Extraction of Roads From Out Door Images

Alejandro Forero Guzmán M.Sc. and Carlos Parra Ph.D.
Departamento de Ingeniería Electrónica, Pontificia Universidad Javeriana, Bogotá Colombia

1. Introduction

The humanitarian demining process is very slow, expensive and most important, because it is done manually, it puts human lives at risk. Deminers are exposed to permanent danger and accidents. Even with the help of dogs, the demining process has not improved much during recent years (UNICEF, 2000).

Figure 1. Left side: Ursula Robot. Right side Amaranta Robot Project

A few separate initiatives from the robotics community to design and prove a mechanical automated solution have taken place. Here at the Pontificia Universidad Javeriana, in Colombia, we are working in this problem: on a previous project the mobile robot Ursula was developed (Rizo et al., 2003); and now we are working in a new mobile robot called Amaranta (Figure1). One part of the humanitarian demining problem is the navigation, in the two projects the autonomous navigation task is executed based on a vision system that uses a camera mounted on the robot.
Landmines could be placed in any type of terrain: deserts, mountains, swamps, roads, forests, etc. This means that when trying to build a robot for demining operations, its workspace has to be previously defined and limited. In this work humanitarian demining, is limited to operate on places that have been modified by man and that represent great importance for a community, for example the roads or paths. Specifically, the systems have been designed for the Colombian territory (Rizo et al., 2003).

Source: Vision Systems: Applications, ISBN 978-3-902613-01-1
Edited by: Goro Obinata and Ashish Dutta, pp. 608, I-Tech, Vienna, Austria, June 2007

Figure 2. Typical images

This chapter presents two approximations made for the vision system in order to enable autonomous navigation in outdoor environments; both based on the road following principle.
Therefore, almost every vehicle facing the problem of autonomous navigation using vision systems solves this problem by following the track. How ever, this technique is widely implemented only over structured roads, because painted lines over the road are a reliable characteristic to exploit. When there are no painted roads to follow or simply no road at all, autonomous navigation based on visual systems, is usually reduced to avoid obstacles.
In the last two decades, autonomous navigations have been a goal sought by different authors (Turk et al., 1988) (Thorpe et al., 1988), yet today still an open area for research (Thrun et al.,2006).
Due to the danger involved in demining efforts, a cheap robot is required, equipped with an autonomous navigation system, in order to minimize the risk for the humanitarian demining team; and an architecture capable of supporting multiple sensors to acquire the most reliable information about the surrounding area. These limitations, along with the special terrain conditions of the Colombia topography guide the present research.
Bellow some concepts from the classic theory are presented and then the complete approach made by the authors is exposed, from the problem of navigation to the extraction of roads or paths in outdoor images as essential part of the autonomous navigation.

2. Important Concepts

2.1 Colour Spaces

In general, colour is the perceptual result of light in the visible region of the spectrum (Jain, 1989). There are many colour spaces reported in the literature, each one has its characteristics. For image processing, its is usually described by the distribution of the colour of the three components R (Red), G (Green) and B (Blue), moreover many other attributes can also be calculated from these components. The colour analysis is more difficult using the three components. In image processing of exterior scenes, the illumination is a very critical parameter. A first approach is to select a colour space expressed as two colour components and one intensity/luminance component (Ohta, HIS, LUV, L*a*b*) (Aviña-Cervantes, 2005)
For example, the CIE L*a*b* colour space was presented by the International Commission of Illumination. The model was based on two properties of an older colour space called CIE XYZ. The first of these properties is that the standard was created from the frequency response of several patients' eyes, making the system independent from electronic devices. The second property, taken from CIEXYZ, is that the mathematical representation of the space allows separating the luminance from the chrominance.
On the other hand, the originality of this colour space is that it introduces the concept of perceptual uniformity. It means that if two colours are similar to each other, in CIE L*a*b* they are close and this distance is measured by the Euclidean metric. In the CIE L*a*b* space L* represents the luminance, a* codifies the reddish and greenish sensation, while b* codifies the yellowish and bluish sensation.
The space transformation from RGB to L*a*b* has two steps. The first one is a linear transformation from RG to CIE XYZ (Eq. 1).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

