3

Vision Based Obstacle Detection Module for a Wheeled Mobile Robot

Oscar Montiel, Alfredo González and Roberto Sepúlveda
Centro de Investigación y Desarrollo de Tecnología Digital del Instituto Politécnico Nacional. México

1. Introduction

Navigation in mobile robotic ambit is a methodology that allows guiding a mobile robot (MR) to accomplish a mission through an environment with obstacles in a good and safe way, and it is one of the most challenging competence required of the MR. The success of this task requires a good coordination of the four main blocks involved in navigation: perception, localization, cognition, and motion control. The perception block allows the MR to acquire knowledge about its environment using sensors. The localization block must determine the position of the MR in the environment. Using the cognition block the robot will select a strategy for achieving its goals. The motion control block contains the kinematic controller, its objective is to follow a trajectory described by its position (Siegwart & Nourbakhsh, 2004). The MR should possess an architecture able to coordinate the on board navigation elements in order to achieve correctly the different objectives specified in the mission with efficiency that can be carried out either in indoor or outdoor environments.

In general, global planning methods complemented with local methods are used for indoor missions since the environments are known or partially known; whereas, for outdoor missions local planning methods are more suitable, becoming global planning methods, a complement because of the scant information of the environment.

In previous work, we developed a path planning method for wheeled MR navigation using a novel proposal of ant colony optimization named SACOdm (Simple Ant Colony Optimization with distance (d) optimization, and memory (m) capability), considering obstacle avoidance into a dynamic environment (Porta et al., 2009). In order to evaluate the algorithm we used virtual obstacle generation, being indispensable for real world application to have a way of sensing the environment.

There are several kind of sensors, broadly speaking they can be classified as passive and active sensors. Passive sensors measure the environmental energy that the sensor receives, in this classification some examples are microphones, tactile sensors, and vision based sensors. Active sensors, emit energy into the environment with the purpose of measuring the environmental reaction. It is common that an MR have several passive and/or active sensors; in our MR, for example, the gear motors use optical quadrature encoders, it uses a high precision GPS for localization, and two video cameras to implement a stereoscopic vision system for object recognition and localization of obstacles for map building and map reconstruction.

This work presents a proposal to achieve stereoscopic vision for MR application, and its development and implementation into VLSI technology to obtain high performance computation to improve local and global planning, obtaining a faster navigation by means of reducing idle times due to slow computations. Navigation using ant colony environment is based on map building and map reconfiguration; in this model, every ant is a virtual MR. The MR system, composed by the MR and the global planner in the main computer, see Fig 1, has the task to construct the map based on a representation of the environment scene, avoiding the use of Landmarks to make the system more versatile. The MR stereo vision transforms the visual information of two 2D images of the same scene environment into deep measure data. Hence, the MR sends this data via RF to the global planner in the main computer; this data is a 3D representation of the MR scene environment and its local position and orientation. By this way, the optimal path in the environment is constantly updated by the global planner.
The MR stereo vision has the advantage, with respect to other navigation techniques, that depth can be inferred with no prior knowledge of the observed scene, in particular the scene may contain unknown moving objects and not only motionless background elements.
For the environment map construction and reconfiguration, the MR makes an inference of the three dimensional structure of a scene from its two dimensional 2D projections. The 3D description of the scene is obtained from different viewpoints. With this 3D description we are able to recreate the environment map for use in robot navigation.
In general, in any stereoscopic vision system after the initial camera calibration, correspondence is found among a set of points in the multiple images by using a feature based approach. Disparity computation for the matched points is then performed. Establishing correspondences between point locations in images acquired from multiple views (matching) is one of the key tasks in the scene reconstruction based on stereoscopic image analysis. This feature based approach involves detecting the feature points and tracking their positions in multiple views of the scene. Aggarwal presented a review of the problem in which they discussed the developments in establishing stereoscopic correspondence for the extraction of the 3D structure (Aggarwal et al., 2000). A few well-known algorithms representing widely different approaches were presented, the focus of the review was stereoscopic matching.
For map construction or reconfiguration of the MR obstacles environment there is not necessary to reconstruct an exact scene of the environment. There are other works in the same line, in (Calisi et al., 2007) is presented an approach that integrates appearance models and stereoscopic vision for decision people tracking in domestic environments. In (Abellatif, 2008) the author used a vision system for obstacle detection and avoidance, it was proposed a method to integrate the behavior decisions by using potential field theory (Khatib, 1985) with fuzzy logic variables. It was used Hue, Saturation, and Intensity (HSI) color since it is perceptually uniform. In (Cao, 2001) was presented an omnidirectional vision camera system that produces spherical field of view of an environment, the continuation of this work was presented in (Cao et al., 2008) where the authors explained several important issues to consider for using fisheye lens in omnidirectional vision, some of them are the lens camera calibration, rectification of the lens distortion, the use of a particle filter for tracking, as well as the algorithms and the hardware configuration that they implemented.
Recently, the company "Mobile Robots" announced a heavy duty high speed stereoscopic vision system for robots called "MobileRanger StereoVision System", that is able to provide processed images at a maximal rate of 60 fps (frames per second) with a resolution of 752×480 pixels.

