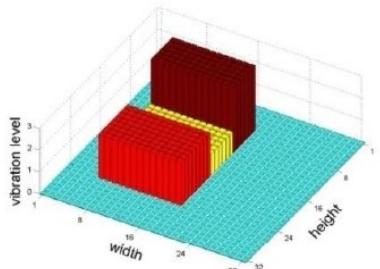


Fig. 5-1. The seven simulated scenarios illustrating the use of the Vibration Array Language (VAL).

## **Chapter 6: Video Stabilization & Motion Detection**

Most of the electronic travel aids (ETAs) process visual information captured from the surrounding space, one of the important challenges that face is to efficiently detect and track motion from the surrounding environment and effectively represent it to visual impaired. Thus, this challenge is directly related to image/ video understanding research, especially to motion analysis, and to haptic technologies as well. Firstly we start 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. Several topics of this field include motion segmentation, object detection and tracking, stereo imaging, 3-D motion and image and video registration. The analysis of dynamic scene is very popular with applications to surveillance, target tracking and navigation systems. In this work a motion analysis scheme is proposed as an assistive vision tool that can be useful for assisting the navigation of visual impaired people [68], [69]. The motion analysis and segmentation approaches may be generally divided into the categories of image differencing, optical flow-based methods and spatiotemporal methods.

The image differencing or subtraction methods appeared first in the literature and include methods such as simple subtraction or more elaborate schemes like background registration [70]. On the other hand, the optical flow

represents an estimate of the motion field of a sequence of image and its computation is based on the luminance conservation constraint. The motion segmentation is completed by detecting the optical flow discontinuities. Several optical flow implementations have been proposed [71]. However it cannot be reliably estimated in the borders of the moving objects and in smoothly textured areas. These pitfalls have led to the introduction of approaches that estimate the global motion and its spatial support simultaneously. Another group of methods is characterized as spatiotemporal [72]. According to these, the image areas are grouped by comparing their spatial and temporal properties.

## 6.1 The Methodology

The proposed scheme [73] is employed for detection and segmentation of moving objects in the image sequence. This is accomplished by a video stabilization algorithm that is followed by spatiotemporal diffusion kernel-based motion detection and the moving objects are segmented using the watershed algorithm.

### 6.1.1 *The Frame Stabilization*

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

In our application since the camera is part of a wearable device, it follows the translational lateral motion that is caused by the human gait. This step is carried out in order to keep the field of view of the camera constant in order to facilitate the subsequent motion analysis of the scene. This problem is re-

solved here, by transforming the original image plane to horizontal and vertical projections and subsequently estimating the motion in these 1-D domains [74], [75].

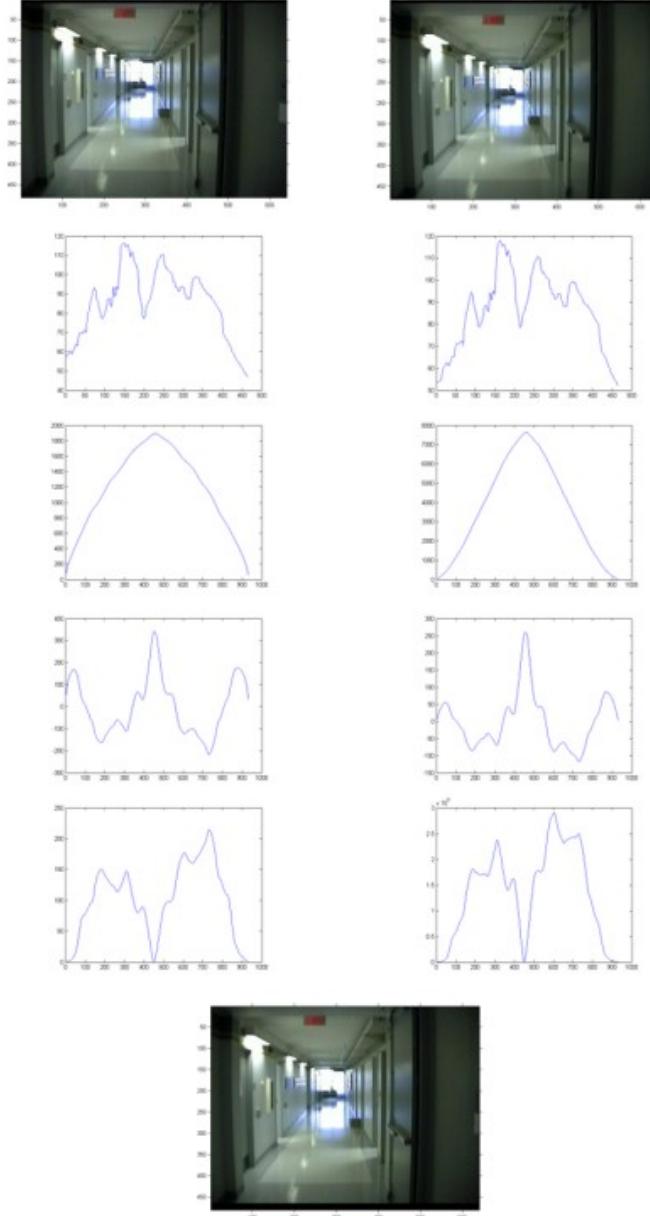


Fig. 6-1. First row: two consecutive frames in the test sequence. Second row: vertical projections of input frames, Third row: normalized (left) and not normalized (right) correlation. Fourth row: normalized (left) and not normalized covariance. Fifth row: normalized (left) and not normalized sum of squared error measure. Sixth row: stabilized frame using sum of squared error.

The projections are estimated by the Radon transform. Assuming a function  $f(x,y)$  the Radon transform is defined in the continuous case as follows:

$$(Eq. 6-1) \quad T(\rho, \theta) = R_\theta[f(x, y)] = \iint f(x, y) \cdot \delta(\rho - x\cos\theta, \rho - y\sin\theta) dx dy$$

where  $\rho$  and  $\theta$  are spatial variables in the radial coordinates;  $\rho$  is the radius and  $\theta$  the angle; and  $\delta$  is the Dirac function defined in (Eq. 6-2):

$$(Eq. 6-2) \quad \delta(i, j) = 1, i = j \text{ and } \delta(i, j) = 0, i \neq j$$

Here are estimated the horizontal and vertical projections and for the discrete image values  $I(x, y)$  we have:

$$(Eq. 6-3) \quad T_{0^\circ}(y) = \sum_x^{N_r} I(x, y)$$

$$(Eq. 6-4) \quad T_{90^\circ}(x) = \sum_y^{N_c} I(x, y)$$

The result of this process is the decomposition of the initial 2-D signal into two 1-D signal. 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 one-dimensional signals  $I$  and  $J$  these measures are estimated as follows:

$$(Eq. 6-5) \quad Corr(u) = \sum_x I(x) \cdot J(x - u)$$

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

$$(Eq. 6-7) \quad SSD(u) = \sum_x [(I(x) - J(x - u))^2]$$

An example of the stabilization process is displayed in Fig. 6-1, where the displacements of a wearable camera are caused by the natural body movement. 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. Correlation and Covariance are normalized by the sum of the signal energy (squared intensity). The Sum of Squared Error 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. After several experiments it was concluded that the scaled Sum of Squared Differences produces better results.

### ***6.1.2 Estimation of Motion Activity***

A spatiotemporal anisotropic diffusion method is first applied that uses the information of the current, previous and next frames in the sequence. This method is based on the anisotropic diffusion theory [76] 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 to that a kernel density estimation approach is proposed to address the presence of outliers and disconnected areas, and produce more robust motion detection results.

#### ***6.1.2.1 Perona's anisotropic diffusion***

Perona has proposed an anisotropic diffusion filtering for scalar images that avoids 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:

$$(Eq. 6-8) \quad \delta_t \cdot I = \operatorname{div}(g(|\nabla I|^2)\nabla I)$$

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

In this work an implementation that uses finite differences, originally proposed in [76], 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:

$$(Eq. 6-9) \quad g(\nabla I) = 1 / \left( 1 + \left( \frac{\|\nabla I\|}{K} \right)^2 \right)$$

where  $K$  denotes the conductance parameter for which a value from 4.5 to 5 is regularly selected for the purpose of segmentation. Equation (Eq. 6-8) can be equivalently written as:

$$(Eq. 6-10) \quad \frac{\partial I(x,y,s)}{\partial s} = c(x,y,s) \cdot \Delta I(x,y,s) + \nabla c(x,y,s) \cdot \nabla I(x,y,s)$$

where  $\Delta$  is the Laplacian operator and  $c(x,y,s)$  is a constant. The discretization of this equation leads to the following PDE for the proof of 2-D case [73]:

$$(Eq. 6-11) \quad I_{i,j}^{s+1} = I_{i,j}^s + \lambda \cdot [c(N) + c(S) \cdot S + c(E) \cdot E + c(W) \cdot W]_{i,j}^t$$

$$(Eq. 6-12) \quad c(k) = g(\|k\|), k = N, S, E, W$$

Here  $N, S, E, W$  denote the difference between the central pixel and its north, south, east and west nearest neighbor respectively. Furthermore  $c(k)$  equals to the value of the diffusivity function for each direction.

#### 6.1.2.2 Estimation of pixel motion activity using spatiotemporal diffusion

In this section is explained the reasoning behind the proposed motion activity measure. This measure is based on the fact that after the spatiotemporal diffusion process the background areas are diffused more than the moving areas.

The main idea for the spatiotemporal non-linear filter is to extend the non-linear diffusion equation using the additional variable of time. The PDE equation now becomes:

$$(Eq. 6-13) \quad I_{i,j,t}^{s+1} = I_{i,j,t}^s + \lambda_s \cdot [c(N) + c(S) \cdot S + c(E) \cdot E + c(W) \cdot W]_{i,j,t}^s +$$

$$\lambda_t [c(PF) \cdot PF + c(NF) \cdot NF]_{i,j,t}^s$$

$$c(k) = g(\|k\|), k = N, S, E, W, PF, NF$$

This equation seems to be a straightforward extension of (Eq. 6-11) 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  $\lambda_s$  and  $\lambda_t$ , which control the amount of spatial and temporal diffusion respectively. In this application it was experimentally found that the temporal diffusion must be significantly bigger than the spatial for motion diffusion.

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

$$(Eq. 6-14) \quad I_{i,j,t}^{s_1+1} = I_{i,j,t}^{s_1} + \Delta I_{i,j,t}^{s_1}$$

Similarly to (Eq. 6-14), for  $s = s_1 + 2$  we have:

$$(Eq. 6-15) \quad I_{i,j,t}^{s_1+2} = I_{i,j,t}^{s_1+1} + \Delta I_{i,j,t}^{s_1+1}$$

By substituting (Eq. 6-14) into (Eq. 6-15) we produce:

$$(Eq. 6-16) \quad I_{i,j,t}^{s_1+2} = I_{i,j,t}^{s_1} + \Delta I_{i,j,t}^{s_1} + \Delta I_{i,j,t}^{s_1+1}$$

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

$$(Eq. 6-17) \quad I_{i,j,t}^{s_2} = I_{i,j,t}^{s_1} + \sum_{i=0}^{s_2-s_1} \Delta I_{i,j,t}^{s_1+i}$$

$$(Eq. 6-18) \quad \sum \Delta I_{i,j,t} = I_{i,j,t}^{s_2} - I_{i,j,t}^{s_1} = \sum_{i=0}^{s_2-s_1} \Delta I_{i,j,t}^{s_1+i}$$

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

$$(Eq. 6-19) \quad PMA_{i,j,t} = 1 - \frac{\sum \Delta I_{i,j,t}}{\max(\sum \Delta I_{i,j,t}^2)}$$

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

#### *6.1.2.3 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 this process is subject to several temporal variations that do not correspond to motion.

In this work this pitfall is resolved by means of non-parametric density estimation using Parzen kernels [77]. 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 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:

$$(Eq. 6-20) \quad f_h(x) = \frac{1}{N \cdot h^p} \sum_{i=1}^N K\left(\frac{x-x_i}{h}\right)$$

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

$$(Eq. 6-21) \quad K(x) = (2\pi)^{-p/2} \cdot \exp(-x^2/2)$$

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 es-

timate 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 set to 20 (Fig. 6-2).

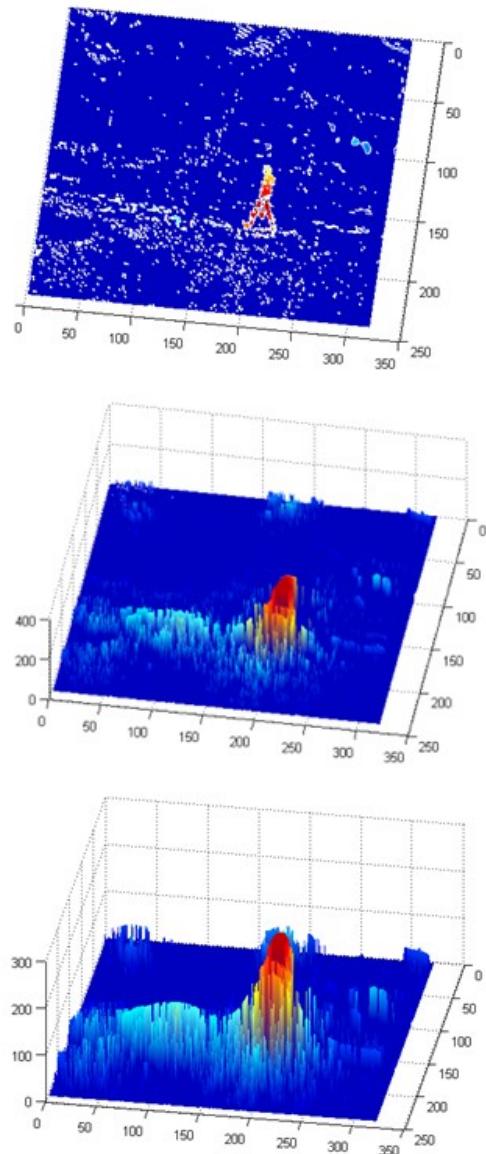


Fig. 6-2. The estimated density of motion activity map using Parzen kernels.

### 6.1.3 *The 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 spatiotemporal 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 is considered here. The watershed algorithm segments a grayscale or color image into different 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 appeared since then, such as iterative, sequential, arrowing, flow line oriented and flooding approaches. A flooding implementation is employed here similar to [68].

This operation can be divided into two stages: 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 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 the probability density values are first inverted.

The lowest minima are subsequently selected and flooded to form the moving areas.

## 6.2 Simple experimental results

Fig. 6-3 shows the results of the motion detection methodology. The three indoor examples are possible navigation examples. The segmented regions that correspond to the moving people are detected successfully. They do not have very fine detail but for the application in a wearable navigation device the accuracy and robustness is good since, explained in the next chapter, the moving areas are areas of interest and the modified pyramidal resolution reduction algorithms are sensitive in keeping areas of interest as the high resolution of the cameras is reduced to match the low resolution of the vibration array.



Fig. 6-3. Three real indoor navigation scenarios. For every scenario a, b and c are the 3 consecutive frames used by the methodology. In d, white areas are the detected pixels that correspond to moving area.

# **Chapter 7: High-to-Low Resolution Representations**

## **7.1 The Resolution of the Vibration Array**

The cameras used in the current Tyflos prototype can capture images of  $640 \times 480$  pixels. In order to reduce the computational cost we capture in  $352 \times 264$ . On the other hand, the resolution of a vibration array undergoes many restrictions. Hardware restrictions are: vibrator motors' dimensions (not less than 1cm, 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  $32 \times 32$ . Our current prototype vibration array is consisted of  $4 \times 4$  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 makes us believe that a  $4 \times 4$  will be a viable solution. The feedback during the experimental phase will direct us for any changes.

## **7.2 Navigation Issues**

During orientation and navigation, the visual environmental information that is normally perceived by the eyes is very complex. The perception of

all this information without the use of sight is very difficult, so a categorization of the visual information is necessary.

According to Welsch and Balsch [4], “*obstacle is defined as an architectural or environmental obstruction in the path travel than can be detected and negotiated with standard long can techniques*” and travel hazard when it cannot. Travel hazards include: public telephone booths, guy wires, wall fixtures (e.g. wall mounted ashtrays or fountains), stairs or escalators (e.g. walking under stairs), planters and shrubbery, street signs, display windows etc. Below there is a list of common objects/ obstacles, or more generally, formations that appear during navigation:

- a. People
- b. Furniture e.g. chair, table
- c. Drop offs: stairs
- d. Over hanging obstacles e.g. signs, tree branches, phone booths
- e. Reclining obstacles: guy wires
- f. Open paths: open door, corridor

For our case, we assume that during navigation there are four basic elements that can appear on the scene:

- **Obstacle (o):** the user needs to avoid for safe navigation.
- **Object (b):** the user may or may not have interest to retrieve.
- **Corridor (c):** informs the user about an open path that can be followed to safely continue his/ her navigation [82].
- **Background (B):** all other elements, not important during navigation.

The distinction of whether an element is an obstacle or object depends on the context, the user's preference and functionality of the system. For example, a system has a face recognition functionality expects from the user to define if a person will be considered as an obstacle or an object; in a navigation mode, the person would be probably be considered as a obstacle while in a people recognition mode, as an object. Generally, the user wants to avoid obstacles and be aware of objects and corridors that may be useful for during his/ her navigation. Thus an image  $I$  can be considered as a set of these elements creating a navigation-based, mathematical definition.

$$(Eq. 7-1) \quad I(x, y) \rightarrow \{(o, b, c, B)\}$$

or as a union:

$$(Eq. 7-2) \quad I(x, y) = \bigcup (o_i, b_j, c_k, B)$$

where  $i, j, k \in \mathbb{N}$  and  $\mathbb{N}$  is the set of natural numbers.

Every element can be defined using specific characteristics (the background's characteristics are not important at this point):

- For the obstacle:  $o \rightarrow (l, d, s, sh, X)$
- For the objects:  $b \rightarrow (l, d, s, sh, X)$
- For the corridor:  $c \rightarrow (l, d, s, w, X)$

where  $l$  is the location on the 2D image,  $d$  is the distance from the user,  $s$  is the size,  $sh$  is the shape and  $w$  is the width. The  $X$  is an open spot for the future declaration of a characteristic. They can be flags that categorize and/ or give more or less importance of an object. Additionally they can be user defined. For example, a flag can characterize an object as a person. The person can be identified

fied and depending on the user's setting a high or low importance can be given (low importance for a stranger, high importance for a familiar person).