The second transformation is a non-linear transformation from CIE XYZ to CIE L*a*b* (Eq. 2).

$$\begin{aligned} &L^* = \begin{cases} 116 \times (Y/Y_n)^{1/3} - 16 & \text{for } Y/Y_n > 0.008856 \\ 903.3 \times Y/Y_n & \text{for } Y/Y_n \le 0.008856 \end{cases} \\ &a^* = 500 \times [f(X/X_n) - f(Y/Y_n)] \\ &b^* = 200 \times [f(Y/Y_n) - f(Z/Z_n)] \\ &\text{with} \\ &f(t) = t^{1/3} \qquad \text{for } t > 0.008856 \\ &f(t) = 7.787 \times t + \frac{16}{116} \qquad \text{for } t \le 0.008856 \end{aligned} \tag{2}$$

The X_n, Y_n, Z_n values are the *tri stimuli* related with the white point. In some cases it can be measured, but there are also standards according to the light conditions. In this case, due to the weather conditions at the capturing moment, the CIE D65 (lightening day) (Eq.3), standard is normally selected (Broek & Rikxoort, 2004).

$$X_n = 0.9502 \quad Y_n = 1 \quad Z_n = 1.0884 \tag{3}$$

This method is not simple for computer implementation. The non-linear transformation takes important time of processing.
A second approach takes the relationship between the components of RGB. For example: R/G and B/G. This approach is also applied over the colour space YCbCr. The colour space YCbCr is used in video systems. Y is the brightness component and Cb and Cr are the blue and red chrominance components.

2.2 Semantic Model

This is an abstract representation; it gives a label, corresponding to a class, to each entity found in the scene (i.e. sky, road, tree, etc.) (Murrieta-Cid et al., 2002). In the semantic model, the classification is based on a priori knowledge given to the system (Fan et al., 2001).
This knowledge consists in:
A list of possible classes that the robot identifies in the environment.
Learning attributes for each class. The region characterization is developed by using several attributes computed from the colour information. Other attributes are texture and geometrical information.
The kind of environment to be analyzed; the nature of the region is obtained by comparing a vector of features with a database composed of different classed, issues of the learning process. The database is a function of the environment and problem restrictions.

3. General System

The approach used to solve the problem of navigation presented here is very similar to the one used with structured roads: find a path and then follow it (Bertozzi et al., 2002) (Aviña-Cervantes et al., 2003). Once the road or path has been identified, following it supposes a known problem in robotics. For this reason, the focus of this exposition is centered on the identification of the area that represents the path the robot will follow, and the extraction of some parameters necessary for the control stage.

Figure 2 shows the kind of images processed by the system. In all of these images there are certain characteristics in common: a set of pixels, mostly connected, represent the road; in a close range image (5 to 25 meters) it is highly probable to have only one road; a road that can be followed goes from the bottom of the picture till some point in the middle upper area of the picture.

Along with these ideas, other facts are implicit: the picture is taken horizontal to the ground, the sky is in the upper portion of the picture, there is sufficient light to distinguish the road or path, there are not objects obstructing the view of the road. All of these assumptions are easily fulfilled in real conditions. Every one of these characteristics is used as semantic information and helps to delimit the scope of the problem. The processing of the images is done to exploit all the semantic characteristics mentioned before. At the end, semantic rules are applied to extract the essential information in the image: the route over the navigable terrain.

Figure 3. Semantic characteristic in the image

3.1 Extraction of characteristics

Different approaches have been made to extract the path or road in outdoor images, colour segmentation is one of those, but this technique is very susceptible to changes in illumination, a big draw back in outdoor vision systems.