The proposed method has some advantages over existing methods, for example it does not need camera calibration for depth (distance) estimation measurements; an improved efficiency in the stereoscopic correspondence for block matching; adaptive candidate matching window concept is introduced for stereoscopic correspondence for block matching resulting in improved efficiency by reducing calculation time, also improves matching accuracy assuring corresponding process only in the areas where there are vertical or corners arranged pixels corresponding to the obstacles selected features. The calculation process is reduced in average 40% corresponding to the surface ground image content which is previously extracted from every image. The areas between edges in the obstacles itself are also excluded from matching process. This is an additional increase in the method efficiency by reducing calculation for matching process. This feature provides the optimal choice of the best component of the video signal giving improvements in precision of architecture based on FPGA implementation of a vision module for obstacle detection, for map building and dynamic map reconfiguration as an extension research of the ant colony environment model described in a previous work (Porta et al., 2009).

This work is organized as follows: In section 2 the general system architecture is explained. Section 3 is dedicated to give a description of the process to extract the surface ground and obstacle edge detection using luminance components, as well as the process when we include the Hue to obtain the ground surface, moreover, in this section we comment some advantages obtained with the implementation of the vision module into an FPGA. In Section 4 some important concepts about stereoscopic vision are given. In Section 5 is explained how the modification of the road map is achieved. Finally, in Section 6 are the conclusions.

2. General System Overview

Figure 1 shows the two main components of the system architecture, the computer, and the MR:

1. The computer contains the global planner based on the SACOdm algorithm, and the communication system.
2. The MR is a three wheels system with frontal differential tracking, it has six main subsystems:
   (a) The stereoscopic vision includes parallel arrange dedicated purpose video decoders controlled via IIC by the FPGA.
   (b) The Spartan-3 FPGA controller board that contains embedded the Microblaze microcontroller, as well as the motors and tracking controllers that were coded in VHDL hardware description language software.
   (c) The power module consists of a high capacity group of rechargeable batteries (not shown in the figure), two H-bridges motor drivers, and the two Pittman DC geared-motor model GM9236S025-R1.
   (d) The communication system based on the XbeePro RF, integrated WiFi communication module.
   (e) A high accuracy GPS module with 1 cm of resolution, 0.05% of accuracy, such as the VBOX 3i from Racelogic (VBOX, 2009), or similar.
   (f) An electromagnetic custom made compass IIC bus compatible, based on the LIS3LV02DL integrated circuit from STMicroelectronics.

The communication between the MR and the computer is achieved using the XBeePro RF Modules that meets the IEEE 802.15.4 standards, the modules operates within the ISM (Industrial Scientific and Medical) 2.4 GHz frequency band. The range of application for indoor/urban range is 100 meters (m), and for outdoor applications with RF line of sight the range is about 1500 m. The serial data rate is in between 1200 bits per second (bps) to 250 kilo bits per second (kbps) (XBee XBee OEM RF Modules, 2007). With no hardware modification it is possible to change the RF module to the XBee-Pro XSC to improve the communication range from 370 m for indoor/urban applications, and 9.6 Km for outdoor line sight applications.

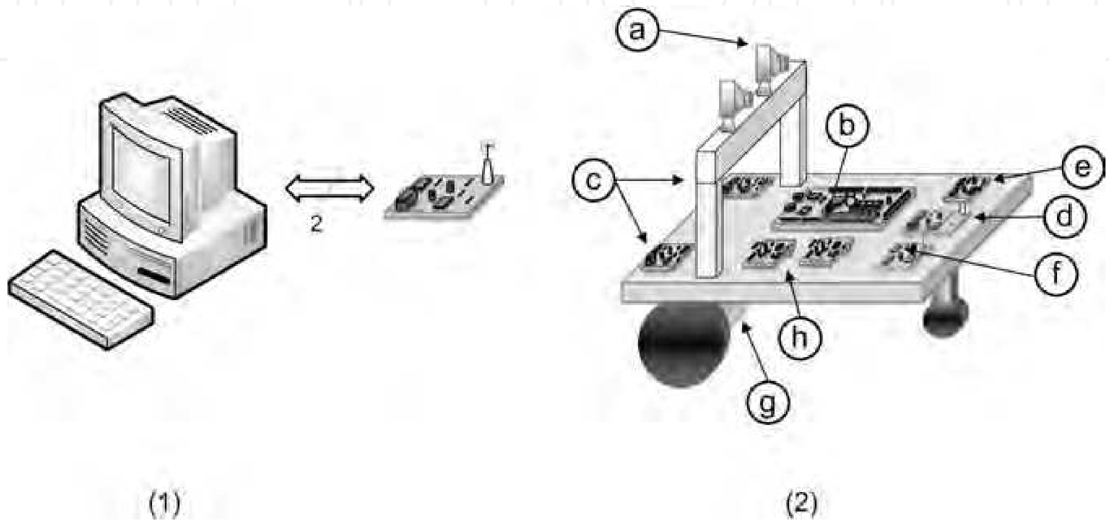

Fig. 1. The global planner is in the computer communicated through RF with the MR, this is shown in 1). In 2) is the MR with its main components: a) the cameras, b) FPGA system board, c) H bridge motor drivers, d) RF communication system based on the Zigbee technology, e) Magnetic Compass, f) GPS module, g) Gear Pittman DC-motors, h) NTSC Composite video to RGB converter cards.