### 7.3 Standard high-to-low algorithms

In order to reduce the resolution of the captured frames, we will be using 4 standard high-to-low resolution algorithms [56], [57], [78] with emphasis on the pyramidal reduction algorithms because this is where the special rules will be applied:

- Bi-linear interpolation
- Nearest neighbor interpolation
- Gaussian pyramid
- **Pyramidal reduction** ( $2 \times 2$  and  $3 \times 3$  kernel)

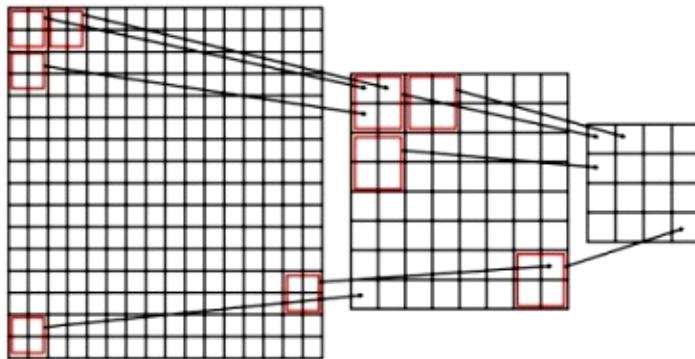


Fig. 7-1. Pyramidal reduction using a  $2 \times 2$  kernel. Every  $2 \times 2$  pixels kernel is reduced to 1 pixel as shown.

**Bi-linear interpolation** is inverse linear interpolation in 2 dimensions.

In the **nearest neighbor interpolation** the new pixel is defined by the nearest pixel value [57]. In **Gaussian pyramid** the image is first convolved with a  $5 \times 5$  Gaussian filter and the downsampled by rejecting even columns and rows. **Pyramidal reduction** is a basic high-to-low resolution algorithm for information

reduction. Recursively,  $k \times k$  pixels (kernel) from the initial matrix/ image are mapped to 1 in the new matrix using a specific function (maximum, average, etc). In Fig. 7-1 there is an example where  $16 \times 16$  image gets to  $4 \times 4$  in 2 iterations, using a  $2 \times 2$  kernel.

The implementation and experimentation of these algorithms was done using C/ C++ programming language and the Intel's open source computer vision library (OpenCV) [56], [57].

## 7.4 Criteria of interest

At this point we use a set of criteria of importance in order to “guide” the high-to-low reduction in a way that important pieces of visual information will “survive” at a reasonable level, so that, the blind user can better understand the 3D surroundings via the vibrating sensations. The criteria are chosen to serve navigation purposes:

- C1: **Distances** of objects from the user.
- C2: **Objects of certain interest** during navigation (e.g. people, chair)
- C3: Open navigation path
- C4: Moving objects

The disadvantage of the standard low resolution methodologies is that they do not take into account any navigational issues and criteria of interest i.e. obstacles, objects of interest, open/ safe navigation paths, motion. Pyramidal reduction is convenient for making modifications. In our experiment, two high-to-low representations are implemented. The first will be called 3D representation and the second, visual representation.

## 7.5 Visual representation

The image captured by the two cameras is the so called *fused-image* (Fig. 7-2a). The *image-of-interests* (Fig. 7-2b) is an image that its pixels correspond one by one with the fused-image. A white pixel in the interests-pixel indicates that the corresponding pixel in fused-image is of-interest i.e. belongs to an object of interest (e.g. familiar person, a Braille sign, any obstacle, moving objects etc). The black pixel corresponds to pixel without interest.

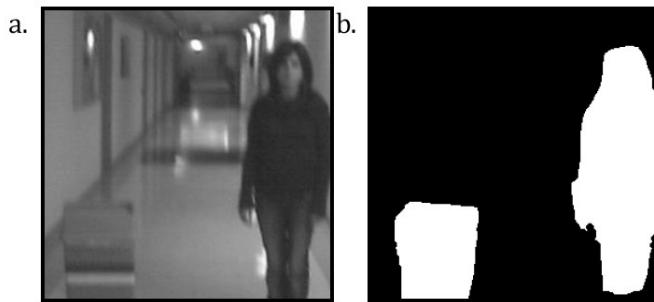


Fig. 7-2. Figure 7-2. a) Fused 256×256 image, presenting a low height close distance obstacle on the left and a walking person in medium distance on the right and b) Image-of-interests representing 2 areas of interest with white color (obstacle and person).

### 7.5.2 Without criteria

The main goal of this paragraph is that the user has to go for collision with the objects of interest. In this section we use The 5 algorithms mentioned before and reduce the resolution of the fused image and the image of interests to 32×32 (Fig. 7-4). The image of interests is complementary to see the behavior of the interest regions (without criteria). From the results in Fig. 7-4, we see that the bi-linear and nearest neighbor interpolations presents better visual outcomes.

### 7.5.3 With criteria

The main idea behind the image reduction with criteria of interest is that the user has to go for a collision path towards the objects of interest, such as a person, a chair, a table, etc. Thus, it is important if we have a methodology to reduce the image information by maintaining as much as possible information related to regions of special interest. Thus, an obstacle in a close distance is an object of special interest (criterion C2); also a walking person is a moving object (C4). In case that C2 and/ or C4 are true, we perform a pyramidal reduction ( $2 \times 2$  and  $3 \times 3$  kernel) using the following rule:

Rule: “*If there is at least one pixel of interest out of the 4 then the next level will be set as “of interest”*” (Fig. 7-3).

The fused image pixel will be the average of the pixels of the corresponding fused pixels. e.g. if only interests pixels (0,0) and (0,1) are of interest then, the next level interest pixel will be “of interest” and the new level fused pixel will be the average of the fused pixels (0,0) and (0,1). (Note: if no pixel of interest is found then the next level pixel of interest will be “of no interest” and the next level fused pixel will be as the average of all the 4.” The processing steps are:

- Initially we take the high fused images.
- Then we apply the rule on set of pixels taking into account the importance of each pixel in order to “save” important pieces of information during the reduction from the higher level to the next lower one.

The results from these steps are presented in figure 7-4 using the pyramidal reduction algorithms. The use of the pyramidal reduction has nothing to

do with the performance of the algorithm for its selection at the end, but only to show how information is reduced by using criteria of interest.

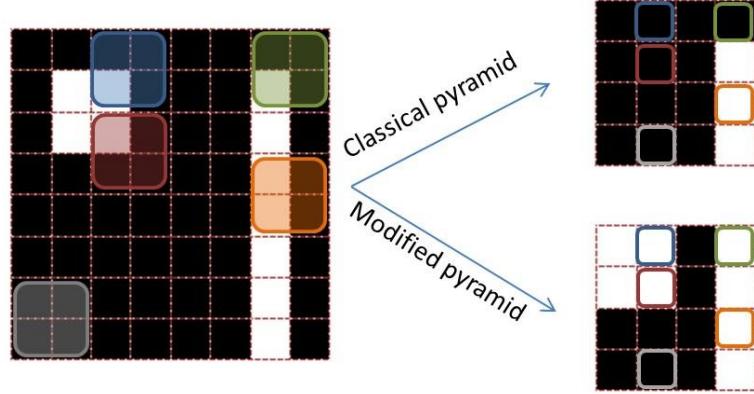


Fig. 7-3. Example of classical and modified pyramidal reductions from an  $8 \times 8$  to a  $4 \times 4$  resolution for visual representation. Four pixels correspond to one pixel in the lower pyramidal level (shown with the same border color); notice that the upper left object of interest (white) is preserved when using the modified pyramid.

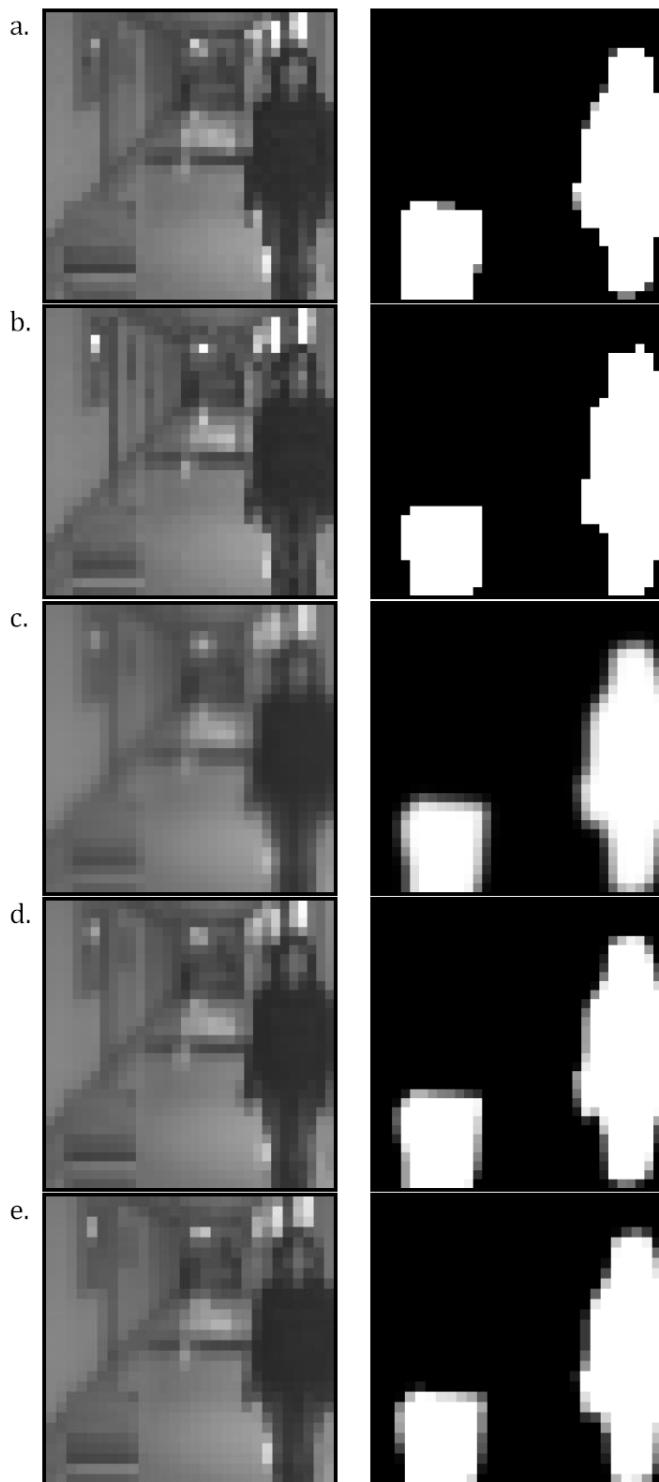


Fig. 7-4. The low resolution results from the images using (a) linear interpolation, (b) nearest neighbor interpolation, (c) Gaussian pyramid, (d) pyramidal reduction  $2 \times 2$ , (e) pyramidal reduction  $3 \times 3$ .

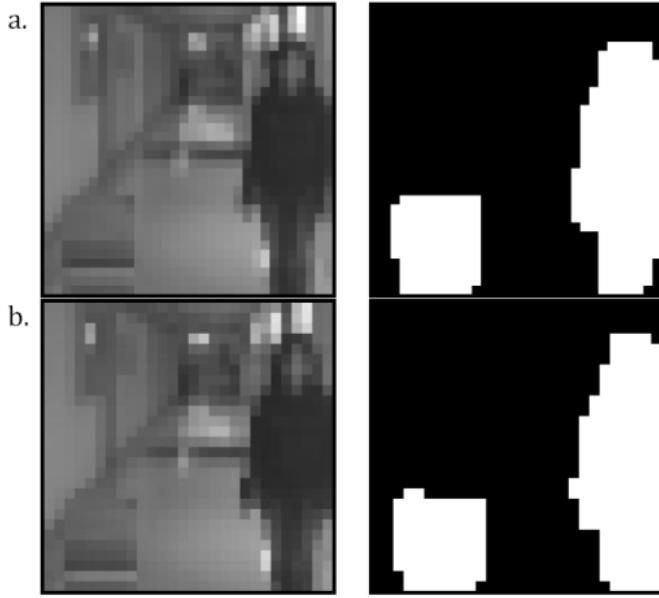


Fig. 7-5. Results for images using criteria of type C2/C4. (a) pyramidal reduction  $2 \times 2$  (b) pyramidal reduction  $3 \times 3$ .

## 7.6 3D representation

The stereo vision system that accompanies the Tyflos prototype is capable of producing depth maps. The depth maps are then simplified to only 4 distance ranges that they correspond to the 4 vibration levels [24] as shown in Table 5-1. The final depth map will be called *image-of-distances* and it 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 5-1. In Fig. 7-6b each segmented region of the initial fused image (Fig. 7-6a) corresponds to its own distance in the 3D navigation space.

In Fig. 7-6a we show the  $256 \times 256$  image of distances captured by the cameras. We can see from that the left object is in a close distance and the person in a medium distance. It's worth noticing that we also have the near part of the ground recognized as an object in a long distance.

### 7.6.1 Without criteria

The five previous mentioned algorithms are applied to the image of distances to get the  $32 \times 32$  resolution for the image (Fig. 7-7).

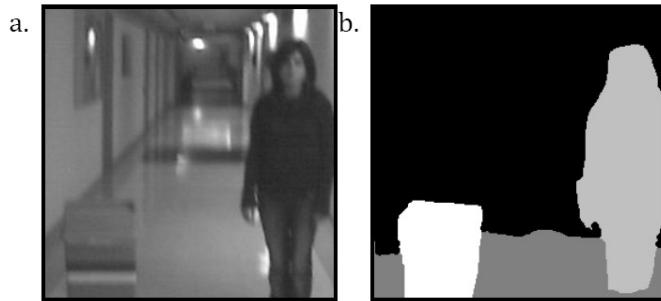


Fig. 7-6. a) Fused image and b) Image of distances.

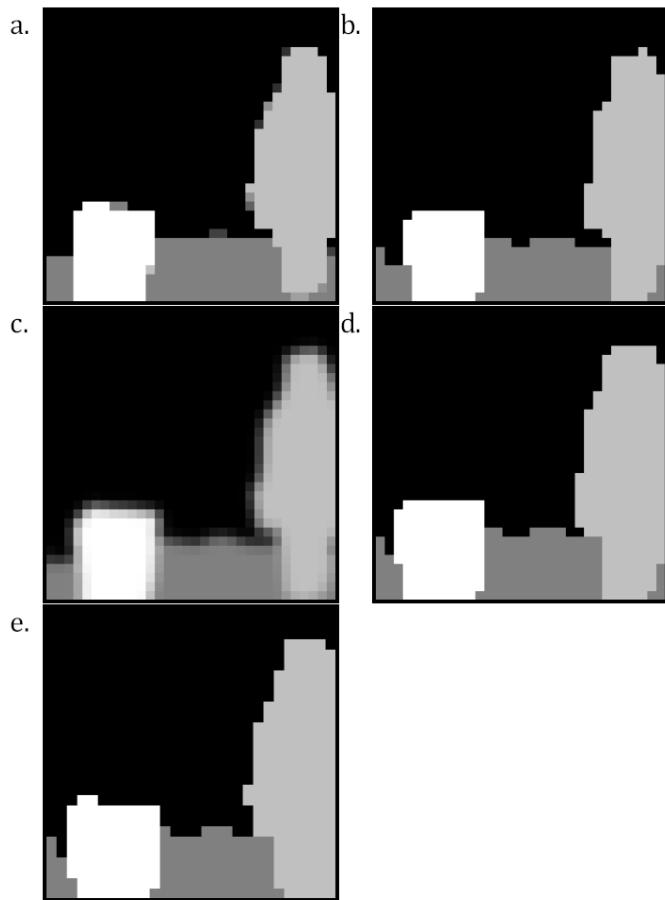


Fig. 7-7. 3D Low resolution results for images of distances: (a) linear interpolation, (b) nearest neighbor interpolation, (c) Gaussian pyramid, (d) pyramidal reduction  $2 \times 2$ , (e) pyramidal reduction  $3 \times 3$ .

A quick observation is that the classical pyramids (d) and (e) offer better results by showing more information about the objects of interest.

### 7.6.2 Geometric characteristics of the system

In order to derive the criteria to our high-to-low representation, we need to study the geometry of a camera-based navigation system such as Tyflos.

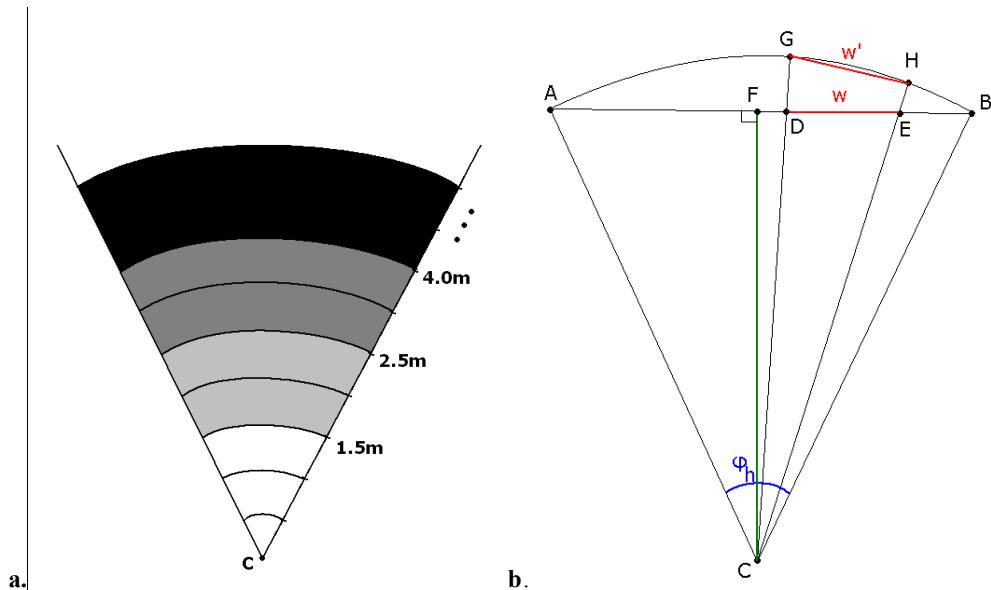


Fig. 7-8. Horizontal navigation space representation (C is the user/cameras), a) different distance ranges (0,1.5m), (1.5m to 2.5m), (2.5m to 4m) and (over 4.0m) with direct correspondence to Table 3-2, b) detail of the geometry for calculating the critical path width.

#### 7.6.2.2 Open path issue

Fig. 7-8a presents the horizontal navigation space where C is the position of the cameras (i.e. the user). The different grayscale levels represent the different distance ranges as defined in Table 3-2.

Assume that the camera has a view angle of  $\phi_h$  degrees, G and H are two obstacles defining an open path w and AB is the horizontal dimension of

the image plane (Fig. 7-8b). We know that the projected width  $w$  on the AB will always be smaller or equal to  $w'$  while for small view angles the difference is smaller. Thus, without loosing a lot in accuracy we assume that the width of the path will be  $w$ . We can underline that the fact that calculated path is usually smaller than the actual, makes the calculation safer for navigation.