To overcome this difficulty many approximations have been presented by numerous authors (Turk et al., 1988) (Thorpe et al., 1988), but some times they are to complex to be implemented in a cheap system. Instead, we propose the use of semantic information, similar to the one we use as humans to follow a path, to reduce the complexity of the colour segmentation by using the possible meanings of the different characteristics found in the image and its relation with the context, hence reducing the universe of possibilities.

3.2 Semantic information

As the camera is upward and horizontal to the ground the representation of a typical image consists of the sky, above the horizon; and the segment below the horizon contains maximum two areas: path and not path regions. Figure 3 shows the typical target image and the principal semantic characteristic in it: sky, in the blue rectangle; horizon, the brown line; and road inside the red segment.

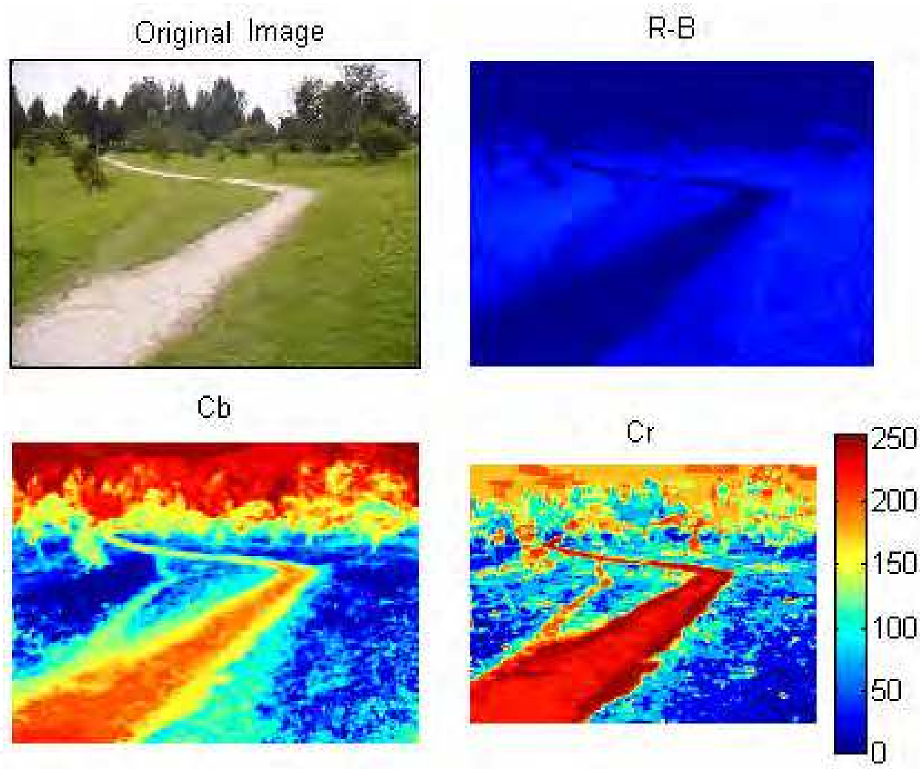

Figure 4. Examples of path visualization in differents color spaces

This simple representation reduces the complexity in the colour segmentation and enables the use of a wider space to separate the two regions below the horizon, what in time, reduces the negative effect of variable shadows and changes of illumination over the objective path.

In addition, the binary classes enable the use of techniques as simple as binarization to make the segmentation: only two classes below the horizon. Herewith the computation problem is reduced; as a result the computational resources can be reduced without increasing the time that the process consumes (Duda et al., 2001).

4. Algorithms

Two approximations were made to solve the problem; both of them use some colour segmentation along with semantic information. The first algorithm, which works in the RGB space, is the initial design (Forero & Parra, 2004). It was tested in Matlab® and designed to work along with the first implementation of the robot Ursula.