In Fig. 2 a more detailed description of the stereoscopic vision system is given, each video camera is connected to a conversion board from NTSC composite video to RGB 24 bits video signals, which in turn are controlled by the FPGA based controller board using IIC communication. The video cards send the video information to the controller board where it is processed.

Fig. 3 shows the Microblaze processor, it is a 32 bit soft core processor with Harvard architecture embedded into a Xilinx FPGA. The Microblaze allows to customize its architecture for a specific application. It can manage 4 GB of memory. The 32 bits Local Memory Bus (LMB) connects the processor's core to the RAM Memory Blocks (BRAM) for data (DLMB) and instruction (ILMB) handling. The Microblaze uses the Processor Local Bus (PLB) also called On-Chip Peripheral Bus (OPB) to connect different slave peripherals (SPLB) with the CPU, for data and instruction exchange it uses the DPLB and IPLB, respectively. In the figure are connected also to the Microblaze core: The peripherals PWM, RS232, IIC, Timer, etc. These last modules were designed for specific application and glued to the Microblaze architecture. An important feature of this processor is that also contains the Microprocessor Debug Module (MDM) that gives the possibility to achieve real time debugging using the JTAG interface. The stereoscopic vision module was programmed using ANSI C/C++ language.

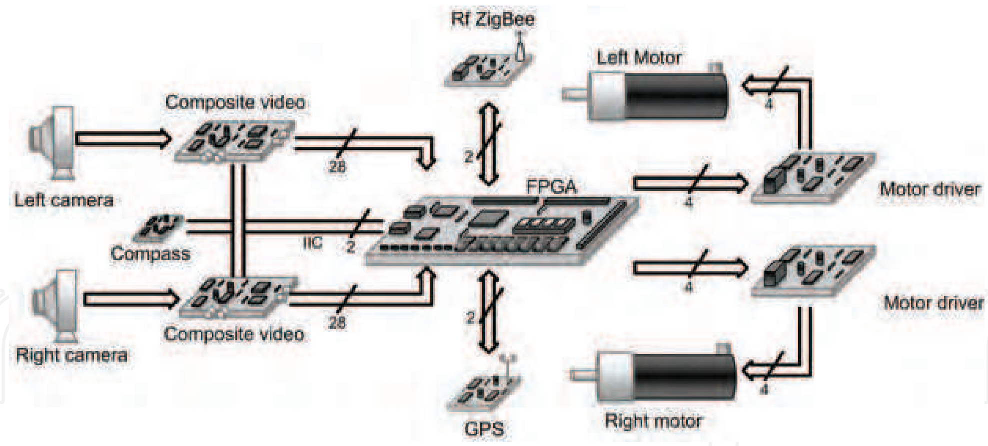

Fig. 2. Detailed overview of subsystems of the Stereoscopic vision stage on board of the MR.