The distance range where the closest obstacle (G or H) is, defines the active distance  $d_a$ . For example if obstacle G is in distance range (0 to 1.5m) and G in (1.5m to 2.5m) then the active distance would be 0.75m which is the middle of the (0 to 1.5m) range. The maximum width that the cameras can capture at the active distance is AB and will be called Field Of View width,  $w_{FOV}$ :

$$(Eq. 7-3) \quad w_{FOV} = 2d_a \tan\left(\frac{\varphi_h}{2}\right)$$

If the horizontal resolution of the image is  $r_h$ ,  $w_{FOV}$  corresponds to  $r_h$  number of pixels. Thus, the path width  $w$  will correspond to  $r_p = \frac{r_h w}{w_{FOV}}$ . If  $w_{cr}$  is the minimum/ critical width that a path should be to be considered as open path, then it will correspond to  $r_{cr}$  number of pixels:

$$(Eq. 7-4) \quad r_{cr} = \frac{r_h w_{cr}}{w_{FOV}}$$

and by substituting (Eq. 7-4)

$$(Eq. 7-5) \quad r_{cr} = \frac{r_h w_{cr}}{2d_a \tan\left(\frac{\varphi_h}{2}\right)}$$

The cameras capture images with 4:3 aspect ratio (352×264 pixels) and the pyramidal reductions work with square 1:1 images, 256×256 pixels (**Error! reference source not found.**), so we need to calculate the active view angle.

$$(Eq. 7-6) \quad \left. \begin{aligned} w_{4:3} &= 2d_a \tan\left(\frac{\varphi_{4:3}}{2}\right) \\ w_{1:1} &= 2d_a \tan\left(\frac{\varphi_{1:1}}{2}\right) \end{aligned} \right\} \Rightarrow \varphi_{1:1} = 2 \arctan\left(\frac{w_{1:1}}{w_{4:3}} \tan\left(\frac{\varphi_{4:3}}{2}\right)\right)$$

where  $w_{4:3}$ ,  $\varphi_{4:3}$  and  $w_{1:1}$ ,  $\varphi_{1:1}$  are the FOV widths and view angle for different aspect ratios.

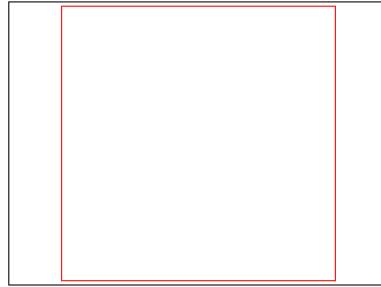


Fig. 7-9. Crop factor of the captured images. The black frame is the cameras' 4:3 and the red is the square 1:1 needed for the pyramidal reductions. The result is a smaller horizontal view angle (the vertical view angle is not affected significantly).

If we define crop factor:

$$(Eq. 7-7) \quad f_c = \frac{w_{1:1}}{w_{4:3}}$$

we can come up with a final formula that can calculate the critical width in pixels for the different active distances (i.e. 0.75m, 2m or 3.25m) and for the different image resolutions (i.e. different pyramidal levels). Substituting (Eq. 7-6) in (Eq. 7-5) we have:

$$(Eq. 7-8) \quad r_{cr} = \frac{r_h w_{cr}}{2d_a f_c \tan\left(\frac{\varphi_h}{2}\right)}$$

The cameras currently used for the prototype have horizontal view angle of  $31^\circ$  and the crop factor will be  $f_c = \frac{256}{352} \approx 0.727$ . We decide that the width for a safe navigation path should be at least 0.5m, wide enough for a person to

pass through.

#### 7.6.2.3 The “ground” issue

One of the important tasks during navigation is to find the open navigation paths. These, are paths/ directions in the 3D space that the user (visually impaired) can navigate safely. In the image of distances we noticed that the floor is represented as an object and, if we translate distances into vibration this would be translated as vibrations. This is something unwanted because the floor it is not an object and if we get vibrations in all the horizontal dimension pixels, the user would have the impression that there are no open paths.

If the camera is mounted at height  $H$  above the ground,  $\sin(\varphi/2) = (H/2)/d$ , where  $d_g$  is the distance from the ground (Fig. 7-10), we have:

$$(Eq. 7-9) \quad d_g = \frac{H}{2\sin(\frac{\varphi}{2})}$$

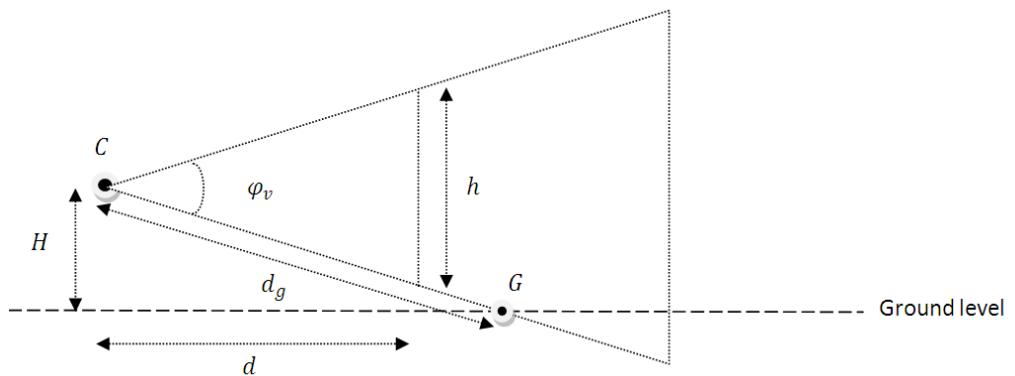


Fig. 7-10. Vertical navigation space representation.

For a person of height  $H = 1.6m$  and  $\varphi = 62^\circ$  we have  $d_g \approx 1.55m$ . In other words, the lower pixels are possible to give us close distances but not very close.

### 7.6.3 With geometric criteria

Two criteria of type C1 (distance related) can be derived from the open path and the “ground” issue and they are applied in every level of the classical pyramidal reductions. We replace the pixels (distances) of every “open” column with distances  $>2\text{m}$  (long distance or no obstacle), see Fig. 7-11a and Fig. 7-11b. For every set of consecutive “open columns”, we replace their pixels with “no obstacle” only if their total width is greater or equal to the critical open path width for that resolution, see Fig. 7-12c and Fig. 7-12d.

The results, as far as the obstacles, are identical but there is a difference in the open paths represented. The small path on the left of the left obstacle is not considered as free when using the critical open path width criterion; it is very narrow to allow navigation. On the other hand the open path between the left obstacle and the standing person; it is wide enough so it is cleared, given as open path.

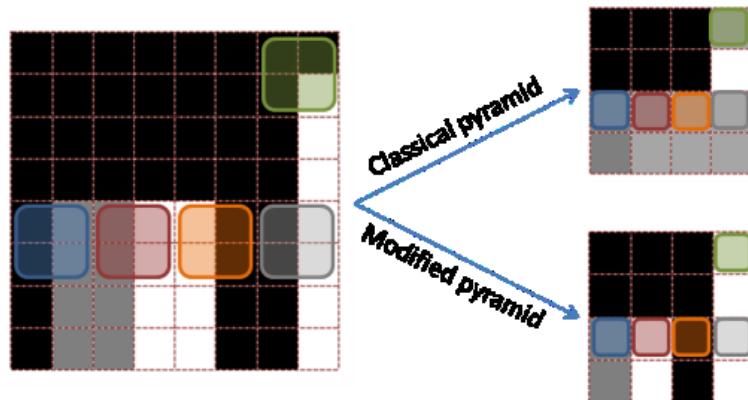


Fig. 7-11. Example of classical and modified pyramidal reductions from an  $8\times 8$  to a  $4\times 4$  resolution for 3D representation. Four pixels correspond to one pixel in the lower pyramidal level (shown with the same border color); notice that the classical pyramid doesn't keep important information for proximity objects (more white) or open paths (black).

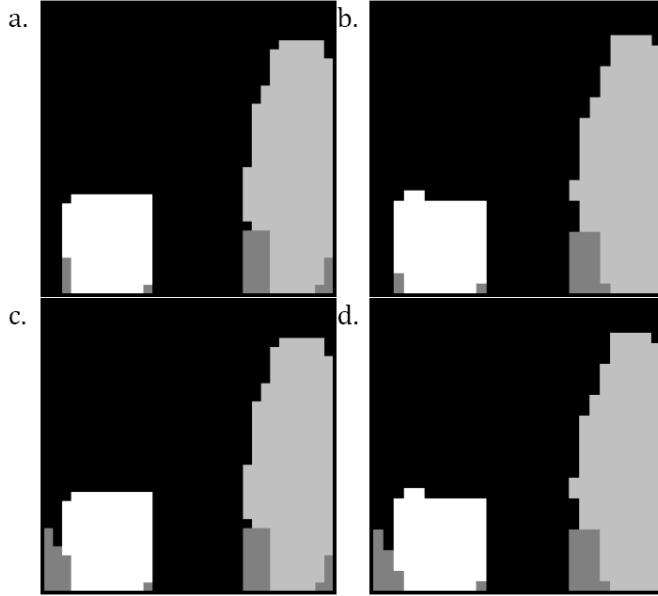


Fig. 7-12. 3D Low resolution results: a) criterion 1, pyramidal reduction  $2 \times 2$ , b) criterion 1, pyramidal reduction  $3 \times 3$ , c) criterion 2, pyramidal reduction  $2 \times 2$ , d) criterion 2, pyramidal reduction  $3 \times 3$ .

## 7.7 Information modeling

Studying the information that an image carries can help us understand if there is a way to extract and preserve, as much as possible, important navigation information (obstacles, objects, corridors) when its resolution has to be reduced.

### 7.7.1 *Information measures*

Image data can be classified using different features such as spectral, textural or contextual. Thirteen basic textural measures are proposed [85] and they are calculated in the spatial domain using image's statistical parameters. The idea behind that is that the information is contained in the average spatial relationship that the pixel's intensities have with each other. For this discussion, three measures will be used:

### Textural Entropy:

$$(Eq. 7-10) \quad ENT = - \sum_i^{N_g} \sum_j^{N_g} p(i,j) \log [p(i,j)]$$

### Textural Correlation:

$$(Eq. 7-11) \quad COR = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i \cdot j) \cdot p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

### Angular Second Moment:

$$(Eq. 7-12) \quad ASM = \sum_i^{N_g} \sum_j^{N_g} [p(i,j)]^2$$

where  $p(i,j)$  is the  $(i,j)$ th element of the normalized spatial-dependence matrix,  $N_g$  is the number of grayscale levels and  $\sigma_x, \sigma_y, \mu_x, \mu_y$  are the means and standard deviations of the marginal-probability matrices [85].

ENT is identical to the classical definition of the entropy in information theory [84] and it is a measure of the images compressibility. COR is a measure of color (in our case, grayscale) linear-dependencies and ASM is a measure of homogeneity in the image. All the three measures are calculated from the so called spatial dependency matrix of the image which can be computed for different “inter-pixel” orientations (0, 45, 90 and 135°).

#### **7.7.2 Experiment**

Four test scenarios from an indoor navigation environment are studied. Every scenario has a fused image, an image-of-distances and an image-of-interest (Fig. 7-13). They include different objects that can appear on the scene such as obstacles (e.g. chair or trash can in scenarios 0 and 1), objects (e.g. persons) and possible corridors.

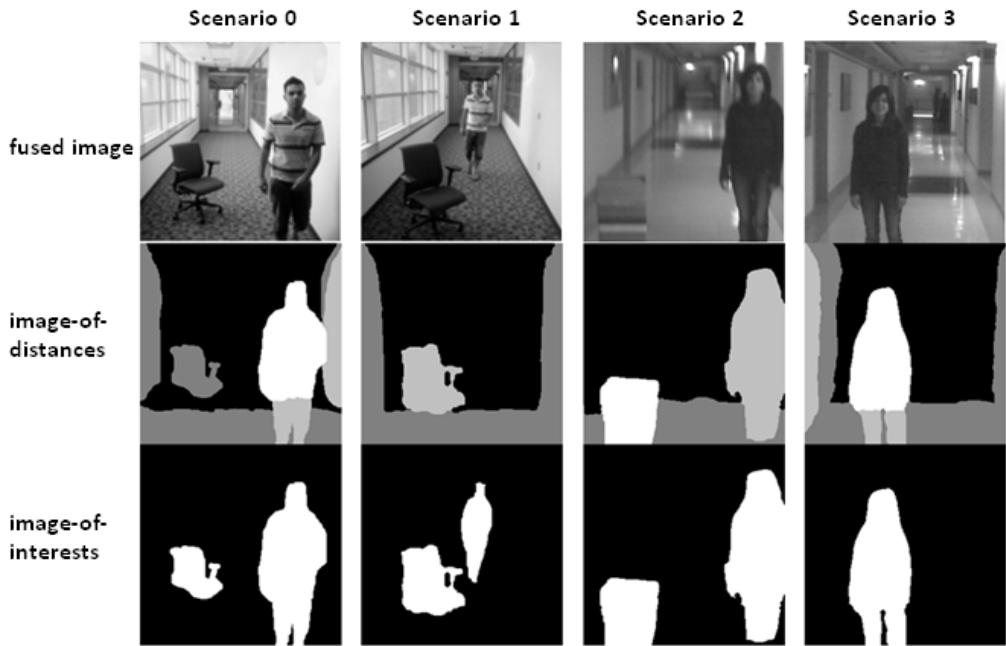


Fig. 7-13. The 4 test navigation scenarios: fused-image, image-of-distances (darker gray-level corresponds to larger distance from the user) and image-of-interests (white pixels belong to regions of interest).

In the first processing step we obtain the depth map of the scene using the two cameras stereo vision system of the Tyflos system. In the second step, we select the objects that appear in less than a desirable distance from the camera, and we assign distances to the objects that correspond to the four vibration levels. In the third step we select moving objects that appear the background. In the fourth step, using object recognition methodologies, we determine and separate objects of interest from obstacles in the 3D scene.

The three measures were computed for every test image following for every pyramidal level, and for different orientations. Scenario 0 and 2 for the different pyramidal levels are shown in Fig. 7-14 and Fig. 7-15. The results follow the same patterns for every measure and some representative are presented in Fig. 7-16, Fig. 7-17 and Fig. 7-18.

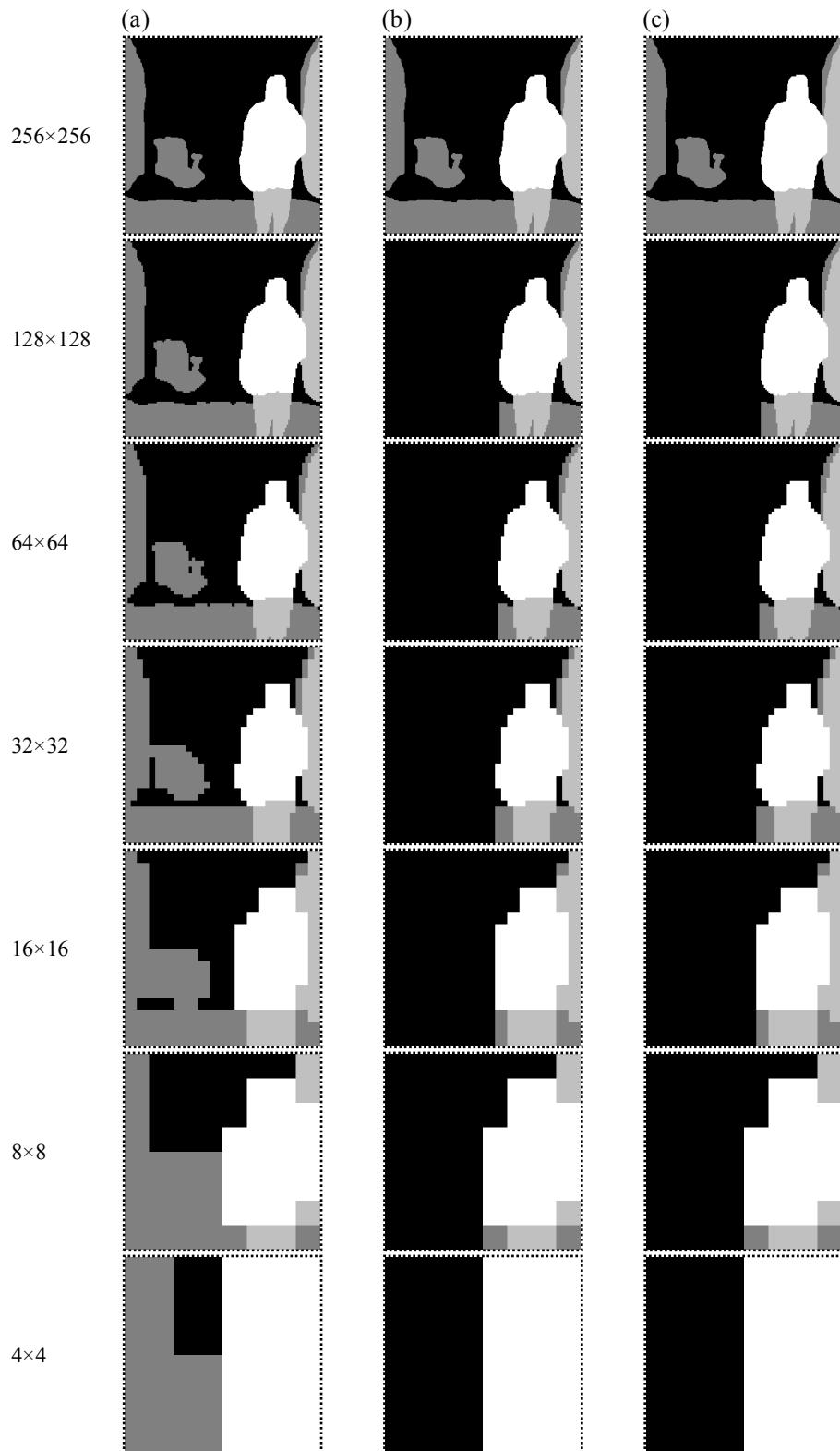


Fig. 7-14. 3D representation and pyramidal levels for scenario 0: a) standard pyramid, b) modified pyramid with criterion 1 and c) modified pyramid with criterion 2.