The second algorithm was conceived to work in an embedded system and it captures the image in the YCbCr colour space. In figure 4, for pictures show the image in different colour spaces. Next, both algorithms are explained in more detail.

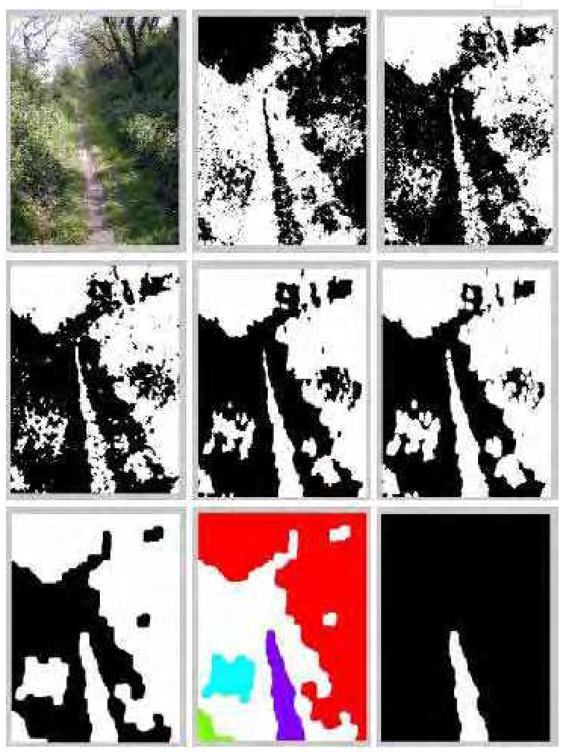

Figure 5. From top to bottom, left to right: original image RGB, R-B plane, inverse R-B, image result from a median filter, image result from the first morphological filter, image thicken, image result from the second morphological filter, image segmented by connectivity, path selected

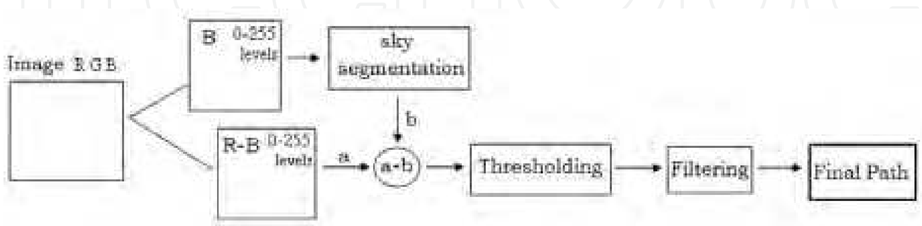

Figure 6. Flow diagram for the RGB implementation

4.1 RGB implementation

This algorithm works in 2 of the three components of the RGB space, the channels red and blue. Chanel blue is used to segment the sky and the projection R −B is used in the rest of the process. Channel green is not used because it has few or not information about the path.
After the sky is identified, it is subtracted from the rest of the image, so the projection R −B does not contain the segment levelled as sky. This special projection $\alpha R - \beta B$, with constants α and β set to 0,5, is used because in it the *dirt* of the paths is enhanced and the effect of shadows is reduced (Turk et al., 1988).
The new image, the projection R −B, is subject to a threshold operation to obtain two classes in it; this is the first approximation to the road and not road classes. The threshold is selected using the Otsu Method (Otsu, 1979) so the separability of the two classes is maximized. Then, morphological filters are applied to reduce holes and increase the connectivity in the biggest region of this image.
After filtering, one of the biggest connected regions in the image should be the road. If more than one region has the characteristics the algorithm is looking for, the semantic rules should help to choose one: centroid, major axis length, minor axis length and orientation of the areas that contains each region are extracted and then the rules are applied to select one as the final path.
Figure 5 illustrates the different stages of the algorithm, and figure 6 illustrates the algorithm with a flow diagram.