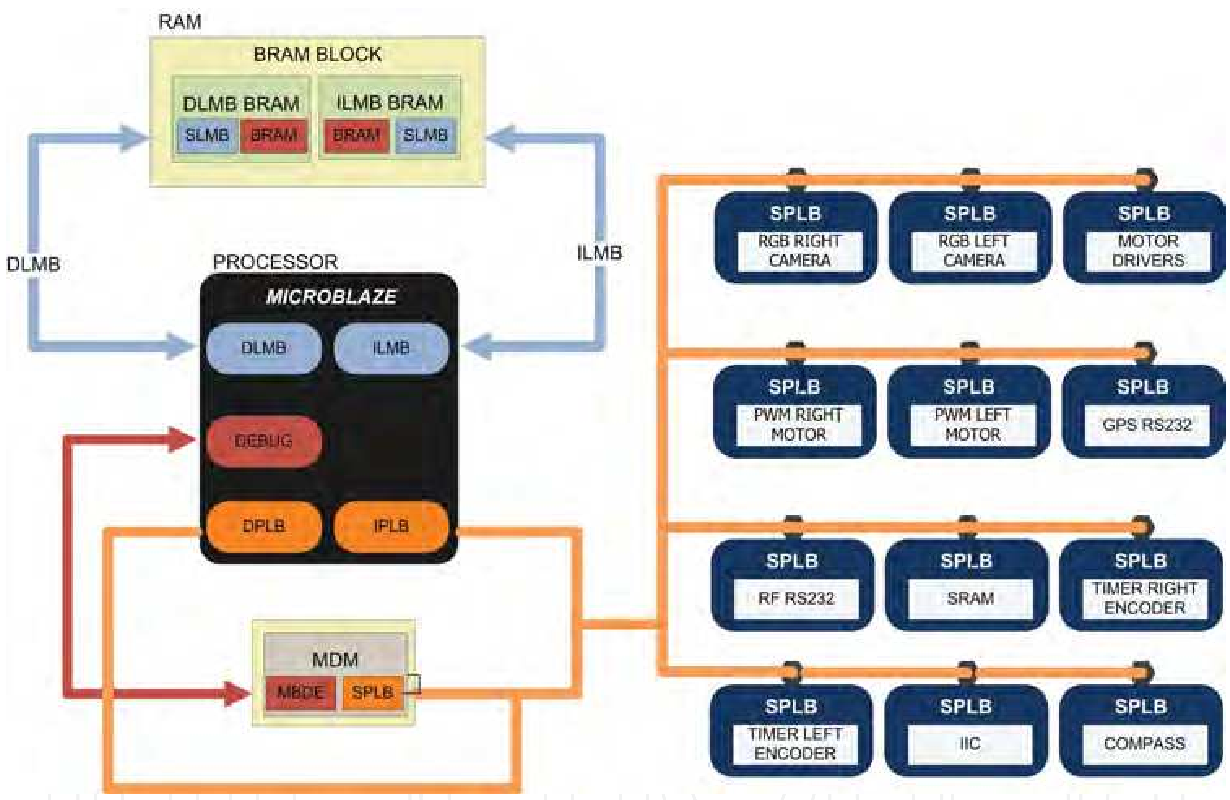

Fig. 3. Microblaze processor embedded into Xilinx FPGA and system peripherals.

3. Description of the Detection Module with Stereoscopic Vision

The navigation task is achieve using the relative depth representation of the obstacles based on stereoscopic vision and the epipolar geometry. The map represents the status at the time of drawing the map, not necessarily consistent with the actual status of the environment at the time of using the map. Mapping is the problem of integrating the information gathered in this case by the MR sensors into a complex model and depicting with a given representation. Stereo images obtained from the environment are supplied to the MR, by applying disparity algorithm on stereo image pairs, depth map for the current view is obtained. A cognitive map of the environment is updated gradually with the depth information extracted while the MR

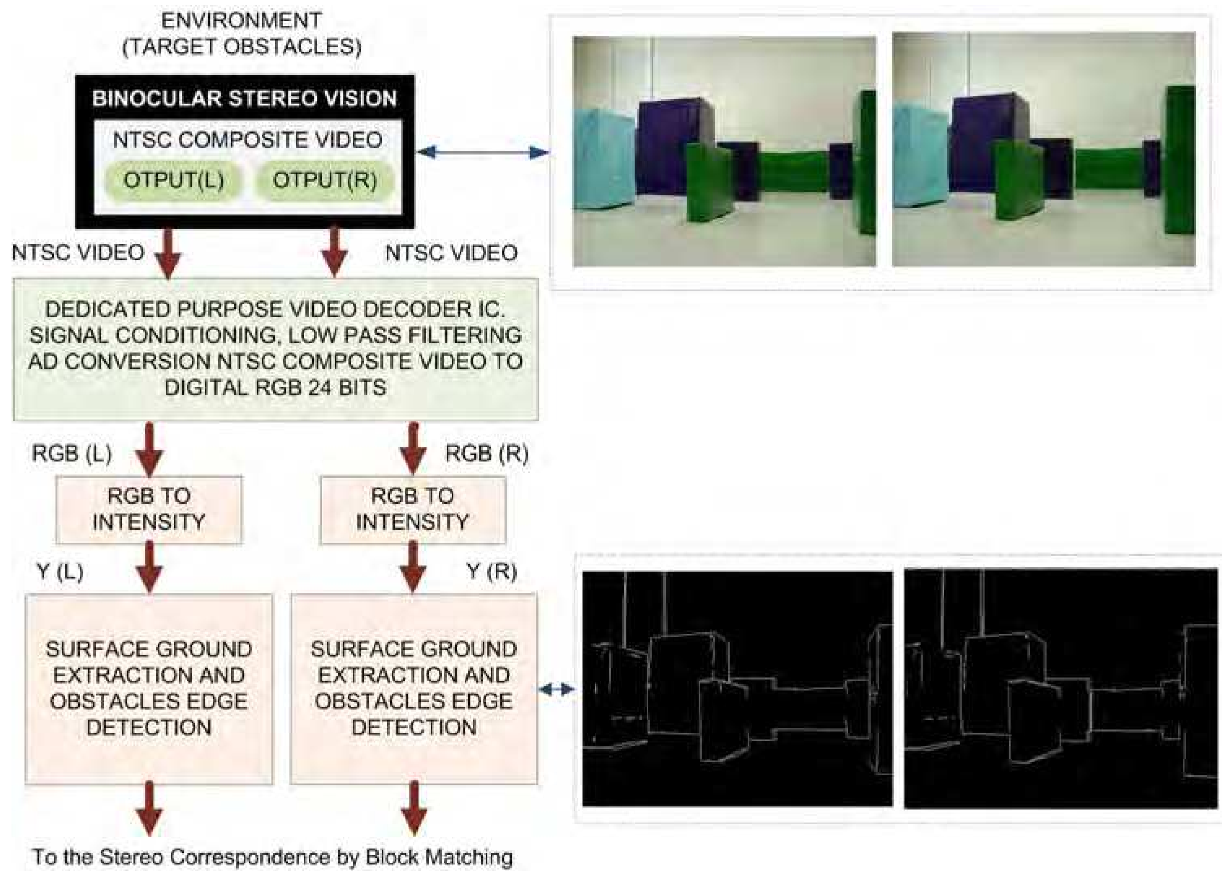

Fig. 4. Process in the detection module for surface ground extraction, and obstacles edge detection using luminance component.

is exploring the environment. The MR explores its environment using the current views, if an obstacle in its path is observed, the information of the target obstacles in the path will be send to the global planner in the main computer. After each movement of the MR in the environment, stereo images are obtained and processed in order to extract depth information. For this purpose, obstacle's feature points, which are obstacle edges, are extracted from the images. Corresponding pairs are found by matching the edge points, i.e., pixel's features which have similar vertical orientation. After performing the stereo epipolar geometry calculation, depth for the current view is extracted. By knowing the camera parameters, location, and orientation, the map can be updated with the current depth information.