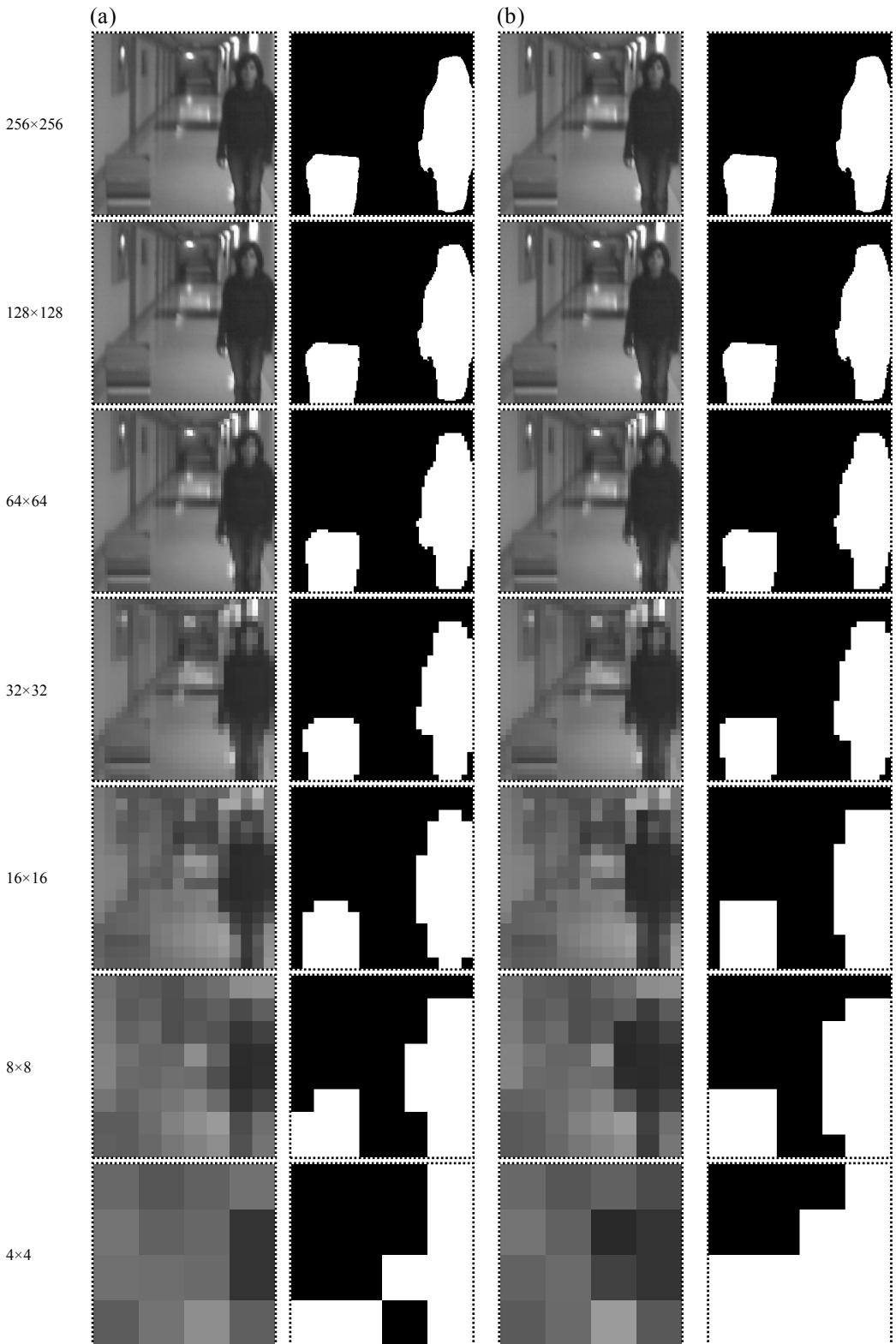


Fig. 7-15. Scenario 2 for different resolutions for the 2 methodologies (a) and (b) for the fused and the image-of-interests. We notice that the modified pyramid (b) emphasizes more on the regions of interest (white).

### 7.7.3 Discussion

Before getting into more detailed discussion, it is important to elaborate about the relation of complexity and information. Complexity is a measure of how heterogeneous/ non-uniform the image is or in other words how much information the image carries. It is well known that the importance of pieces of information is subjective and relevant to the application. Thus, depending on the application, some pieces of information is useful and some not and since usually the non-useful information is called noise, from now on, when we mention information we mean “useful information”.

Textural entropy (ENT) is a measure of the image’s complexity-compressibility.

- The fused image is a highly complex image (many regions, many grayscale values etc) and so the reduction of the resolution reduces the ENT of the image.
- On the other hand, the reduction of the resolution increases the ENT not because the complexity is increased but because the information that the image carries is denser. The reason for that is that the image-of-distances and the image-of-interests are more simple images since they are outcomes of processing of the fused image. The regions of interest such as obstacles, objects, corridors retain their basic characteristics as the resolution is reduced and thus the increasing the desirable information density in the reduced image.

Textural correlation (COR) follows the same patterns as the textural entropy with the only difference that the angle here doesn’t affect the results. An-

gular second moment (ASM) is a measure of homogeneity (entropy is a measure of heterogeneity) and this explains the reversed results:

- ASM for the fused images is increased as resolution is decreased.
- ASM for image-of-distances and image-of-distances is decreased as the resolution is decreased.

The above results show that the different entropy measures are capable of giving a **quick, real-time and accurate indication about the complexity and the information of the image**. They are indication of the different environmental situation that the user can face e.g. a complex image with few information can possibly be an un-safe environment for the user to navigate. Consequently, the user or the system can decide for the next step, which can be either a halt of the system or a use of more advanced computer vision methodologies for a more detailed (but also computationally expensive) interpretation of the environment. Additionally, in all the cases the modified methodologies return images with lower or at least equal complexity as the standard methodology. This means that the images are less complex but still carrying the same amount of important information. Note here that the difference in scale doesn't affect those qualitative results.

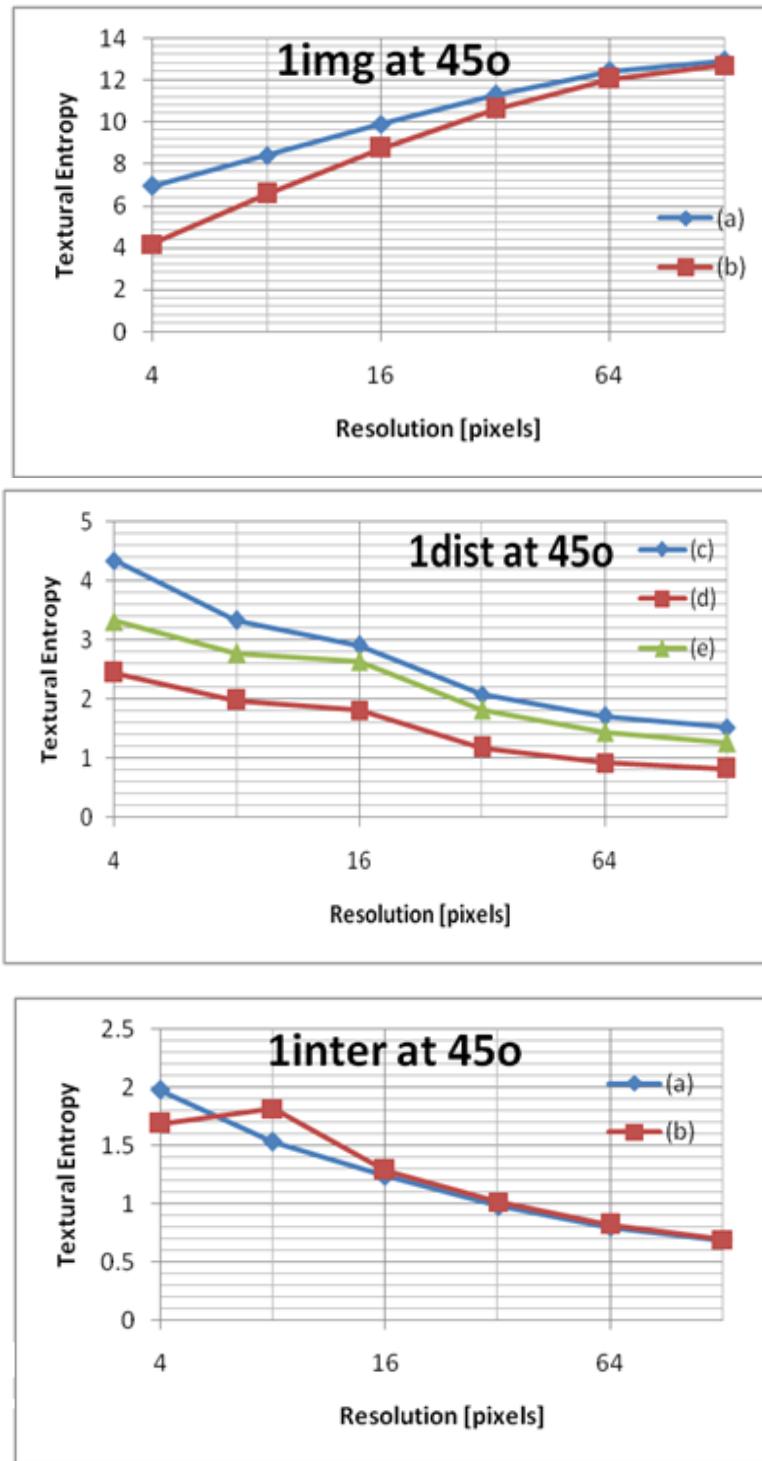


Fig. 7-16. Textural Entropy for scenario 1 at 45° orientation (left to right: fused-image, image-of-distances, image-of-interests). The blue lines represent the classical approach (a) and (c) and the red and green the modified methodologies (b), (d) and (e).

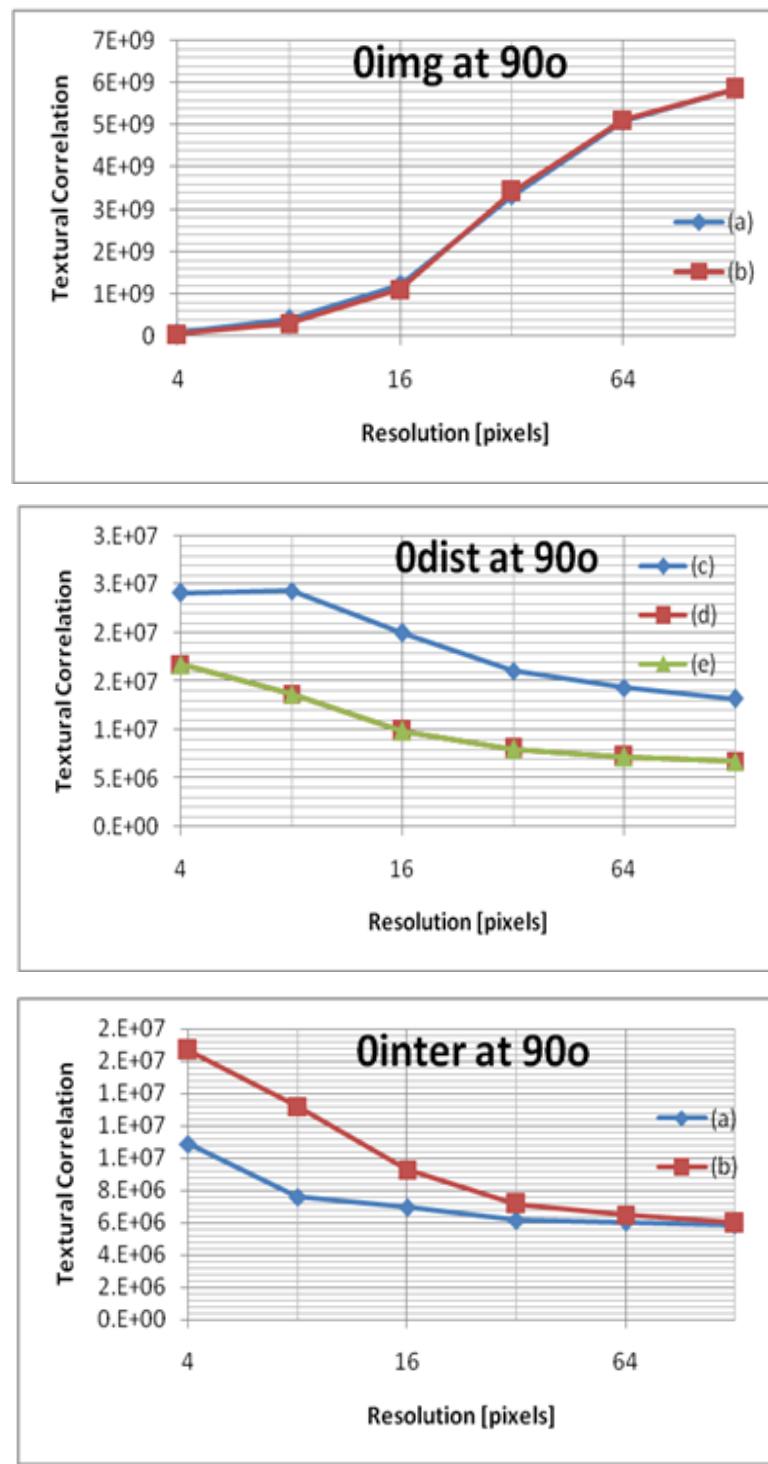


Fig. 7-17. Textural Correlation for scenario 0 at 90° orientation (left to right: fused-image, image-of-distances, image-of-interests).

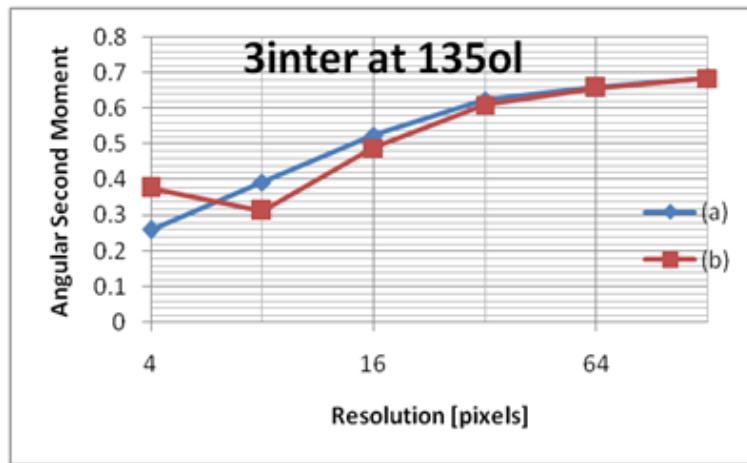
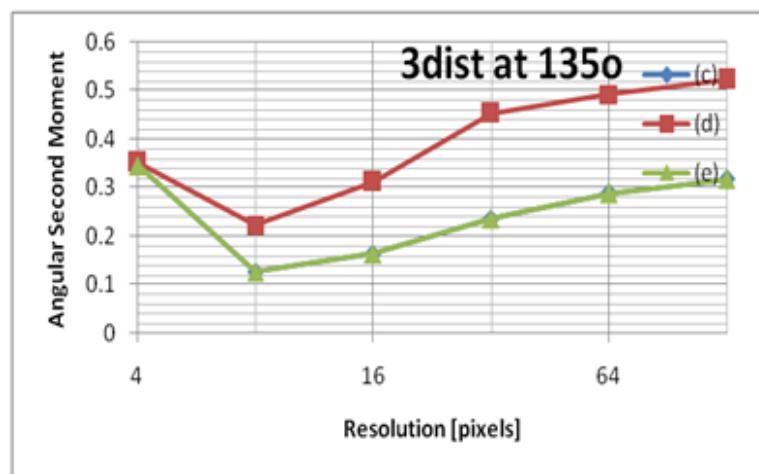
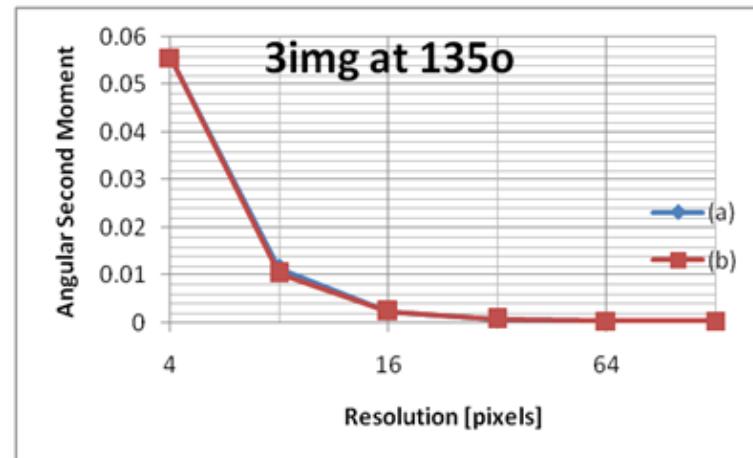


Fig. 7-18. Angular Second Moment for scenario 3 at 135° orientation (left to right: fused-image, image-of-distances, image-of-interests)

## **Chapter 8: Tactile Vocabulary and Testing**

While haptic interfaces have been troubling scientists from different fields since the 1960s [41], to our knowledge, the exploitation of such tactile interface for representing the 3D environment for mobility purposes has not been performed. In this final chapter the prototype is tested on sighted and visually impaired users while a special emphasis will be given to the work towards the creation of a tactile vocabulary. The statistical experimental results and the feedback from the users will be used to evaluate the software and hardware components of the prototype and open possible avenues for future changes and enhancements.

The 16 vibrating elements are capable of producing  $2^{16}=65,536$  vibration patterns; words, in terms of the Vibration Array Language (VAL). Experimental results in haptic devices show that the users need to some extent to learn those patterns, thus the large number of pattern would correspond to an undesirable heavy cognitive load.

We work towards the selection of a set of patterns that will constitute the 2D tactile vocabulary, a set that the user will be able to learn and distinguish between them with. In a more formal way, we try to create the dictionary of the VAL; what are the combinations of symbols that create “valid” words. This set will be selected using criteria based on different possible navigation scenarios, safe navigation and simplicity and using two rule-generation ap-

proaches: the vertical and the horizontal. A pattern vibrotactile recognition work has been done also by Jones et. Al [90] but the patterns were simple and they had the form of directional or instructional cues. Directional cues is the most often use of tactile interfaces such as Van Erp's tactile navigation display [55]. Finally, in order to match the generated patterns with one from the tactile vocabulary, we propose a pattern matching methodology using a modified Euclidean dissimilarity measure.

## 8.1 Experimental design

A series of experiments were performed and described in the next sections for testing the user's ability to recognize symbols and patterns on the 2n prototype's  $4 \times 4$  vibration array. For the first set of experiments described in 8.2, 10 sighted subjects (6 male, 4 female) volunteered with their ages varying from 14 to 60 and they have never been trained with or used the 2D vibration array or other tactile feedback devices. They were asked to wear the vest and we made sure that all the vibrating motors are in contact with their body.

For the second set described in 8.5, Two congenitally visually impaired students participated: the first subject was a 25 years old male, totally blind and the second was a 23 years old female with severe visual impairment. This first attempt, despite the only two subjects will give us a first insight on whether there are any major differences between sighted and visually impaired users.

We are mostly interested in demonstrating that the users are able to perceive the basic structures of the 3D environment via the vibration array and the tactile vocabulary (as described later). The proof-of-concept goal makes us

followed the repeated measures design with straightforward strong manipulations [91], [92]

For every participant we set the stage of the experiment by explaining why the experiment was being conducted and run a trial run of the different experimental phases to see if the participant understood the instructions or had any questions. The instructions were given mostly verbally to avoid experimental bias between the sighted and visually impaired.

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. Additionally a few rest periods were added between the different experimental phases. Finally, a debriefing of the experiment let the participants talk about their experience, providing important feedback.

## 8.2 Testing on sighted users

### 8.2.1 *Testing vibration levels*

The vibration array has the ability to represent the 3D navigational space. The 3<sup>rd</sup> dimension, which is the distance of the subjects from the user, is represented with the different vibration levels of the motors. The correspondence of the vibration levels to vibration frequencies and distance is shown in Tables 1 and 2.

#### 8.2.1.1 Experimental results (vibration levels)

An experiment was performed to validate the user's capability to recognize the different vibration frequencies. Three elements were randomly se-

lected and random vibration levels were sent to them for every trial. The subjects were asked to identify the vibration level for each element (vibration level 0 was excluded due to its simplicity).