4.2 YCbCr implementation

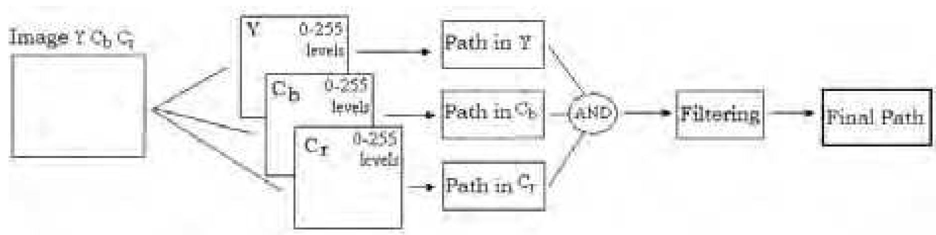

Figure 7. Flow diagram for the YCbCr implementation

This algorithm was developed after the RGB implementation (Maldonado et al., 2006), thinking of the online application; it takes the three channels of the space YCbCr, and uses each one as 256 level images. The dimension in all channels is reduced, so the portion of the sky is taken out. This reduction in the image is done by a geometric calculation, and if some pixels from the ground are also lost, they would be the far away pixels close to the horizon; this is not a problem because in future images the problem will be corrected, before the robot reaches that point.
Afterward, a threshold is applied to each channel, following the same idea used in the RGB implementation; then all the information in the three components its put together with an AND operator. This way the information is merging by adding the coincident pixels and extracting those who only represent a hit in one or two channels.
Finally, the result image is filtered to reduce the effect of noise and obtain a single path. A median filter with a 5 x 5 window is used for this purpose.

After one region is selected as the path, his centroid along with the direction of the path are extracted. This enables the control system to plan the path and overtake autonomous navigation. Figure 7 illustrates the algorithm with a flow diagram, and figure 8 illustrates the different stages of the algorithm.