3.1 Surface Ground and Obstacles Detection Using Luminance and Hue

The vision based obstacle detection module classifies each individual image pixel as belonging either to an obstacle or the ground. Appearance base method is used for surface ground classification and extraction from the MR vision module captured images, see Fig. 4. Any pixel that differs in appearance from the ground is classified as an obstacle. After surface ground extraction, remaining image content are only obstacles. A combination of pixel appearance and feature base method is used for individual obstacle detection and edge extraction. Obstacles edges are more suitable for stereo correspondence block matching in order to determine the disparity between left and right images. For ground surface extraction purpose, two assumptions were established that are reasonable for a variety of indoor and

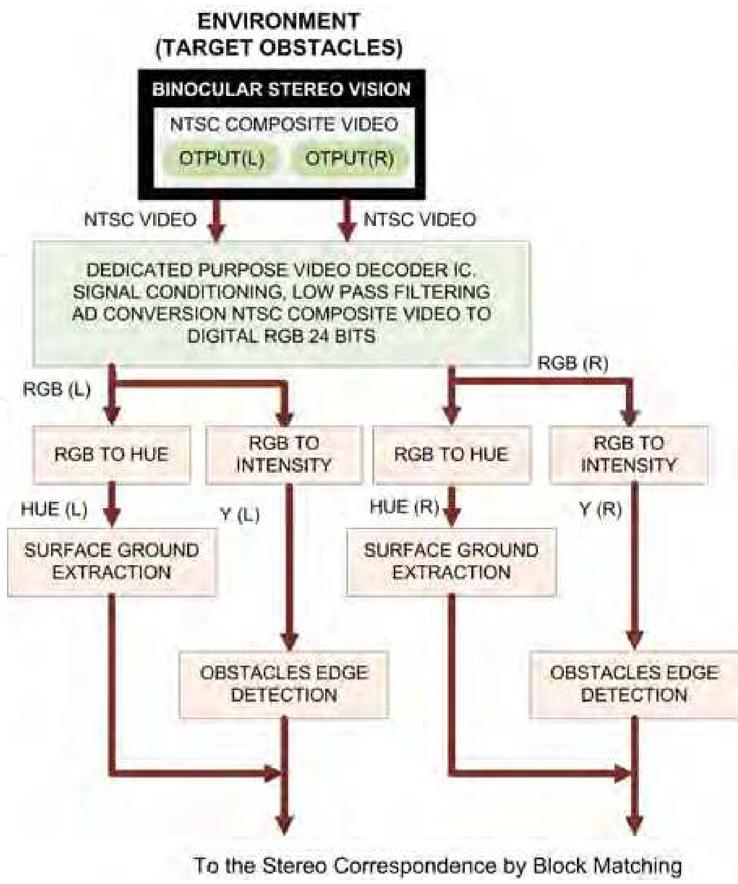

Fig. 5. Process in the detection module for surface ground extraction using Hue, and obstacles edge detection using luminance components.

outdoor environments:

1. The ground is relatively flat.
2. Obstacles differ in color appearance from the ground. This difference is reasonable and can be subjectively measured as Just Noticeably Difference (JND), which is reasonable for a real environment.

Above assumptions allow us to distinguish obstacles from the ground and to estimate the distances between detected obstacles from the vision based system. The classification of a pixel as representing an obstacle or the surface ground can be based on local visual attributes: Intensity, Hue, edges, and corners. Selected attributes must provide information so that the system performs reliably in a variety of environments. Selected attributes should also require low computation time so that real time system performance can be achieved. The less computational cost has the attribute, the obstacle detection update rate is greater, and consequently the MR travel faster and safer.

For appearance classification we used Hue as a primary attribute for ground surface detection and extraction, see Fig. 5. Hue provides more stable information than color or luminance based on pixel gray level. Color saturation and luminance perceived from an object is affected by changes in incident and reflected lightness. Also compared to texture, Hue is more local attribute and faster to calculate. In general, Hue is one of the main properties of a color, defined as the degree of perceived stimulus described as Red, Green, and Blue. When a pixel is classified as an obstacle, its distance from the MR stereo vision cameras system is estimated. The considerations for the surface ground extraction and obstacle edge detection for correspondences block matching are:

1. Color image from each video camera is converted from NTSC composite video to RGB 24 bits color space.
2. A typical ground area in front of the MR is used as a reference. The Hue attributes from the pixels inside this area are histogrammed in order to determine its Hue attribute statistics.
3. Surface ground is extracted from the scene captured by the MR stereo vision by means of a comparison against the reference of point 2 above, and based on Hue attribute. Hue limits are based in JND units.
4. Remaining content in images are only obstacles. Edges are extracted from individual obstacles based on feature and appearance pixel's attributes.
5. Correspondence for block matching is established in pixels from the obstacle vertical edges.
6. Disparity map is obtained from the sum of absolute differences (SAD) correlation method.