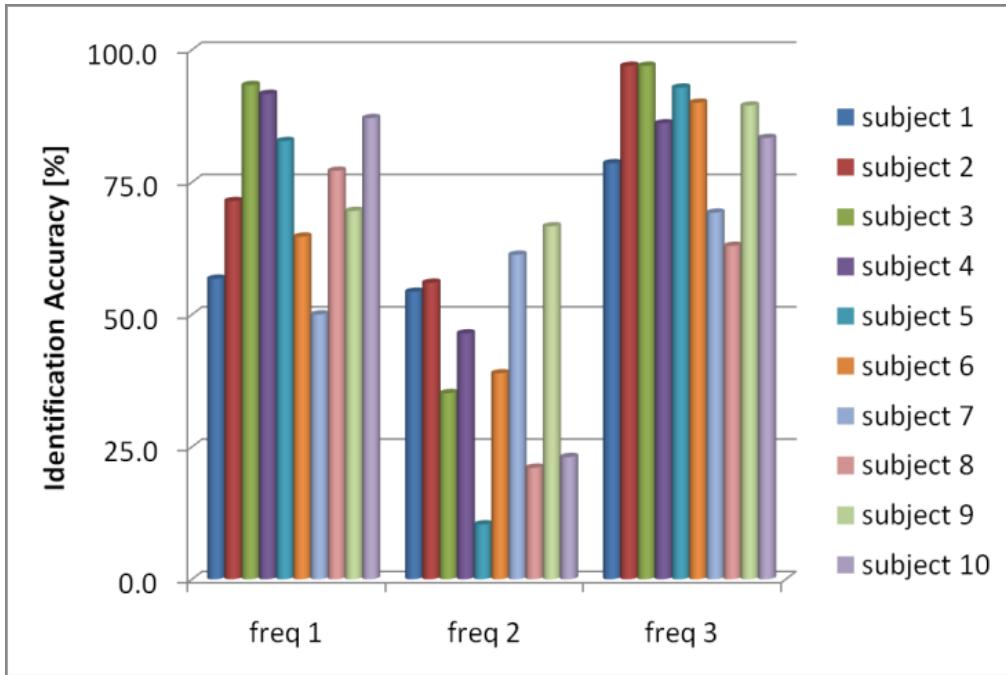


Fig. 8-1. Vibration levels/frequencies Identification accuracy (sighted users).

Fig. 8-1 presents the identification accuracy for the 3 vibration levels. We notice that for levels 1 and 3, most of the subjects responded in a good manner showing identification accuracy over 60% and many times over 75-80%. The subjects had difficulties identifying level 2. There are two possible explanations for that. The first is that the frequencies 1 and 2 are relatively close (1Hz and 2Hz correspondingly), compared to frequency 3 (10Hz) and this was mentioned by the subjects during the experiment.

The second explanation is that that none of the users was trained with the vibration array. Studies show that training of the users is a very crucial part during the development of an assistive device such as Tyflos and we strongly

believe that the users trained with the system will dramatically increase their performance.

### 8.2.2 Testing symbols (*vertical rules*)

During navigation the most important information is whether there is an open path to navigate and this 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 can be set to reduce the number of patterns that can appear in the vertical dimension (y).

Table 8-1. The 6 vibration symbols (activated vibrating motors are filled with black).

S0 : A	S1: B	S2 : C	S3 : D	S4 : E	S5 : F
○	○	○	●	●	●
○	○	●	●	○	●
○	●	●	○	○	●
○	●	●	○	○	●

Table 8-2. The four different positions that the symbols can appear on the 4×4 VA.

a: left	b: left center	c: right center	d: right
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○

We select 6 types of VAL column-type symbols shown in Table 8-1. They can appear in four different positions on the array (Table 8-2) thus, the possible navigation patterns are  $6^4=1296$ .

### ***8.2.2.2 Experiments (symbols)***

We experiment with the users' ability to recognize the different vertical symbols in different positions.

#### **Position recognition experiments**

For the first set of experiments (#1 to #5) vertical symbol S1-S5 is sent to the VA in random position (a, b, c or d). For example, as we can see in table 3, for experiment #3, vibration patterns 09, 10, 11 and 12 are selected and they correspond to symbol S3 in all possible positions. The user was asked to identify the correct position by naming it (a, b, c or d) and as soon as he/ she identified it, a new random pattern (out of the 4) was sent. One hundred measurements/ trials were performed for each experiment.

Table 8-3. Experiment #1 to #5 for position identification accuracy for the different symbols (B, C, D, E, F)

<b>subject</b>	<b>Position accuracy [%] for different patterns</b>				
	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>#1</b>	99.0	100.0	98.0	100.0	100.0
<b>#2</b>	99.0	100.0	100.0	99.0	100.0
<b>#3</b>	99.0	99.0	97.0	98.0	98.0
<b>#4</b>	97.0	100.0	98.0	85.0	98.0
<b>#5</b>	100.0	100.0	98.0	100.0	99.0
<b>#6</b>	96.0	100.0	98.0	98.0	100.0
<b>#7</b>	99.0	99.0	97.0	97.0	99.0
<b>#8</b>	96.0	100.0	99.0	92.0	95.0
<b>#9</b>	99.0	97.0	96.0	100.0	96.0
<b>#10</b>	100.0	100.0	98.0	94.0	99.0

### Symbol and symbol/ position recognition experiments

In experiment #6, a random vertical symbol (S0 to S5) is sent to position “a” and the user was asked to identify the correct symbol. In the final experiment #7, the user was sent a random symbol in a random position and the user was asked to identify the correct symbol and the correct position. 300 measurements/ trials were taken for each experiment.

Table 8-4. Experiment #6: symbol identification accuracy

Symbol identification accuracy [%] for the ten subjects										
#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	

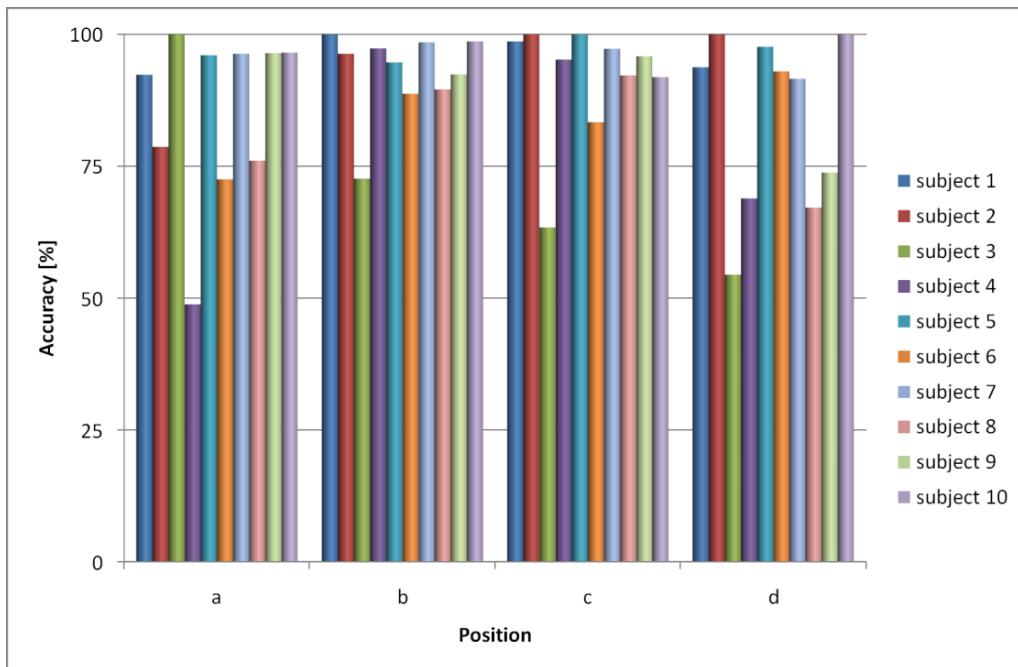


Fig. 8-2. Experiment #7: User's accuracy in identifying the correct position of a random symbol (sighted users).

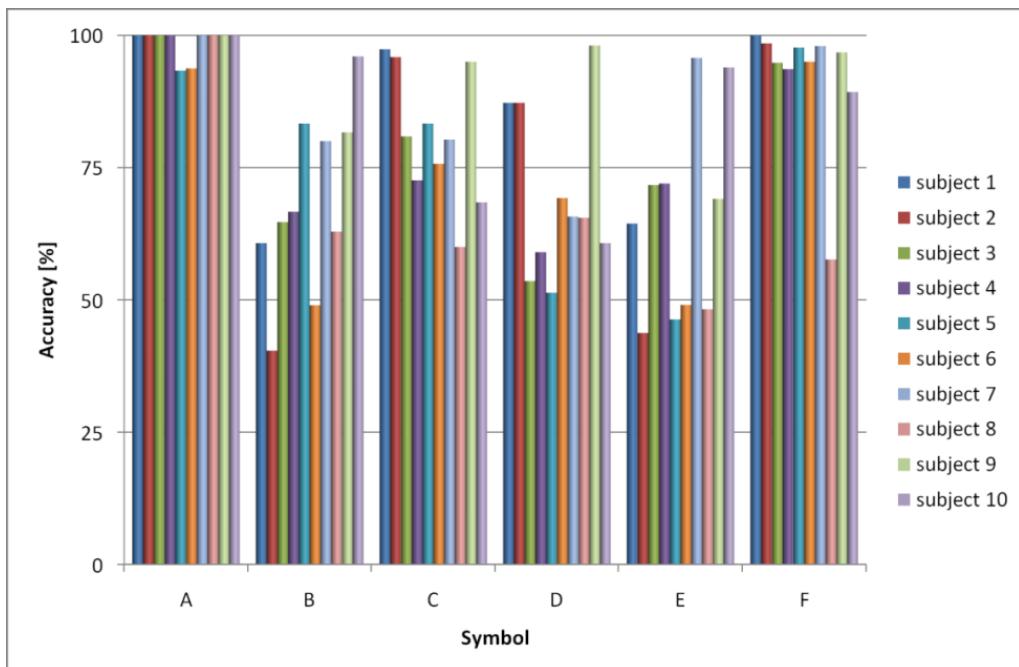


Fig. 8-3. Experiment #7 (sighted users): Users' accuracy in identifying the different symbols in random positions.

#### 8.2.2.3 Results – Discussion (vertical rules)

Table 8-3 shows that users had at least **92% accuracy in identifying the position/direction of a predefined symbol**. On the contrary in experiment 7, were the symbol was random, those percentages are considerably smaller (Fig. 8-2). Possible explanations for this difference are:

- The users didn't only have to identify the patterns and/ or positions of symbols but they also had to say corresponding letter for each one of them. This requires more thinking and so possible mistakes. Indeed, many subjects said that that many times they were confused and used they were using the wrong letter to describe a pattern. Thus in experiment 7, that symbols and positions had to be identified, the possibility for saying the incorrect letter was higher.
- **Fatigue effect;** many subjects complained that they got tired, especially

during the final experiment which is also the longest. This can probably result to more incorrect letter selection (as discussed before).

On the contrary, Table 8-4 and Fig. 8-3 show that they had some problems identifying the correct symbol. The confusion matrix (Table 8-6) shows that the major problem was distinguishing if there is one or two consecutive vibrators active; difficulties identifying the symbol B and D where most of the times they were incorrectly identified as C and E accordingly (28.3% and 30.2% misidentification).

Table 8-5. Experiment #6 (sighted users): Average symbols' confusion matrix [%](sighted users).

Actual pattern	Guessed pattern					
	A	B	C	D	E	F
<b>A</b>	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
<b>B</b>	3.3	<b>85.2</b>	<b>10.1</b>	0.5	0.9	0.0
<b>C</b>	0.0	9.9	<b>83.2</b>	3.7	0.2	3.1
<b>D</b>	0.0	2.2	4.4	<b>76.3</b>	<b>11.6</b>	5.5
<b>E</b>	5.1	0.3	1.0	<b>12.4</b>	<b>80.8</b>	0.4
<b>F</b>	0.4	0.2	9.5	<b>10.4</b>	0.4	<b>79.1</b>

Table 8-6. Experiment #7 (sighted users): Average symbols' confusion matrix [%].

Actual pattern	Guessed pattern					
	A	B	C	D	E	F
<b>A</b>	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
<b>B</b>	0.7	<b>68.6</b>	<b>28.3</b>	0.9	0.9	0.6
<b>C</b>	0.0	<b>10.4</b>	<b>81.2</b>	2.6	0.4	5.5
<b>D</b>	0.2	0.3	2.4	<b>69.8</b>	9.1	<b>18.3</b>
<b>E</b>	0.2	1.3	1.8	<b>30.2</b>	<b>65.7</b>	0.7
<b>F</b>	0.0	0.3	3.4	3.8	0.2	<b>92.3</b>

From bibliography [45], [48], [53] we know that the spatial acuity on the torso is higher than the one that our array has. As far as navigation purposes, this confusion between symbols B-C and D-E is not of major concern since the subject can still identify the low-obstacle (for example if there was a confusion between B and F then it will be more important since the symbols represent different situations i.e. low-obstacle and tall obstacle. Some possible explanation for that misidentification is that the vibrators were not making proper contact with the user; the abdomen area is not uniform so not all the vibrators were placed as firmly.

Despite the problem 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.

### ***8.2.3 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.

The six horizontal rules:

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 show in Table 8-7. Now, by applying the above vertical and horizontal rules we reach a **tactile vocabulary of 298 words** which is **0.45% of the initial 65,536 patterns**.

Table 8-7. Horizontal rules experimental patterns. After application of the rules, white motors with X are enabled and black motors with X are disabled.

Rule 1			
Rule 2			
Rule 3			
Rule 4			
Rule 5			
Rule 6			

## 8.3 Pattern matching

As shown in the software and architecture in the second chapter, the final step before sending the pattern to the vibration array is a pattern matching. This is because the output of the high-to-low methodologies doesn't necessarily match with one of the patterns from the VAL vocabulary, so we need a methodology that maps the low resolution images to words from the VAL vocabulary.

### 8.3.1 Modified dissimilarity measure

Various similarity and dissimilarity measures have been used for pattern recognition/ matching applications [89]. The Euclidean distance is a fundamental dissimilarity measure. In the case of grayscale images, the Euclidean distance of two pixels is:

$$(Eq. 8-1) \quad d = \sqrt{p_1^2 - p_2^2}$$

where  $p_1$  is the grayscale value of pixels 1 and  $p_2$  the grayscale value of pixel two. Larger distance means larger difference of intensity/ color which means high dissimilarity of the pixels.

We can define a basic Euclidean dissimilarity measure between two images A and B of the same dimension  $m \times n$ :

$$(Eq. 8-2) \quad D = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \sqrt{(A_{ij} - B_{ij})^2} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |A_{ij} - B_{ij}|$$

where  $A_{ij}$  and  $B_{ij}$  are the values of pixels (i,j) of images A and B.

Two identical images will result in  $D=0$  which is zero dissimilarity. The larger the D, the more different the images are but this measure doesn't incorporate any pixel's spatial distribution and in our case, this is crucial so a mod-

ification is necessary.

The Vibration Array Language is based on vertical patterns, which correspond to the different vibration symbols and in terms of navigation to the directions of the different navigation paths. Thus, a solution for a better pattern matching would be, not to calculate the dissimilarity for the whole image but for every column separately, mapping every column to a symbol.

A parameter  $\lambda$  is introduced to modify the Euclidean distance dissimilarity measure of (Eq. 8-2) where now the new measure will be called  $D_t$  where  $t$  stands for “tactile”:

$$(Eq. 8-3) \quad D_t = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I_{ij} - P_{ij}|(1 + \lambda)$$

where  $I$  is the initial image,  $P$  the pattern with which  $I$  is compared and

$$(Eq. 8-4) \quad \lambda = \begin{cases} \frac{1}{\sqrt{m \cdot n}}, & \text{if } I_{ij} > 0 \wedge P_{ij} = 0 \\ 0, & \text{in any other case} \end{cases}$$

where  $m$  and  $n$  are the dimensions of the images.

Parameter  $\lambda$  reflects if the “active” pixels are matched correctly. Active pixels we call the pixels that carry distance information, excluding the pixels that correspond to greater than the maximum distance (i.e. no-vibration pixels). If an active pixel is not represented in the pattern then the dissimilarity is increased. For larger images the dissimilarity is increased less because the pixels carry less spatial information since the pixels correspond to smaller environmental space.

Table 8-8 shows an example of how the modified dissimilarity measure applies in comparison with the standard. the standard Euclidean dissimilarity

brings two possible matches: symbol S3 and S5 with dissimilarity 1. For a  $4 \times 1$  image  $\lambda=0.5$  and the modified dissimilarity matches only with S5. The incorporation of spatial characteristics on the new measure is evident because the new dissimilarities are more dispersed e.g. S2 has smaller dissimilarity with the image than S0 because it has less active pixel incorrect matches. From a navigation point of view symbol S5 represents better the image because represents better the active pixels (3 out of 3). Finally symbol S2 matches better than S0 because it represents at least one of the three active pixels.

Table 8-8. Example of symbol matching using the standard and modified dissimilarities. The black circles correspond to active pixels i.e. corresponding motors vibrate.

Image column	Possible matching symbols					
	S0	S1	S2	S3	S4	S5
●	○	○	○	●	●	●
●	○	○	○	●	○	●
●	○	○	●	○	○	●
○	○	●	●	○	○	●
<b>Standard dissimilarity</b>	3	4	3	<b>1</b>	2	<b>1</b>
<b>Modified dissimilarity</b>	4.5	5.5	4	<b>1.5</b>	3	<b>1</b>

Fig. 8-4 present the outline of the proposed pattern matching methodology: The initial image after the high-to-low methodologies is first updated using the modified dissimilarity measure so that its columns correspond to valid language symbols. The new image is mapped to a pattern with the smaller dissimilarity value. Finally the selected pattern is updated so that the distance information corresponds to the initial image; active pixels in the matched pattern inherit the distance information from the corresponding pixels in the initial image. If the initial image pixel is not active then it inherits it from the closest active pixel.

tive pixel of the same column. If there are more than one pixels then the pixel that is in the same half (bottom or up) is selected.

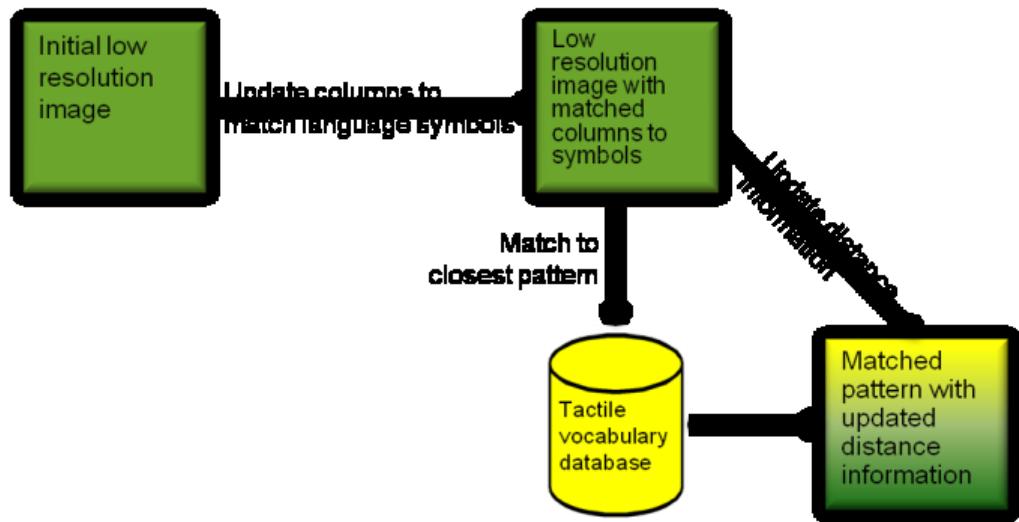


Fig. 8-4. Pattern matching methodology flow.