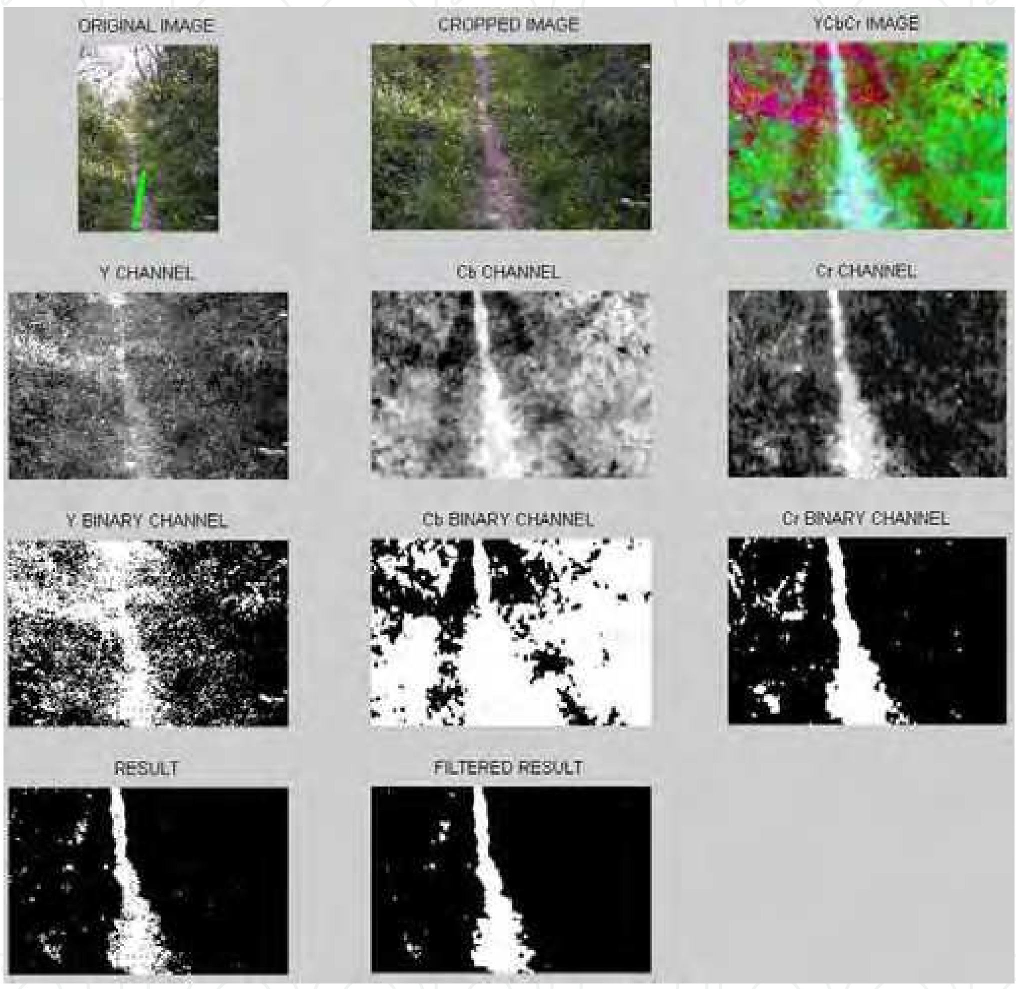

Figure 8. Real Time System Processing

5. Hardware realization

The algorithm was tested in Matlab®, and then implemented in a high-level language, C++ with the Intel's library OpenCV®. These steps were to help develop and analyse the algorithm. After concluding this stage, the complete process was implemented in embedded systems to achieve portability along with real time processing. Here after, the implementation in OpenCV® and the implementation in the Blackfin® Processors are explained with more detail.

5.1 OpenCV® implementation

The implementation in OpenCV® was done in a PC with Fedora Linux operative system, where the algorithm was programmed in C/C++. The Intel's compiler used came along with the OpenCV library. The final application takes as input an image in AVI (Audio Video Interleave) format and then executes the recognition algorithm, previously tested in Matlab®. The objective of using C++ to program the concluding version was to facilitate the migration of the application into the embedded version in the DSP EZKIT- LITE BF533 Blackfin®.