3.2 Vision System Module FPGA Implementation

When a robot has to react immediately to real-world events detected by a vision system, high speed processing is required. Vision is part of the MR control loop during navigation. Sensors and processing system should ideally respond within one robot control cycle in order to not limit their MR dynamic. An MR vision system equipped, requires high computational power and data throughput which computation time often exceed their abilities to properly react. In the ant colony environment model, every ant is a virtual MR full equipped, trying to find the optimal route, eventually, weather there exist, it will be obtained. Of course, the ACO based planner will give the best route found, and the real ant, the MR, which is equipped on board with the vision system, will update the global map in the planner. There are many tasks to do at the same time, however, a good feature of using FPGAs is that they allow concurrently implementation of the different tasks, this is a desirable quality for processing high speed vision. High parallelism is comprised with high use of the FPGA resources; so a balance between parallelization of task, and serial execution of some of them will depend on the specific necessities.

The vision system consists of stereoscopic vision module implemented in VHDL and C codes operating in a Xilinx based FPGA, hence a balanced used of resources were used. Video information is processed in a stereo vision system and video interface. The NTSC composite video signals from each camera after properly low pass filtering and level conditioning, are converted to RGB 24 bits color space by a state of the art video interface system HDTV capable. The rest of the video stage was programmed in C for the Microblaze system embedded into the FPGA. Other tasks, such as the motion control block are parallel implementation to the video system.

4. Design of the Stereoscopic Vision Module

The two stereo cameras parallel aligned, capture images of the same obstacle from different positions. The 2D images on the plane of projection represent the object from camera view. These two images contain the encrypted depth distance information. This depth distance information can be used for a 3D representation in the ant colony environment in order to build a map.

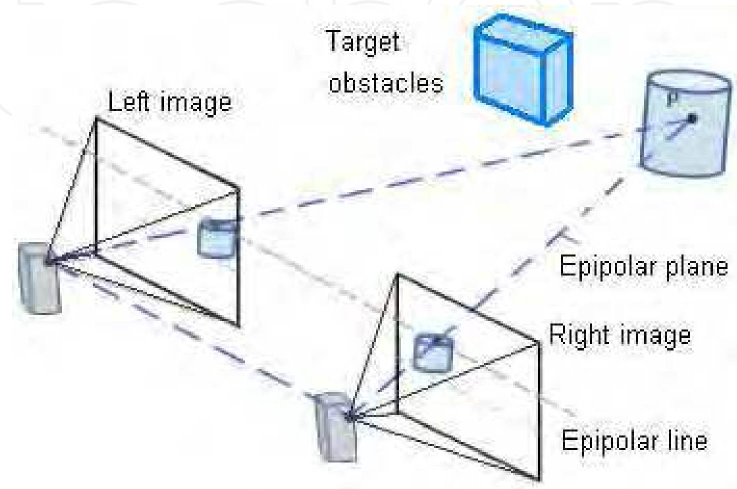

Fig. 6. Projection of one point into left and right images from parallel arrange stereo cameras.

4.1 Stereoscopic Vision

The MR using its side by side left and right cameras see the scene environment from different positions in a similar way as human eyes, see Fig. 6. The FPGA based processing system finds corresponding points in the two images and compares them in a correspondence matching process. Images are compared by shifting a small pixels block "window". The result is a comparison of the two images together over top of each other to find the pixels of the obstacle that best match. The shifted amount between the same pixel in the two images is called disparity, which is related to the obstacle depth distance. The higher disparity means that the obstacle containing that pixel is closer to the cameras. The less disparity means the object is far from the cameras, if the object is very far away, the disparity is zero, that means the object on the left image is the same pixel location on the right image.
Figure 7 shows the geometrical basis for stereoscopic vision by using two identical cameras, which are fixed on the same plane and turned in the same direction, parallax sight. The position of the cameras is different in the X axis. The image planes are presented in front of the cameras to model the projection easier. Consider the point P on the object, whose perspective projections on the image planes are located at P_L and P_R from left and right cameras respectively. These perspective projections are constructed by drawing straight lines from the point to the center lens of the left and right cameras. The intersection of the line and image plane is the projection point. The left camera's projection point P_L is shift from the center, while the right camera's projection point P_R is at the center. This shift of the corresponding point on left and right camera can be computed to get the depth information of the obstacle.

4.2 Depth Measure from Stereo Image

In order to calculate the depth measure of the obstacles in the scene, the first step is to determine the points of interest for correspondence matching between the two images. This corresponding points are selected based on the obstacle edge feature. Then calculate the depth distance based on the shifting "disparity". The disparity is calculated based on the amount of pixel's shifting in a particular corresponding point. There are stereo image constraints to be assume for solving the correspondence problem:

1. Uniqueness. Each point has at most one match in the other image.
2. Similarity. Each intensity color area matches a similar intensity color area in the other image.
3. Ordering. The order of points in two images is usually the same.
4. Continuity. Disparity changes vary slowly across a surface, except at depth edges.
5. Epipolar constraint. Given a point in the image, the matching point in the other image must lie along a single line.