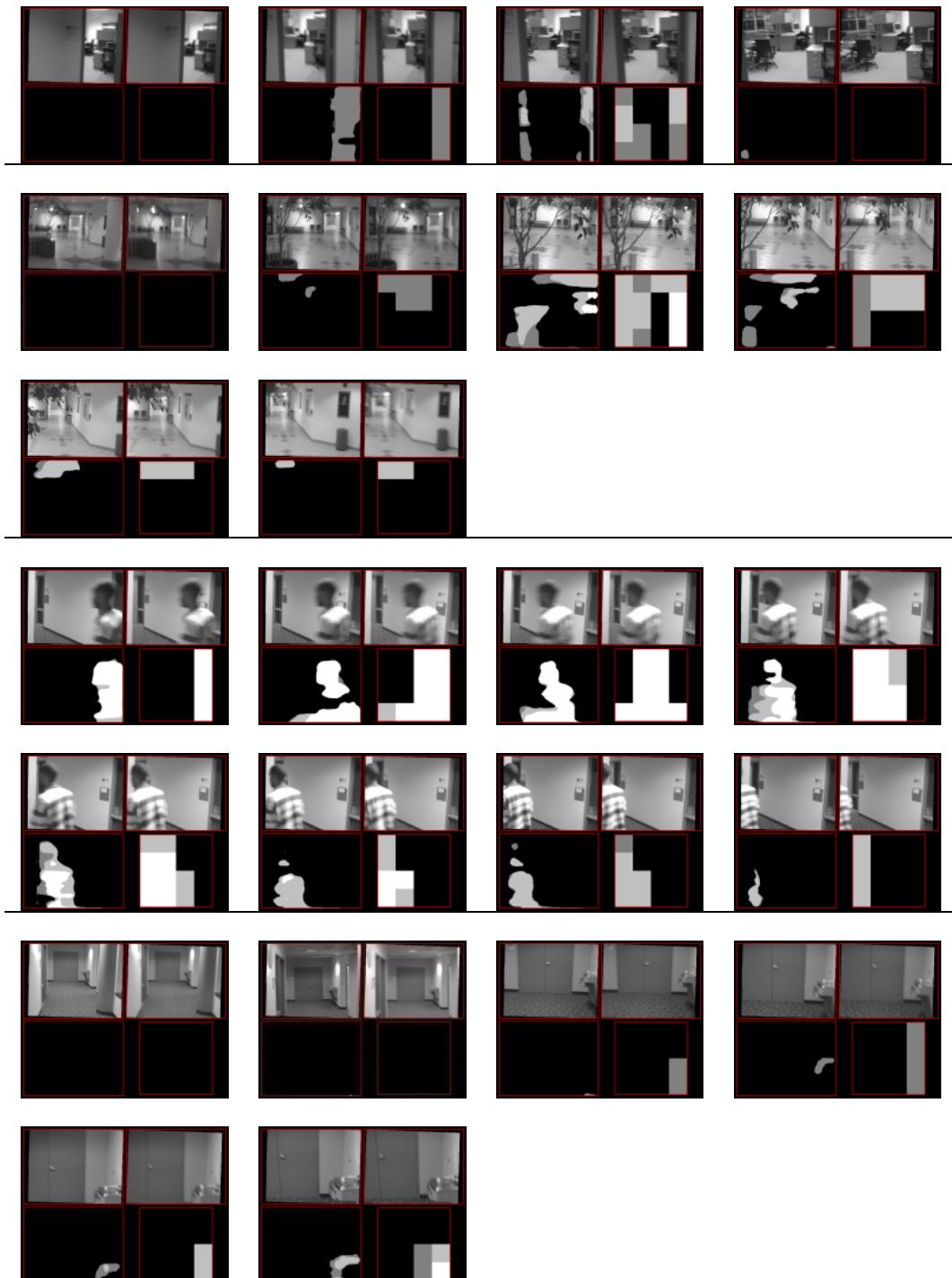


Fig. 8-5. Examples of selected frames. Every frame is consisting of 4 images: Top left and top right are left and right high resolution camera frames. Bottom left is the high resolution disparity map. Bottom left is the ( $4 \times 4$ ; low resolution) tactile vocabulary pattern sent to the vibration 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).

## 8.4 Testing Navigation Scenarios

The last experimental part involves testing with real navigation scenarios. 11 videos were recorded from inside the school of engineering. The videos include different possible navigation scenarios, including moving or static people, over-hanging obstacles, low height obstacles, doors etc. From the left and right cameras video frames 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. 8-5.

The users were asked to do a quick sketch of the vibration pattern sent to them which is the direction of open paths and obstacles with distance/ frequency information.

Table 8-9. Correct identification of: patterns without vibrations; open paths within patterns; open areas; obstacles; distance of obstacles; non-zero frequencies (sighted users).

<b>subject</b>	<b>empty patterns [%]</b>	<b>open paths [%]</b>	<b>open areas [%]</b>	<b>obstacles [%]</b>	<b>distance of obstacles [%]</b>
<b>0</b>	97.0	87.4	89.4	68.1	26.6
<b>1</b>	100.0	96.4	95.2	67.1	34.4
<b>2</b>	97.0	93.2	92.1	64.7	32.5
<b>3</b>	97.0	86.5	91.6	55.1	27.6
<b>4</b>	97.0	93.7	86.4	68.8	35.2
<b>5</b>	100.0	89.6	95.1	43.6	18.5
<b>6</b>	90.9	95.0	94.3	70.5	26.7
<b>7</b>	97.0	91.4	87.7	76.5	30.9
<b>8</b>	100.0	86.9	89.9	67.7	42.6
<b>9</b>	100.0	94.6	94.1	61.8	30.5
<b>Average</b>	<b>97.6</b>	<b>91.5</b>	<b>91.6</b>	<b>64.4</b>	<b>30.6</b>

Table 8-10. Navigation scenarios: average confusion matrix for the different frequencies (sighted users).

actual	guess			
	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

#### 8.4.2 Discussion (navigation scenarios)

The experimental results from the navigation scenarios are presented in Table 8-9 and Table 8-10. They give us some important feedback for the evaluation and further development of the Tyflos Navigation prototype but before further discussion we want to emphasize on two points:

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

The most important information that a visually impaired user needs to know while navigating, 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 8-9 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 (“ceiling effect”) but important to notice. We also notice that 91.6% of the open areas (inactive vibrating elements) were identified and this includes a **91.5% of the open paths**.

**A 64.4% of the obstacles were accurately identified as far as the x-y**

**position but as far as their distance only 30.6%.** A first look that we have a “floor effect” (a task that hardly any participant performs well) but a closer look on the confusion matrix (Table 8-10) shows that the users misidentified vibration levels with close frequencies. For example when the level was 3, 35.7% of the users guessed correctly level 3 but 29.3% guessed level 2, while a small 4.6% guessed level 1.

We also notice a misidentification of non-zero levels as zeros with 36.8%, 34.1% and 30.4% correspondingly. The most possible explanation is that the users misidentified the symbol which also means incorrect frequency identification (Note: as seen in the previous experiments, the confusion matrix of symbols (Table 8-6) shows that the users misidentified the symbols with similar percentages of up to 30.2%)

## 8.5 Testing on visually impaired users

Three experiments were performed: the testing of the vibration levels (earlier described in 8.2.1), the random-symbol-random-position experiment (as in 8.2.2.2, experiment #7) and the navigation scenarios (as in 8.4). The results are shown in Fig. 8-6, Fig. 8-7, Fig. 8-8 and Table 8-11, Table 8-12, Table 8-13. The average of the results for the sighted subject will be shown with a green bar.

A quick look on the results reveals that there are no systematic and major differences between the sighted and visually impaired users. Note here that subject 1, shows a systematic identification error during the frequencies identification experiment (Fig. 8-6) since vibration level #2 and #3 are hardly recog-

nized. The most possible explanation is we should give more time to him to get familiar with the frequencies. The argument is that at navigation scenarios experiment the identification is drastically improved (Table 8-13). On the contrary subject 2 shows frequency identification accuracy similar with the sighted users.

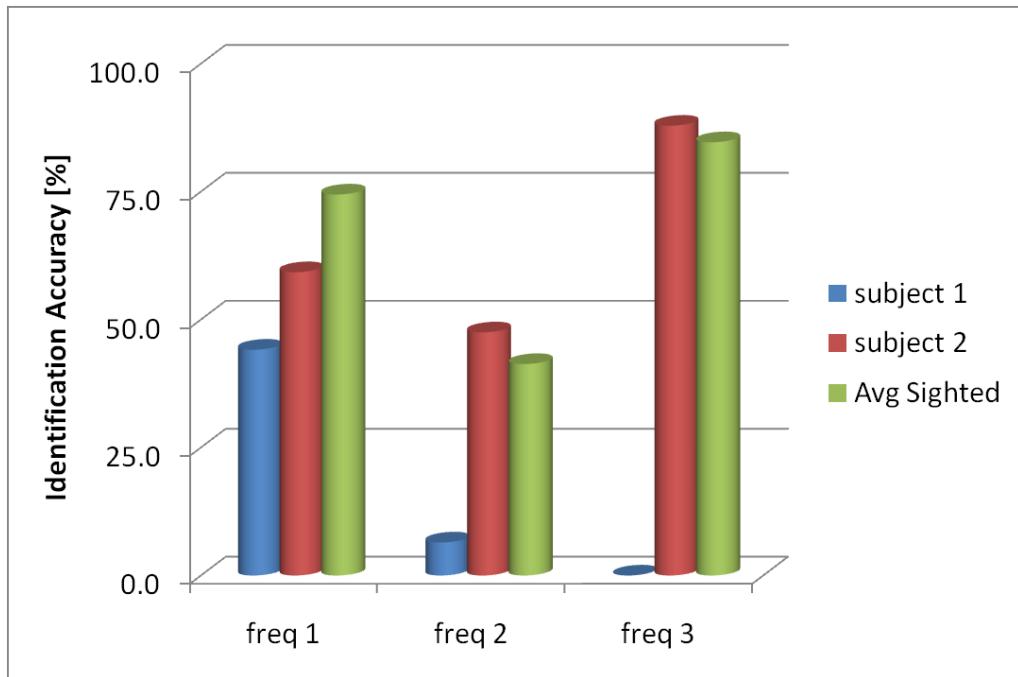


Fig. 8-6. Frequencies/Vibration levels identification accuracy (visually impaired users), similar to Fig. 8-1.

The position identification results (Fig. 8-7) show that the visually impaired users had a little more difficulty identifying the extreme left and right columns while at the symbol identification accuracy (Fig. 8-8), the percentages are very similar. The confusion matrix verifies the difficulty in distinguishing between symbols with one or two consecutive activated vibrating elements.

Regarding the navigation experiments we notice that the visually impaired users had difficulty distinguishing the various vibration levels (Table 8-

13) and this complies with the results of the frequency identification experiment. On the contrary, as seen in Table 8-12, the basic features of the patterns (open paths, empty patterns and non-zero frequencies) are identified with a very comparable accuracy. Fusing the results from the previous paragraphs and the oral remarks that the users made we can have a first evaluation of the system, presented in the next paragraph.

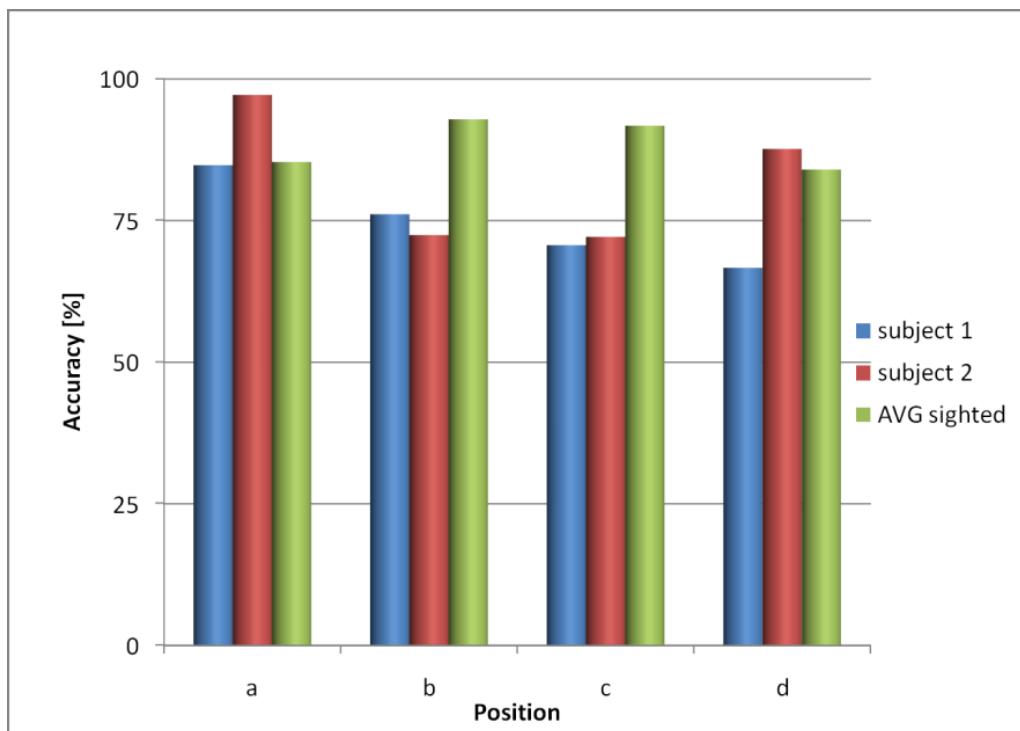


Fig. 8-7. Position identification accuracy for random symbol and position (visually impaired users), similar to Fig. 8-2.

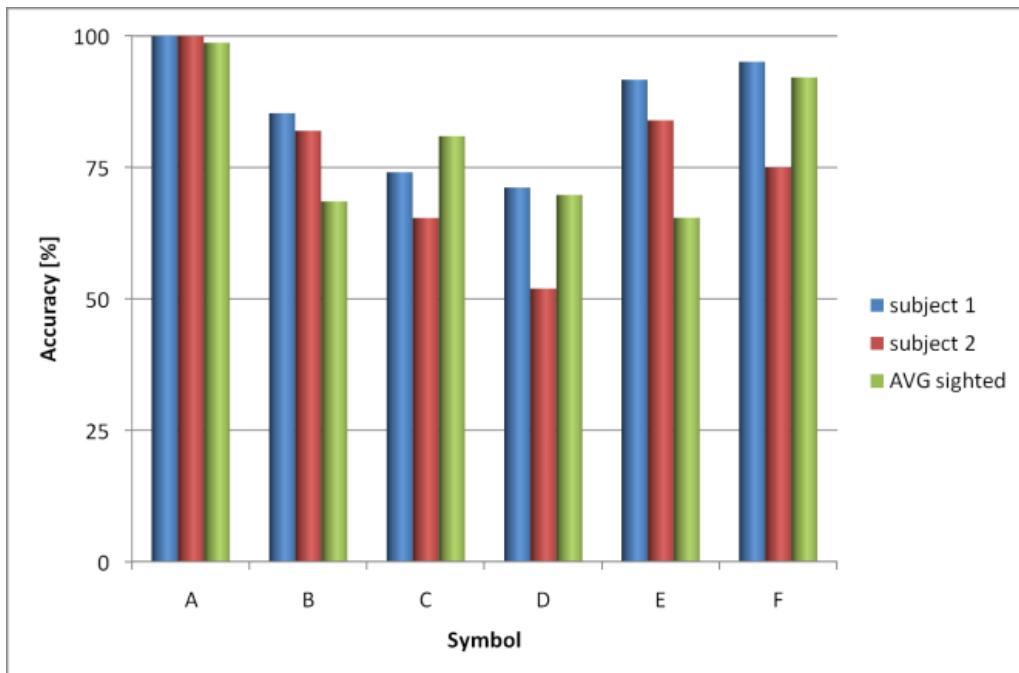


Fig. 8-8. Symbol identification accuracy in random position (visually impaired users), similar to Fig. 8-3 (shown on the right).

Table 8-11. Average confusion matrix for symbol identification (visually impaired), similar to Table 8-6 (shown on the right).

Actual pattern	Guessed pattern					
	A	B	C	D	E	F
A	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
B	0.0	<b>83.6</b>	<b>13.2</b>	2.5	0.7	0.0
C	1.0	<b>16.3</b>	<b>70.4</b>	2.9	0.0	9.5
D	1.0	0.0	1.0	<b>61.5</b>	<b>29.8</b>	6.7
E	0.0	2.7	0.0	8.6	<b>87.8</b>	0.9
F	0.0	0.0	4.5	9.7	0.7	<b>85.0</b>

Actual pattern	Guessed pattern					
	A	B	C	D	E	F
A	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
B	0.7	<b>68.6</b>	<b>28.3</b>	0.9	0.9	0.6
C	0.0	<b>10.4</b>	<b>81.2</b>	2.6	0.4	5.5
D	0.2	0.3	<b>2.4</b>	<b>69.8</b>	9.1	<b>18.3</b>
E	0.2	1.3	1.8	<b>30.2</b>	<b>65.7</b>	0.7
F	0.0	0.3	3.4	3.8	0.2	<b>92.3</b>

Table 8-12. Correct identification of: patterns without vibrations; open paths within patterns; non-zero frequencies (visually impaired users), similar to 0.

subject	empty patterns [%]	open paths [%]	open areas [%]	obstacles [%]	distance of obstacles [%]
1	100.0	93.2	86.7	87.3	31.0
2	100.0	94.1	97.3	86.7	41.3
Average	<b>100.0</b>	<b>93.7</b>	<b>92.0</b>	<b>87.0</b>	<b>36.2</b>
Average (sighted)	97.6	91.5	91.6	64.4	30.6
<b>Average (sighted+visually impaired)</b>	<b>98.8</b>	<b>92.6</b>	<b>91.8</b>	<b>75.7</b>	<b>33.4</b>

Table 8-13. Navigation scenarios: average confusion matrix for the different frequencies (visually impaired users), similar to Table 8-10 (shown on the right).

		guess						guess			
		0	1	2	3	actual	0	1	2	3	
actual	0	87.0	2.9	6.0	4.1	0	91.6	4.0	3.2	1.3	
0	1	8.6	<b>27.2</b>	<b>45.6</b>	<b>18.6</b>	1	<b>36.8</b>	<b>27.6</b>	<b>31.8</b>	<b>3.8</b>	
1	2	8.2	<b>12.4</b>	<b>44.4</b>	<b>35.0</b>	2	<b>34.1</b>	<b>15.0</b>	<b>35.8</b>	<b>15.1</b>	
2	3	0.0	0.0	8.2	<b>91.8</b>	3	<b>30.4</b>	4.6	<b>29.3</b>	<b>35.7</b>	

## 8.6 Evaluation

The differences between sighted and visually impaired people appear to be not systematic and relatively small. This is not a final conclusion but a hint for the future experiments.