5.2 Blackfin® implementation

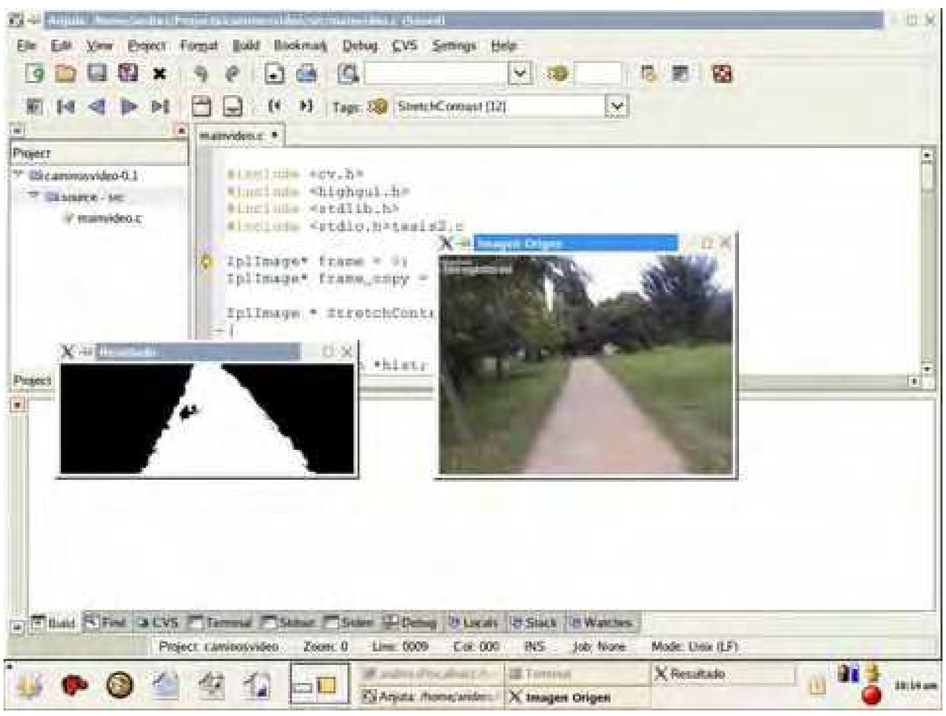

Figure 9. Real Time System Processing

To accomplish real time processing, as shown in figure 9, the DSP EZKIT- LITE BF533 Blackfin® was used. Three special functions were implemented to bring about the operation in the desire hardware:

Initialization, the Blackfin BF 533 processor and the developed card ADSP EZKIT – LITE BF533 are prepared to capture video and generate the interrupt and synchronism signals with the peripheries: a video decoder ADV7183 and a the embedded DMA. The first one transforms the analog video NTSC into digital video ITU-656, and the processor's DMA transfers the video information into the RAM memory in the developed card.

The processing stage starts with the extraction of each channel YCbCr from the image in memory. Them, it calls all procedures that execute the algorithm: Contrast estimation, histogram calculation, threshold calculation, and adding the three channels. Finally, it calculates the moments of the region of interest to extract the centroid and the orientation of the path.

In the last step, Transmission, the information concerning the path (centroid and the orientation) is transmitted by a RS-232 serial interface to a navigation module.

Besides these functions, other considerations had to be taken to run the algorithm in the embedded system:

New data types were created in C++ in order to be compatible with ADSP EZKIT- LITE BF533. These data structures manage the information in the image and handle all the parameters that the algorithm uses.

All the variables are defined according with the size and physical position that each one will take in the physical memory in the development kit. This execution allows a better use of the hardware resource and enables simultaneous processing of two images, one image is acquired by the DMA, and other is processed in the CPU.

Finally, The Blackfin's ALU only handles fixed-point values, so floating-point values have to be avoided in order to maintain the performance of the whole system.

6. Conclusion

Even when there has been an extensive development of works on road detection and road following during the last two decades, most of them are focused on well structured roads, making difficult its use for humanitarian demining activities. The present work shows a way to use the natural information in outdoor environment to extract the roads or paths characteristics, which can be used as landmarks for the navigation process.

Other important observation is that the information combines of two colors, (i.e. the projection R─ B, Cb or Cr channels) hence reducing the harmful effect of the changing illumination in natural environment.

Good results were also achieved in the path planning process. The robot executes a 2½ D trajectory planning, which facilitates the work of the vision system because only the close range segmentation has to be correct to be successful in the path planning.

With regard to the semantic information, the results show how semantic characteristics make possible the use of low-level operations to extract the information required without spending too many time and hardware resources.

Finally, the system implemented is part of a visual exploration strategy which is being implemented in the robot Amaranta, and has other visual perception functions like the detection of buried objects by color and texture analysis. When the whole system will be functional it will integrate techniques of control visual navigation and would be a great tool to test how all the system can work together (Coronado et al., 2005).

7. References

Aviña-Cervantes, G. Navigation visuelle d'un robot mobile dans un environnement d'extérieur semi-structuré. *Ph.D. Thesis*. INP Toulouse. France. 2005.

Broek, E.L. van den; Rikxoort, E.M. van. Evaluation of color representation for texture analysis. *Proceedings of the 16th Belgium-Netherlands Conference on Artificial Intelligence*. University of Groningen. 21-22 October, 2004.

UNICEF- Colombia. *Sembrando Minas, Cosechando Muerte.* UNICEFBogotá. Colombia. September 2000.