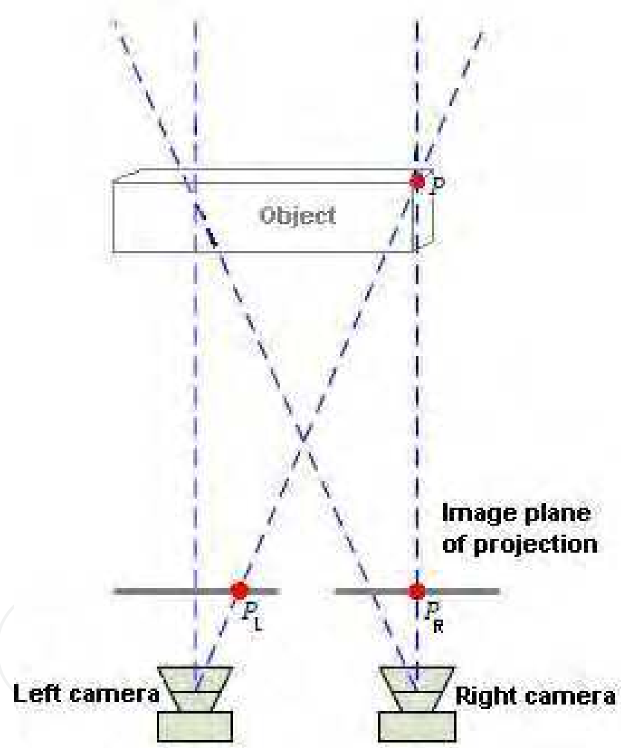

Fig. 7. Points P_L and P_R are the perspective projections of P in left and right views.

5. Modifying Road Maps

The modification of the Road Maps is achieved using the information of disparity in pixels, where the distance of the MR from the obstacle is estimated using disparity measures, the less disparity measure means that the obstacle is far from the visual system of the MR as can be seen in Fig. 8. Moreover, the MR uses a high accuracy GPS and a digital compass. For every capture scene, the MR sends the location, orientation (x, y, θ) and the corresponding disparity map with all the necessary (x, y, d) coordinates and corresponding disparities, which in reality are a 3D representation of the 2D obstacles images captured from the stereoscopic visual system. After pixel's scaling and coordinates translation, the global planner is able to update the environment, its representation includes the visual shape and geographical coordinates. Once the global planner in the main computer has been modified using the new information about new obstacles and current position of the MR, the global planner performs calculations using ACO to obtain an updated optimized path, which is sent to the MR to achieve the navigation. The MR has the ability to send new information every 100ms via RF from every scene captured; however, times in the global planner are bigger since it is based on a natural optimization method, and it depends on the actual position of MR with respect to the goal. Hence, most of times a new path can be obtained every 3 seconds.

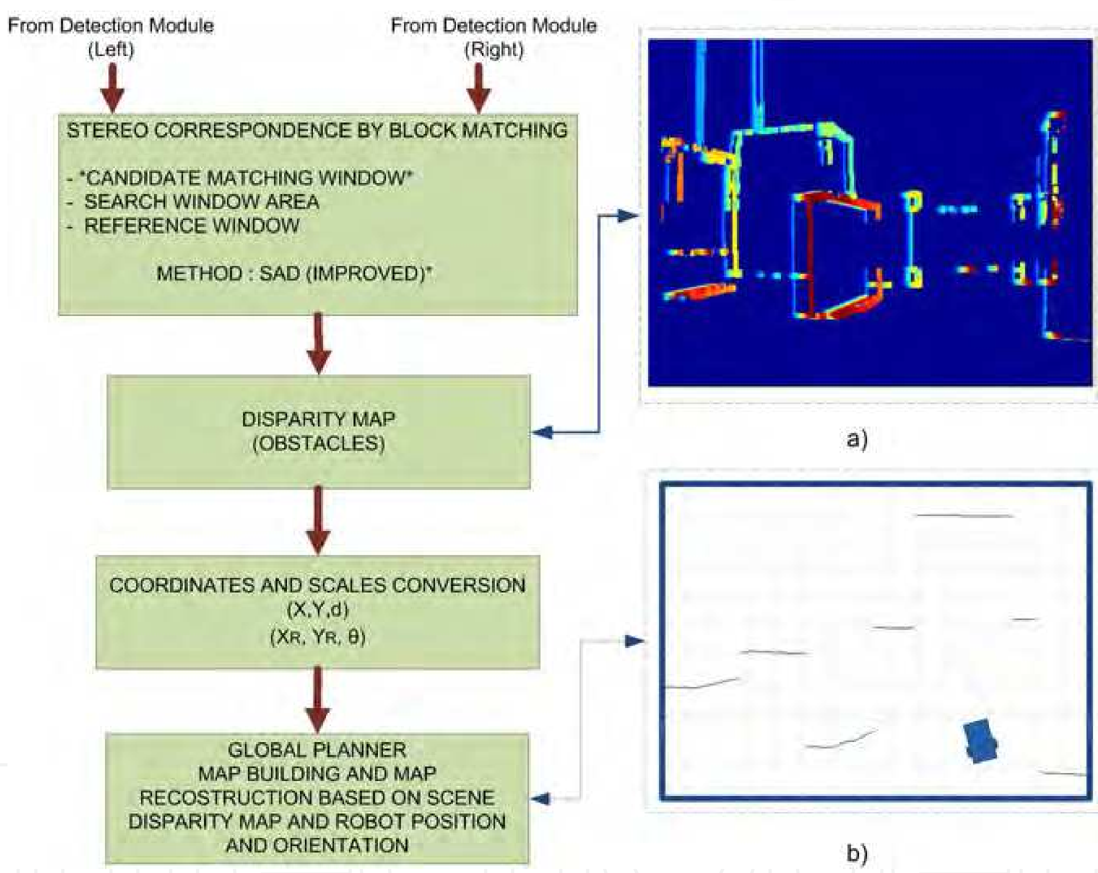

Fig. 8. Process for map building and map reconfiguration.