We will emphasize on the results from the navigation scenarios experiments (Table 8-9, Table 8-12) since they provide an almost real-world testing of the system. Both user groups (sighted and visually impaired) performed really well in identifying the **open paths in a scene with average accuracy of 92.6%**. This is a very important result since the existence and the direction of open paths is the most crucial navigation information that that users need to be aware.

The identification of the vertical vibration symbols is equivalent to identifying the size and the shape of the obstacles in the scene. The correct identification of the vibration levels means that the distance of the obstacle is also identified. The experiments show that on average, **75.7% of the obstacles were identified but with a low 33.4% accuracy about their distance**. Since obstacle/ object detection is the second more crucial information during navigation this percentage is very promising.

During the experiments and their debriefing the users made various comments regarding the prototype:

- **Spatial acuity:** many users had difficulty distinguishing between consecutive vibrating elements and they proposed to have bigger distances between the elements.
- **Synchronization:** the frequencies of vibrating elements were not synchronized which means that e.g. the rising edges of two pulses that drive two motors with the same frequency, were not synchronous. This asynchrony was pointed out repeatedly. An example of a problem that can appear is that if two consecutive elements were vibrating in the same frequency they often gave the impression as if being one element vibrating at the double frequency.
- **Motors tactility:** due to the shape of the abdomen area, some times the motors did not make a proper contact with the users' abdomen. Many users often kept adjusting their torso to facilitate a better contact with the motors.

- **Frequencies:** most of the users mentioned that the vibration levels 1 and 2 are very similar and proposed to make them more distinguishable by increasing the frequency range.

# **Chapter 9: Research Contributions & New Experimental Design**

## **9.1 The new system/prototype (pub. 1, 2)**

The system/ prototype, outcome of this dissertation is a novel and **unique** assistive device for blind and visually impaired individuals. It is a wearable Electronic Travel Aid (ETA) that collects environmental data using the stereoscopic cameras, processes them and presents them to the user via a tactile interface. Its main purpose is to assist users towards their independent mobility, complementing their main mobility aid; usually the white cane or the guide dog. Currently the **Tyflos Navigator prototype is the ETA offering the most complete, appropriate (based on NRC specs) and unique collection of features and functionalities.**

The most crucial elements that affect navigation and can appear in an environmental scene are obstacles and open (safe) navigation paths. The obstacles are environmental or architectural obstructions in the navigation path. If they cannot be detected with the standard long cane techniques they are often called travel hazards. The prototype has an obstacle detection (and safe navigation paths detection) functionality but it provides a more overall sensation of the 3D environment of the travel scene including other elements that are of interest for the user such as: moving objects, people, objects of special interest such as a signs etc.

The custom built inexpensive stereoscopic vision system serves as the main sensor unit of the prototype. It gathers visual information and by using state of the art computer vision techniques we obtain depth data (correspond to the 3D structure) and detect motion activity areas (moving objects).

The 2D tactile display (i.e. vibration array) consists of 16 vibrating elements arranged in a four-by-four manner and is the main interface between the prototype and the system. The two dimensions (x,y) of the display, in combination with the varying frequencies of each vibrating element provide a new and appropriate interface for projecting processed and carefully selected information about the environment, including the 3D structure and moving objects.

Both the sensor (stereo cameras) and the interface unit (vibration array) are light, small in size and thus wearable and cosmetically accepted. The cameras are attached on conventional eye-glasses and the vibration array is attached on an elastic cloth worn below the outerwear on the abdomen.

The novel and new methodologies used are: frame stabilization and motion detection methods are followed by an also novel high-to-low resolution methodology that preserves the important navigation information as the resolution of the sensor's output images is decreased to accommodate the low resolution of the tactile array. Finally, a context-free tactile formal language is developed with symbols and words generated using continuous experimental feedback from the users.

**The following full-paper refereed publications support the new scientific contributions of the Tyflos prototype:**

- 1) D. Dakopoulos, N. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: a Survey", **IEEE Trans. Systems Man and Cybernetics, Part C** (in press).
- 2) D. Dakopoulos, S. K. Boddhu, N. Bourbakis, "A 2D Vibration Array as an Assistive Device for Visually Impaired", **7th IEEE International Conference on Bioinformatics and Bioengineering**, Vol. 1, 14-17 October 2007, Boston, MA, USA, pp.930-937.
- 3) D. Dakopoulos, N. Bourbakis, "Preserving Visual Information in Low Resolution Images During Navigation for Visually Impaired", In Proc. of the 1st **Int'l Conf. on PErvasive Technologies Related to Assistive Environments (PETRA '08)**, Vol. 282, 15-19 July 2008, Athens, Greece, pp. 1-6.
- 4) N. Bourbakis, A. Esposito, R. Keefer, D. Dakopoulos, "A Multimodal Interaction Scheme Between a Blind User and the Tyflos Assistive Prototype", Proceedings 20th **IEEE Int'l Conference on Tools with Artificial Intelligence (ICTAI '08)**, Vol. 2, 3-5 November 2008, Dayton, OH, USA, pp. 487-494.
- 5) N. Bourbakis and D. Dakopoulos, "The algebraic modeling of the VAL language", **Pattern Recognition Journal**, (under revision)
- 6) N. Bourbakis, S. Makrogiannis, D. Dakopoulos, "Detecting, Tracking and Representing Motion via Vibrations for Navigation of Visually Impaired", **IEEE Trans on Haptics**, tbs

## **9.2 Analysis of the new research contributions**

### **9.2.1 Vibration Array Language & Tactile Vocabulary (pub. 2,4)**

A context-free formal language called vibration array language (VAL) was developed to model the information presented on the vibration array. It is produced by a context-free formal grammar; the left-hand side of the rewriting rule is only a single non-terminal symbol. In other words, the non-terminal symbols are rewritten without regard to the context in which they occur.

Its symbols are column-type and the string of symbols in any given moment on the vibration array consist a VAL word. For an  $m \times n$  vibration array there can exist  $2^m$  distinct symbols (terminating) and  $(2^m)^n$  distinct words. For the current prototype's  $4 \times 4$  vibration array, there can be 16 symbols and 65,536 words.

A major characteristic of the VAL is its evolving nature; that not all words are valid because its vocabulary is defined experimentally and is adjusted to the user's needs. The application of the special vertical and horizontal rules discussed in earlier chapter produces a reduced-size vocabulary of 298 words and more work is expected to be done towards the definition of the optimum VAL vocabulary by redefining those rules, removing some of them and adding new ones. The evolving characteristic of the VAL is that those are not yet finalized but they will continue to change using the feedback from the users.

The final goal will be to create a vocabulary that will optimize the user's perception of the environment and that means a balance between environmental information and perception accuracy. This is to be defined experimentally.

### **9.2.2 Motion detection (6)**

Discussions with visually impaired users revealed that moving objects (e.g. moving people and vehicles) are very important for their mobility safety and thus is necessary to be early detected (note: an early detection is necessary even before it gets in the detection range of the stereo vision sensor unit).

A pre-processing step for eliminating vertical video frames movements

due to the human gait is necessary. This is being done by developing a technique based on the Radon transform and statistical “similarity” measures (e.g. sum of squares differences).

The proposed **new motion detection methodology** is based on the argument that moving areas in successive video frames are diffused more than the static ones. The partial differential equation describing anisotropic diffusion, proposed by Perona, was extended by inserting the variable of time; three successive frames. A new pixel-based motion activity measure (PMA) was defined and Parzen kernels were used to eliminate statistical variabilities. Finally, a standard watershed (minima piercing and flooding) segmentation method segments the moving areas. Finally, the experimental results validate the efficiency of both the stabilization and motion detection methods.

### **9.2.3 High-to-Low resolution (pub. 3, 6)**

The output of the computer vision methods used (motion detection, stereoscopic vision) is high-resolution images (in our case,  $352 \times 264$  pixels) and the vibration array has a low resolution of  $4 \times 4$  vibrating elements. In our case the output depth and motion activity images (mentioned as image-of-distances and image-of-interest correspondingly) contain some information if using standard down-sampling methods it may be reduced or eliminated.

The novel developed method takes into consideration navigation criteria: distances of obstacles, open/ safe navigation paths, moving objects and objects of interest. A geometrical study of the wearable camera sensor unit detects any open paths in the scene (wider than a predefined width) and eliminates the

issue of the ground as being represented as an obstacle. The standard pyramidal reduction ( $2 \times 2$  kernel) was modified so that the parsing kernel for every pyramidal preserves proximity obstacles, safe paths and objects of interest.

In order to evaluate if this approach produces better results than classical image reduction methods, we used three probabilistic measurements (entropy, complexity, angular second moment) to model the High-to-Low information representation. From this evaluation we concluded that the information reduction by preserving important and information of interest has generated better results than the classical pyramidal method. **This was also a new piece of scientific knowledge generated during our studies.**

#### *9.2.4 The synergy of the methods*

An additional unique contribution to new sciences is the synergistic co-operation of these three methodologies mentioned above as one integrated method. In particular, the VAL language constrained by technological limitations relevant to its resolution has triggered the development of a **new approach-solution** to the blind's navigation problem. More specifically, the low resolution of the actual vibration array has triggered the study of the representation and preservation of important and users-selected information at high resolution levels (images) to the low levels of resolution of the vibration array. This challenging problem (High-to-Low with preservation of information) became more difficult and unique since motion detection was involved in it. Thus, we had to developed our **new motion detection technique** at the high resolution level and appropriately preserve it at the low resolution level. Finally, in order

to put all together, we developed a **new High-to-Low representation scheme** with preservation capabilities of selected pieces of information, such as motion, obstacle in close range, and open space.

The overall synergy has given to us a unique prototype methodology for further scientific studies, such as (i) feedback from the users testing the prototype on a series of navigation scenarios, (ii) using the feedback from the users to appropriately model and improve the vibration array resolution for better representation of information, (iii) **creating new knowledge** through the testing process about the 3D sensation of the environment for blind and visually impaired using low resolution 3D representations.

### 9.3 A new experimental design

The experiments performed and presented in the last chapter are not extensive but they provide some early statistical results and first feedback regarding the validity of our prototype's design with an emphasis on its tactile interface. Future experimental work can be organized in three hierarchical parts with each one requiring the previous to be completed:

1. Interface testing and modeling

The goal of the first set of experiments is to validate the hypothesis that the 2D vibration array is a sound interface for delivering spatio-temporal information. They will be performed off-line with synthetic scenarios with focus on answering more technical questions related with the interface:

- Increase of the **motors' tactility** by creating a new thinner vest, using different material (elastic vs. non-elastic) or by attaching motors directly

on the users skin.

- Define **optimum distance** between two consecutive vibrating elements considering the 2D characteristic of the array and the specification of the array.
- Compare of **synchronous or asynchronous frequencies**.
- Define of the optimum **range of vibration frequencies** in terms of perception.

## 2. Vocabulary testing and modeling

The goal of the second set of experiments is to validate the hypothesis that an appropriate reduced-size VAL vocabulary is more appropriate for delivering 3D environmental information as opposed to the raw camera/ high-to-low output. Here are the steps.

- Define the **optimum set of VAL symbols**.
- Define the **optimum set of horizontal rules** for the given VAL symbols.
- Off-line comparison for **obstacles and open paths identification accuracy** with real-word navigation scenarios using reduced-size and full-size vocabulary.
- **Revisit** (number and type of VAL symbols, horizontal rules) reduced-size vocabulary and compare again.

## 3. Prototype testing

The goal of the third set of experiments is to validate the hypothesis that the prototype can improve independent mobility. The participants will be asked to navigate in controlled in-lab environments using different combina-

tions of mobility aids: Tyflos with cane, only Tyflos and only cane. The following parameters will be measured and compared for the different aids combinations to prove the hypothesis:

- Obstacles identification accuracy.
- Open paths identification accuracy.
- Travel speed and time.
- Additional feedback information will be extracted concerning other software or hardware parts the prototype such as the computer vision system (e.g. cameras angle, stereo correspondence algorithm, speed).

### 9.3.1 *Objectives*

Due to the complexity of the prototype, the proposed experimental design will be a step towards the first experiments and some fundamental experimental research questions (the **hypotheses** in every experimental set) will be investigated.

The most unique hardware component of the Tyflos prototype is the 2D vibration array interface. As far as we know it is the first time that a 2D tactile display is used to provide information about the the structure of the 3D environment as well as other information useful during navigation. The novelty is that it delivers information dynamically in time (following the user's head movement), exploiting spatial (2D) and temporal (vibration levels) characteristics. The modeling of the information presented on the array is via the Vibration Array Language using the VAL vocabulary which must both have to be defined experimentally.

Taking the above into consideration our approach aims to define the characteristics of the VAL and the current vibration array design that optimize usability. In that way we answer questions that have been unanswered by the first set of experiments and regard the usability of the Tyflos 2D vibration array interface. More details will be presented in the next paragraph.

### 9.3.2 *Variables*

The second step for every experimental design, after the definition of the objectives, is to “...specify the variables in a way that no other definitions will be necessary during all the steps of the project” [94]. For our approach we focus on the interface related variables which are:

- Motors' tactility.
- Resolution of the 2D ( $m \times n$ ) vibration array.
- Motors density: for a 2D ( $m \times n$ ) array the motors density is  $1/ m * n$ .
- Stimulus Onset Asynchrony (i.e. synchronous or asynchronous frequencies)
- Number of frequencies.
- Frequencies range.
- BD (Burst Duration).
- VAL symbols
- VAL horizontal rules

The motors tactility is a specification of the design of the current vibration array vest and it is a controlled variable. For this design the motors will be placed directly on the user's skin with adhesive tape in order to achieve maxi-

mum tactility. As soon as the design parameters are determined the motors will be placed again on the vibration array vest.

For the current experimental design, no changes in the current vibration array design will be performed, thus the following variables are controlled:

- The array resolution and density: 4×4 square lattice with 5.5cm center-to-center distance of the motors (**0.033 motors/cm<sup>2</sup>**)
- SOA: asynchronous; **random**.
- Frequencies range; **1.25-10.5Hz**
- Duty cycle (i.e. burst duration; specified by the frequency).

The independent variables are the following:

- The **3 non-zero frequency levels**.
- The **VAL symbols**; form of activated motors in the same column.
- The **VAL vocabulary horizontal rules**; forms of 2 or 3 consecutive VAL symbols.

The dependent variables are the **frequency identification accuracy**; for each frequency and the **VAL symbol identification accuracy**. The latter can be measured indirectly by measuring the identification accuracy for the 3D environmental structure when using the VAL vocabulary as opposed to not use a reduced size vocabulary. The environmental characteristics are:

- **Open paths** identification accuracy.
- **Open space** (i.e. more than one consecutive paths) identification accuracy.
- **Obstacles identification accuracy** (i.e. x-y and frequency information)

Four experiments are proposed:

### 9.3.3 Experiment 1: Definition of vibration frequencies

As seen in the first experimental results, the users had many difficulties identifying the different frequencies (vibration levels). The percentages were close to 30% which is close to the random guess accuracy, although we know that the major confusion was in distinguishing similar frequencies. This proposed experiment aims identify the best triplet of frequencies for the current interface design but the technique could be extended for different designs.

Table 9-1. Experiment 1: layout

<b>Hypothesis</b>	The best perception comes when the frequencies are more distributed and for our hardware design this is: 1.3Hz, 5.9Hz and 10.5Hz.
<b>Independent variables</b>	Frequency levels.
<b>Dependent variables</b>	Frequency identifications accuracies for the three frequency levels.
<b>Tactic</b>	Create different possible triplets of vibration frequencies and by checking the identification accuracy for the different frequency, decide which set is the most appropriate
<b>Other goals</b>	Reveal general frequency accuracy patterns.

A literature review [45], [48] reveals that the spatial acuity of the abdomen area varies slightly on the position: the navel areas has the higher acuity compared to the left or right abdomen, while there is no significant different on the vertical direction.

For a given frequency step  $f_{step}$ , range  $r = f_{max} - f_{min}$  and number of levels N, there would be  $D = \frac{r}{f_{step}}$ . With the constrain,  $f_1 < f_2 < f_3$  the total sets of frequencies is  $N_{sets} = \frac{D!}{N!(D-N)!}$

In our case  $r = 10.5 - 1.3 = 10.2 \approx 10\text{Hz}$ , the number of levels is  $N = 3$ ,

we decide on  $f_{step} = 1.0\text{Hz}$  and so we have a total of 84 possible frequency sets.

We select three consecutive motors placed vertically on the left (or right) abdomen area and a random set of frequencies is sent and ask the users to identify the frequencies of the motors (relative to each other).

The average results will be plotted in three dimensions (as a cube) where the axes are the frequencies ( $f_1, f_2, f_3$ ) and the intensity of every 3D point will be the average frequency accuracy. A third order non-linear fit will create surfaces that correspond to the same frequency accuracy. The study of this 3D contour plot is a good visual representation of the experimental data and it will possibly reveal:

- Frequency set(s) with the best accuracy (highest intensity)
- The effect of frequencies distribution i.e. testing the hypothesis. Do spread out frequencies results better accuracy and if yes, in which frequency areas?
- The frequency sensitivity (e.g. dense contours in the higher frequencies would mean that the small frequency changes affect performance)

#### 9.3.4 *Experiment 2: Definition of VAL symbols*

Table 9-2. Experiment 2: layout

<b>Hypotheses</b>	Some vertical VAL symbols have overlapping meaning and the elimination of some will increase the identification accuracy.
<b>Independent variables</b>	Column-type symbols.
<b>Dependent variables</b>	Symbol identification accuracy.
<b>Tactic</b>	Reduce symbols that are usually confused with each other and compare the users accuracy in identifying basic environmental structure before and after this reduction.

The goal is to create a VAL vocabulary, a set of VAL words that we can assure that the users understand precisely. During the initial experimental part, six symbols (including the empty symbol) were chosen taking into consideration some navigational issues but now all possible VAL symbols will be tested.

The users will be presented all the possible 16 symbols in random [92] positions and their identification accuracy will be measured. A  $16 \times 16$  symbols confusion matrix will reveal which (if any) symbols have this overlapping identification. The kappa coefficient or other correlation coefficient [x] will be used to measure the importance of incorrect identification and thus eliminate symbols with low statistical significance.

$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$ , where  $P(A)$  is the correct identification percentage of that symbol and  $P(E)$  is expected percentage by chance which is  $1/16=6.25\%$ . For example, if symbol A is misidentified with symbol B with  $\kappa < 0.7$  this corresponds to low statistic correlation and the symbol with the less identification accuracy will be eliminated (0.7 is the standard experimental value used).