Murrieta-Cid, R; Parra, C. & Devy M. Visual Navigation on Natural Environments. *Journal on Autonomous Robots*. Vol. 13. July 2002. pp 143-168. ISSN 0929-5593

Rizo, J.; Coronado, J.; Campo, C.; Forero A.; Otálora, C.; Devy, M. & Parra, C. URSULA : Robotic Demining System. *Proceedings of International Conference on Advanced Robotics ICAR*. ISBN: 9729688990. Coimbra. Portugal. 2003.

Jain, A. *Fundamental of Digital Image Processing*. Prentice-Hall International Editions. ISBN: 0-13-332578-4. United State of America. 1989.

Forero, A. & Parra C. Automatic Extraction of Semantic Characteristics from Outdoor Images for Visual Robot Navigation. *Proceedings of International Conference IEEE/ International Symposium on Robotics and Automation*. ISBN: 9709702009. Querétaro. Mexico. 2004.

Aviña-Cervantes, G.; Devy, M. & Marín, A. Lane Extraction and Tracking for Robot Navigation in Agricultural Applications. *Proceedings of International Conference on Advanced Robotics ICAR*. ISBN: 9729688990. Coimbra. Portugal. 2003.

Maldonado, A.; Forero A. & Parra, C. Real Time Navigation on Unstructured Roads. *Proceedings of Colombian Workshop on Robotics And Automation (CWRA/IEEE)*. Bogotá. Colombia. 2006.

Turk, M. A.; Morgenthaler, D. G.; Gremgan, K. D. & Marra, M. VITS- A vision system for autonomous land vehicle navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 10. No. 3 May 1988. ISSN: 0162-8828.

Thorpe, C.; Hebert, M.; Kanade, T. & Shafer, S. Vision and navigation for the Carnegie-Mellon Navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 10. No. 3 May 1988. ISSN: 0162-8828.

Fan, J.; Zhu, X. & Wu, L. Automatic model-based semantic object extraction algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*. Vol. 11. No. 10. October 2001. ISSN: 1051-8215.

Duda, R.; Hart, P. & Stork, D. *Pattern Classification. Second Edition*. John Wiley & Sons, Inc. ISBN: 0-471-05669-3. Canada. 2001.

Berttozzi, M.; Broggi, A.; Cellario, M.; Fascioli, A.; Lombardi, P. & Porta, M. Artificial Vision on Roads Vehicles. *Proceedings of the IEEE*. Vol. 90. Issue 7. July 2002. ISSN: 0018-9219.

Otsu, N. A threshold selection method from grey-level histograms. *IEEE Transactions on System, Man and Cybernetics*. Vo. SMC. 9. No. 1. January 1979. ISSN: 1083-4427.

Thrun, S et al. Stanley: The Robot that Won the DARPA Grand Challenge. *Journal of Field Robotics*. Vo. 23 No.9. Published online on Wiley InterScience. 2006.

Coronado, J.; Aviña, G.; Devy, M. & Parra C. Towards landmine detection using artificial vision. *Proceedings of International Conference on Intelligent Robots and Systems. IROS/IEEE´05*. Edmonton. Canada. August 2005. ISBN: 0-7803-8913-1.

8. Acknowledgement

The present work was partially founded by Colciencias and Ecos-Nord Program.

Vision Systems: Applications

Edited by Goro Obinata and Ashish Dutta

ISBN 978-3-902613-01-1

Hard cover, 608 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

Computer Vision is the most important key in developing autonomous navigation systems for interaction with the environment. It also leads us to marvel at the functioning of our own vision system. In this book we have collected the latest applications of vision research from around the world. It contains both the conventional research areas like mobile robot navigation and map building, and more recent applications such as, micro vision, etc.The fist seven chapters contain the newer applications of vision like micro vision, grasping using vision, behavior based perception, inspection of railways and humanitarian demining. The later chapters deal with applications of vision in mobile robot navigation, camera calibration, object detection in vision search, map building, etc.