6. Conclusion

In this work was shown the design of an stereoscopic vision module for a wheeled mobile robot, suitable to be implemented into an FPGA. The main purpose of the onboard system of the MR is to provide the necessary elements for perception, obstacles detection, map building and map reconfiguration in a tough environment where there are no landmarks or references. The stereoscopic vision system captures left and right images from the same MR scene, the system is capable of using both appearance based pixel descriptors for surface ground extraction, luminance or Hue depending of the environment particular characteristics. In an environment with constant lightness, minimum reflections and proper setting in the edge detector threshold level, luminance can be suitable because surface ground and obstacles edge detection can be performed at the same time. For environment with variable light conditions or uncertain, Hue is the primary attribute for pixel appearance descriptor in the surface ground extraction process due to its invariance to changes in luminance and color saturation. After surface ground extraction and obstacles edge detection, stereoscopic corresponding by block matching is performed, the correspondence is found among a set of points in the left and right images by using a feature based approach. Disparity computation for the matched points is then performed. Establishing correspondences between point locations in images acquired from multiple views (matching) is one of the key tasks in the reconstruction based on stereo image analysis. This feature based approach, involves detecting the feature points and tracking their positions in multiple views of the environment. Stereoscopic camera calibration is not required due to the improvements in matching process. Disparity maps which are the depth measure of the obstacles position in the environment are obtained after the stereo correspondence process. The MR sends this data, including its position and orientation via RF to the global planner located in the main computer outside the environment. With this information the global planner is able to constantly update the environment map.

7. References

Abellatif M. (2008). Behavior Fusion for Visually-Guided Service Robots, Xiong Zhihui *In*: In-Teh, *Computer Vision*, Croatia, pp. 1-12.

Aggarwal J. K., Zhao H., Mandal C., Bemuri B. C. (2000). 3D Shape Reconstruction from Multiple Views, *in* Alan C. Bovik, Editor, *Handbook of Image and Video Processing*, Academic Press, pp. 243-257.

Calisi D., Iocci L., Leone G. R. (2007). Person Following through Appearence Models and Stereo Vision using a Mobile Robot, *Proc. of International Workshop on Robot Vision*, pp. 46-56.

Cao Z. L. (2001). Ommi-vision based Autonomous Mobile Robotic Platform, *Proceedings of SPIE Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision*, Vol. 4572, Newton USA, pp. 51-60.

Cao Z., Meng X., & Liu S. (2008). Dynamic Omnidirectional Vision Localization Using a Beacon Tracker Based on Particle Filter, *In:* Xiong Zhihui, *Computer Vision*, Ed. In-Teh, Croatia, pp. 13-28.

Khatib O. (1985). Real-Time Obstacle Avoidance for Manipulators and Mobile Robots, *Proceedings of IEEE International conference on Robotics and Automation*, pp. 500-505.

Porta García M. A., Montiel O., Castillo O., Sepúlveda R., Melin P. (2009). Path planning for autonomous mobile robot navigation with ant colony optimization and fuzzy cost function evaluation, *Applied Soft Computing*, Vol. 9 (No. 3): 1102-1110.

Siegwart R., & Nourbakhsh I. R. (2004). *Introduction to Autonomous Mobile Robots, A Bradford Book*, The MIT Press, Cambridge Massachusetts, London, England.

Tsai R. Y. (1986). An efficient and accurate camera calibration technique for 3D machine vision, *IEEE Conference on Computer Vision and Pattern recognition.* pp. 364-374.

VBOX product (2009). Web page available at: http://www.racelogic.co.uk/?show=VBOX

Woods A., Docherty T., Koch R. (1993). Image distortions in stereoscopic video systems, *Proceedings of the SPIE*, San Jose, Ca. USA., Vol. 1925.

Xbee XBee-Pro OEM RF Modules (2007) , Product Manual v1.xAx - 802.15.4 Protocol, MaxStream, Inc.

Mobile Robots Navigation
Edited by Alejandra Barrera

ISBN 978-953-307-076-6
Hard cover, 666 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

Mobile robots navigation includes different interrelated activities: (i) perception, as obtaining and interpreting sensory information; (ii) exploration, as the strategy that guides the robot to select the next direction to go; (iii) mapping, involving the construction of a spatial representation by using the sensory information perceived; (iv) localization, as the strategy to estimate the robot position within the spatial map; (v) path planning, as the strategy to find a path towards a goal location being optimal or not; and (vi) path execution, where motor actions are determined and adapted to environmental changes. The book addresses those activities by integrating results from the research work of several authors all over the world. Research cases are documented in 32 chapters organized within 7 categories next described.