Finally, a repeated measures paired t-test [94] will be performed to test hypothesis that the VAL symbols improve the average symbols identification accuracy.

### 9.3.5 *Experiment 3: Definition of horizontal rules*

The VAL symbols defined in experiment 2 are presented to the users in doublets in random consecutive columns on the array. For N symbols (excluding the empty symbol) there can be  $N^2$  possible doublets. An  $N^2 \times N^2$  confusion matrix will be created and then the method as described in experiment 3 will

be applied using the kappa coefficient to define which symbol doublets have overlapping identification.

Table 9-3. Experiment 3: layout

<b>Hypothesis</b>	A set of horizontal rules will improve the symbol identification accuracy.
<b>Independent variables</b>	Doublets of VAL symbols
<b>Dependent variables</b>	Symbol identification accuracy.
<b>Tactic</b>	Present various groups of symbols (couples or triplets). Measure the symbol identification accuracy and find groups that are identified similarly. Create the horizontal rules.

Similarly with experiment 2, a repeated measures paired t-test can be performed to test the hypothesis that horizontal rules improve the VAL symbols identification.

### 9.3.6 *Experiment 4: VAL vocabulary testing*

Table 9-4. Experiment 4: layout

<b>Hypothesis</b>	The VAL vocabulary improves the identification of 3D environmental characteristics.
<b>Independent Variables</b>	- Real-world navigation scenarios: VAL words
<b>Dependent Variables</b>	- Important 3D environmental characteristics: - Open paths identification accuracy. - Open space (i.e. more than one consecutive paths) identification accuracy. - Obstacles identification accuracy (i.e. x-y and frequency information)

The final experiment will involve testing with real-world indoor navigation scenarios similar to the last chapter. All the Tyflos methodologies will be applied off-line (stereo, motion detection, high-to-low, pattern matching) and a selection of frames will be made that includes all the VAL symbols and all poss-

ible elements that can appear during navigation (as in 7.2).

Finally, to test the hypothesis, the multiple dependent variables require the use of repeated measures Multivariate Analysis Of VAriance (MANOVA) [92].

### 9.3.7 *Conclusions*

The proposed experimental design aims to answer questions regarding the usability of the 2D vibration array and the VAL vocabulary. A careful experimentation would provide feedback and confidence about the usability of the whole Tyflos prototype

The first experiments performed in the previous chapter “aim to define a good starting point in the experimental space and include all the previously identified variables” [94] while the proposed design of this chapter aims to identify some of the important variables. Despite the complexity of the prototype and the strong correlation of the various variables, one-at-a-time-strategies were used avoiding advance multivariate techniques which will be used for final experimental stages.

## REFERENCES

- [1] National Federation of the Blind: <http://www.nfb.org/>
- [2] American Foundation for the Blind: <http://www.afb.org/>
- [3] International Agency for Prevention of Blindness: <http://www.iapb.org/>
- [4] Blasch B., Wiener W. and Welsh, R., Foundations of Orientation and Mobility, 2<sup>nd</sup> Edition, American Foundation for the Blind, New York, NY, 1997.
- [5] D. Dakopoulos, N. Bourbakis, "A Comparative Survey on Wearable Systems for Blinds' Navigation, 1<sup>st</sup> International IEEE-BAIS Symposium on Assistive Technologies, 16 April 2005, Dayton, OH, USA, pp. 3-12.
- [6] D. Dakopoulos, N. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids: a Survey", IEEE Trans. Systems, Man and Cybernetics, 2009.
- [7] T. Ifukube, T. Sasaki, C. Peng, "A Blind Mobility Aid Modeled after Echolocation of Bats", IEEE Trans. On Biomedical Engineering, Vol. 38, No. 5, May 1991, pp. 461-465.
- [8] S. Shoval, J. Borenstein, Y. Koren, "Mobile robot obstacle avoidance in a computerized travel aid for the blind", Proceedings of the 1994 IEEE Robotics and Automation Conference, San Diego, CA, USA, May 8-13, 1994, pp. 2023-2029.
- [9] P. B. L. Meijer, "An experimental System for Auditory Image Representations", IEEE Transactions on Biomedical Engineering, Vol.39, No2, February 1992. <http://www.seeingwithsound.com/>
- [10] A. Hub, J. Diepstraten, T. Ertl, "Design and development of an indoor navigation and object identification system for the blind", ACM SIGACCESS Accessibility and Computing, Issue 77-78 , Sept. 2003 - Jan. 2004.
- [11] D. Aguerrevere, M. Choudhury, A. Barreto, "Portable 3D Sound / Sonar Navigation System for Blind Individuals", Second LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCET 2004), 2-4 June 2004, Miami, FL, USA.
- [12] J. L. González-Mora, A. Rodríguez-Hernández, L. F. Rodríguez-Ramos, L. Díaz-Saco, N. Sosa, "Development of a new space perception system for blind people,

- based on the creation of a virtual acoustic space”, Technical Report <http://www.iac.es/proyect/eavi>
- [13] G. Sainarayanan, R. Nagarajan, and S. Yaacob, “Fuzzy image processing scheme for autonomous navigation of human blind”, *Appl. Soft Comput.* 7, 1 (Jan. 2007), pp. 257-264.
  - [14] R. Audette, J. Balthazaar, C. Dunk, and J. Zelek, “A stereo-vision system for the visually impaired,” Tech. Rep. 2000-41x-1, School of Engineering, University of Guelph, Guelph, ON, Canada.
  - [15] I. Ulrich, J. Borenstein, “The Guidecane – Applying Mobile Robot Technologies to Assist the Visually Impaired People”, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol.31, No2, March 2001, pp. 131-136.
  - [16] S. Meers, K. Ward, “A Substitute Vision System for Providing 3D Perception and GPS Navigation via Electro-tactile Stimulation”, 1<sup>st</sup> International Conference on Sensing Technology, November 21-23, 2005, Palmerston North, New Zealand.
  - [17] K. Ito, M. Okamoto, J. Akita, T. Ono, I. Gyobu, T. Tagaki, T. Hoshi, Y. Mishima, “CyARM: an Alternative Aid Device for Blind Persons”, CHI05, Portland, OR, USA, April 2-7, 2005, pp. 1483-1488.
  - [18] M. Bouzit, A. Chaibi, K. J. De Laurentis, C. Mavroidis, “Tactile Feedback Navigation Handle for the Visually Impaired”, 2004 ASME International Mechanical Engineering Congress and RD&D Expo, November 13-19, 2004, Anaheim, CA, USA.
  - [19] C. Shah, M. Bouzit, M. Youssef, Vasquez L., “Evaluation of RU-Netra – Tactile Feedback Navigation System for the Visually Impaired”, International Workshop on Virtual Rehabilitation, 2006, New York, NY, USA, pp. 71-77.
  - [20] L. A. Johnson and C. M. Higgins, "A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution" in Proceedings the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006, New York, NY, USA, pp 6298-6292.
  - [21] S. Cardin, D. Thalmann, F. Vexo. “A wearable system for mobility improvement of visually impaired people”. *Vis. Comput.* 23, 2 (Jan. 2007), pp. 109-118.

- [22] N.G. Bourbakis, D. Kavraki, "Intelligent assistants for handicapped people's independence: case study, 1996 IEEE International Joint Symposia on Intelligence and Systems (IJSIS '96), 1996.
- [23] N. Bourbakis, D. Kavraki, "A 2D Vibration Array for Sensing Dynamic Changes and 3D Space for Blinds' Navigation", 5<sup>th</sup> IEEE Symposium on Bioinformatics and Bioengineering, 19-21 October 2005, Minneapolis, MN, USA, pp. 222-226.
- [24] D. Dakopoulos, S. K. Boddhu, N. Bourbakis, "A 2D Vibration Array as an Assistive Device for Visually Impaired", 7<sup>th</sup> IEEE International Conference on Bioinformatics and Bioengineering, Vol. 1, 14-17 October 2007, Boston, MA, USA, pp.930-937.
- [25] D. Dakopoulos, N. Bourbakis, "Preserving Visual Information in Low Resolution Images During Navigation for Visually Impaired", In Proc. of the 1st Int'l Conf. on PErvasive Technologies Related to Assistive Environments (PETRA '08), Vol. 282, 15-19 July 2008, Athens, Greece, pp. 1-6.
- [26] N. Bourbakis, "Sensing surrounding 3-D space for navigation of the blind - A prototype system featuring vibration arrays and data fusion provides a near real-time feedback." IEEE Engineering in Medicine and Biology Magazine, v. 27, Issue 1, 2008, p. 49-55.
- [27] M. Adjouadi, "A man-machine vision interface for sensing the environment", Journal of Rehabilitation Research and Development Vol.29, No2, 1992, pp. 57-56.
- [28] D. Yuan, R. Manduchi, "A Tool for Range Sensing and Environment Discovery for the Blind", Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (Cvprw'04) Vol.3, June 27 - July 02, 2004, CVPRW. Washington, DC, USA.
- [29] Bay Advanced Technologies Ltd., <http://www.batforblind.co.nz/index.htm>
- [30] BestPluton World Cie, <http://www.bestpluton.com/>
- [31] GDP Research, Australia, <http://www.gdp-research.com.au/>
- [32] Nurion-Raycal, <http://www.lasercane.com/>
- [33] Sound Foresight Ltd., <http://www.soundforesight.co.uk/>

- [34] T. C. Davies, C. M. Burns, S. D. Pinder, "Mobility Interfaces for the Visually Impaired: What's missing?", CHINZ '07, Vol. 254, July 2-4, 2007, Hamilton, New Zealand, pp. 41-47.
- [35] R. D. Easton, "Inherent Problems of Attempts to Apply Sonar and Vibrotactile Sensory Aid Technology to the Perceptual Needs of the Blind", Optometry and Vision Science, Vol. 69, No. 1, 1992, pp. 3-14.
- [36] J. M. Loomis, R. L. Klatzky, R. G. Gooledge, "Navigating Without Vision: Basic and Applied Research", Optometry and Vision Science, Vol. 78, No. 5, 2001, pp. 282-289.
- [37] J. Sanchez, M. Elias, "Guidelines for Designing Mobility and Orientation Software for Blind Children", INTERACT 2007, LNCS 4662, Part 1, 2007, pp. 375-388.
- [38] N. N. Bourbakis, A. Esposito, R. Keefer, D. Dakopoulos, "A Multimodal Interaction Scheme Between a Blind User and the Tyflos Assistive Prototype", In Proc. of the 20th IEEE Int'l Conference on Tools with Artificial Intelligence (ICTAI '08), Vol. 2, 3-5 November 2008, Dayton, OH, USA, pp. 487-494.
- [39] Haas, P. J., Stochastic Petri Nets: modeling, stability, simulation, Springer, New York, 2002.
- [40] Young, Steve, et. al., *The HTK Book*, Cambridge University Engineering Department, 2006.
- [41] V.G. Chouvardas, A. N. Miliou, M. K. Hatalis, "Tactile Displays: Overview and recent advances", Displays, Vol. 29, 2008, pp. 185-194.
- [42] Handy Tech Elektronik GmbH, <https://www.handytech.de/>
- [43] GW Micro, <http://www.gwmicro.com/SyncBraille/>
- [44] J. B. F. Van Erp, "Guidelines for the Use of Vibro-Tactile displays in Human Computer Interaction", Eurohaptics 2002, Edinburgh, Ed. by S. A. Wall et al., pp. 18-22.
- [45] J. B. F. Van Erp, "Vibrotactile Spatial Acuity on the Torso: Effects of Location and Timing Parameters", 1<sup>st</sup> Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005, pp. 80-85.
- [46] J. B. F. Van Erp, "Absolute localization of vibrotactile stimuli on the torso", Perception & Psychophysics, Vol. 70, No. 6, 2008, pp. 1016-1023.

- [47] J. B. F. Van Erp, "Cross-modal visual and vibrotactile tracking", *Applied Ergonomics*, Vol. 35, 2004, pp. 105-112.
- [48] R. W. Cholewiak, A. A. Collins, J. C. Brill, "Spatial Factors in Vibrotactile Pattern Perception", *Eurohaptics 2001*, 1-4 July 2001, Birmingham, England.
- [49] Wilentz, J. S., *The Senses of Man*, Thomas Y Crowell Company, NY, 1968.
- [50] Gescheider, G. A., Temporal relations in cutaneous stimulation, in *Cutaneous communication systems and devices*, F. A. Geldard, Austin, TX: The Psychonomic Society.
- [51] L. Petrosino, D. Fucci, "Temporal resolution of the aging tactile sensory system", *Perceptual and Motor Skills*, Vol. 68, No. 1, 1989, pp. 288-290.
- [52] Weinstein, S., Intensive and extensive aspects of tactile sensitivity as a function of body-part, sex and laterality, in: *The Skin Senses*, edited by D. R. Kenshalo, Springfield, C. Thomas, pp. 195-218.
- [53] K. O. Johnson, J. R. Philips, "Tactile spatial resolution. I. Two point discrimination, gap detection, grating resolution and letter recognition", *Journal of Neurophysiology*, Vol. 46, No. 6, 1981, pp. 1177-1191.
- [54] Shcrope, M., *Simply Sensational*, New Scientist, Issue 2293, June 2, pp. 30-33, 2001.
- [55] Van Erp, J. B. F., *Tactile Navigation Display*, in Brewster, S., Murray-Smith, R., (Eds.), *Haptic Human Computer Interaction*, Lecture Notes in Computer Science, Vol. 2058, Springer, Berlin, pp. 165-173, 2001.
- [56] Intel's Open Source Computer Vision Library (OpenCV):  
<http://www.intel.com/technology/computing/opencv/>
- [57] Bradski, G. R., Kaehler, A., *Learning OpenCV*, O'Reilly Media, Inc., 2008.
- [58] Creative Labs Inc., <http://www.creative.com/>
- [59] Forsyth, D. A., Ponce, J., *Computer Vision: A Modern Approach*, Prentice Hall, 2002.
- [60] S. Birchfield and C. Tomasi, "Depth discontinuities by Pixel-to-Pixel Stereo", IEEE International Conference on Computer Vision, Bombay, India, p. 1073–1080, 1998.
- [61] S. Birchfield and C. Tomasi. "A Pixel Dissimilarity Measure That Is Insensitive To Image Sampling", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 4, 401–406, 1998.

- [62] Jinlong Machinery and Electronics Co., Ltd., <http://www.vibratormotor.com/>
- [63] Maxim Integrated Products, Inc., <http://www.maxim-ic.com/>
- [64] Dallas Semiconductors, Application Note 139, "Controlling a DS1803 Digital Potentiometer Using an 8051 Microprocessor to Generate 2-Wire Signals".
- [65] PJRC.COM, LLC., <http://www.pjrc.com/>
- [66] N. Bourbakis, C. Alexopoulos, "A fractal-based image processing language: formal modeling", Pattern Recognition, Vol. 32, 1999, pp. 317-338.
- [67] Hopcroft, J.E. and Ulmann, J. D., *Formal Languages and their Relation to Automata*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1969.
- [68] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 6, pp. 583-597, June 1991.
- [69] D. Wang, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 5, pp. 539-546.
- [70] S.-Y. Chien, S.-Y. Ma and L.-G. Chen, "Efficient Moving Object Segmentation Algorithm Using Background Registration Techniques", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 7, pp. 577-586, July 2002.
- [71] B. K. P. Horn, *Robot Vision*, the MIT Press, 1986.
- [72] J. Shi and J. Malik, "Motion Segmentation and Tracking Using Normalized Cuts", Proceedings of International Conference on Computer Vision (ICCV 1998), pp. 1154-1160, 1998.
- [73] S. Makrogiannis and N. Bourbakis, "Motion Analysis with Application to Assistive Vision Technology". 16<sup>th</sup> International Conference on Tools with Artificial Intelligence (ICTAI 2004), pp. 344-352, 15-17 Nov. 2004.
- [74] S. Alliney and C. Morandi, "Digital image registration using projections", IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 8, no. 2, pp. 222-233, 1986.
- [75] S. C. Cain, M.M. Hayat, and E. E. Armstrong, "Projection-based image registration in the presence of fixed-pattern noise", IEEE Transactions on Image Processing, vol. 10, no. 12, pp. 1860-1872, 2001.

- [76] P. Perona, J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629-639, July 1990.
- [77] E. Parzen, "On estimation of a probability density function and mode", Ann. Math. Stat. vol. 33, pp. 1065-1076, 1962.
- [78] Pavlidis, T., *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD, USA, 1982.
- [79] N. Bourbakis, A. Klinger, "A Hierarchical Picture Coding Scheme", Pattern Recognition, Vol.22, No. 3, 1989, pp. 317-329.
- [80] N. Bourbakis, "Interpretation of Brain Patterns Associated with 3D Vibration Sequences", International Biomedical Engineering Conference, also ATRC-TR, 2007.
- [81] N. Bourbakis, S. Makrogiannis, D. Dakopoulos, "Detecting, Tracking and Representing Motion via Vibrations for Visual Impaired Navigation", IEEE Trans. SMC.
- [82] N. Bourbakis. "A traffic priority language for collision-free navigation of autonomous mobile robots in dynamic environments", IEEE Transaction on Systems, Man, and Cybernetics Part B - Cybernetics, vol. 27, Issue 4, 1997, p. 573-587.
- [83] Pitas, I., *Digital Image Processing Algorithms and Applications*, Wiley, New York, NY, USA, 2000.
- [84] Jain A. K., *Fundamentals of Digital Image Processing*, Prentice Hall Inc., Englewood Cliffs, NJ, 1989.
- [85] Haralick, R. M., Shanmugam, K., Dinstein, I., "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC 3, No. 6, Nov 1973.
- [86] D. Dakopoulos, N. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: a Survey", IEEE Trans. Systems Man and Cybernetics Part C (in press).
- [87] D. Dakopoulos, N. Bourbakis, "High to Low Mapping and Vibration Array Language", Pattern Recognition Journal (tbs)

- [88] N.G. Bourbakis, D. Kavraki, "Intelligent assistants for handicapped people's independence: case study, 1996 IEEE Int'l Joint Symposia on Intelligence and Systems (IJSIS '96), 1996.
- [89] Theodoridis, S., Koutroumbas, K., *Pattern Recognition*, 3<sup>rd</sup> Edition, Academic Press, 2006.
- [90] L. A. Jones., B. Lockyer, E. Piateski, "Tactile display and vibrotactile pattern recognition", Advanced Robotics, Vol. 20, No.12, 2006, pp.1359-1374.
- [91] Cozby, P.C., *Methods in Behavioral Research*, 9<sup>th</sup> Ed., McGraw Hill, 2006.
- [92] Ryan, T. R., *Modern Experimental Design*, Wiley, Hoboken, NJ, 2007.
- [93] Eberts, R. E., *User Interface Design*, Prentice Hall, Englewood Cliffs, NJ, 1994.
- [94] Diamond, W. J., *Practical Experiment Designs*, 3rd Ed., Wiley, New York, NY, 2001.