International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:5, No:7, 2011

FPGA based Relative Distance Measurement using Stereo Vision Technology

Manasi Pathade, Prachi Kadam, Renuka Kulkarni, Tejas Teredesai

***Abstract*—**In this paper, we propose a novel concept of relative distance measurement using Stereo Vision Technology and discuss its implementation on a FPGA based real-time image processor. We capture two images using two CCD cameras and compare them. Disparity is calculated for each pixel using a real time dense disparity calculation algorithm. This algorithm is based on the concept of indexed histogram for matching. Disparity being inversely proportional to distance (Proved Later), we can thus get the relative distances of objects in front of the camera. The output is displayed on a TV screen in the form of a depth image (optionally using pseudo colors). This system works in real time on a full PAL frame rate (720 x 576 active pixels @ 25 fps).

***Keywords*—**Stereo Vision, Relative Distance Measurement, Indexed Histogram, Real time FPGA Image Processor

I. Introduction

AS the world progresses towards human comfort and automation, depth perception forms an integral part of automated navigation and motion. There are several techniques that are implemented for the purpose of obstacle detection and distance perception. Some popular methods include use of infrared sensors, ultrasonic sensors, common RADAR technology, a combination of digital camera and Laser, etc. Most of these methods involve recording of time between transmission and receiving of the signal. Other systems may use Laser stripping, optical flow, etc. However the sensors used in these methods are very strongly affected by environmental conditions like temperature, fog etc. Also, these methods provide information only about the distance of the object and not its geometry. The image sensor that we use, i.e. the camera is less affected by environmental parameters and provides information about the geometry of the object too which can be further utilized for navigation and other such purposes.

A. Author is with Electronics & Telecommunication Department of Cummins College of Engineering for Women, Pune, India (e-mail: manasi.pathade@cumminscollege.in)
B. Author is with Electronics & Telecommunication Department of Cummins College of Engineering for Women, Pune, India (Phone: +91-20-25400275; e-mail: prachi.kadam@cumminscollege.in)
C. Author is with Electronics & Telecommunication Department of Cummins College of Engineering for Women, Pune, India (Phone: +91-20-25430157; e-mail: renuka.kulkarni@cumminscollege.in)
D. Author is with Electronics & Telecommunication Department of Cummins College of Engineering for Women, Pune, India. (e-mail: tejas.teredesai@cumminscollege.in)
Financial Support was provided by Wavelet Group, Pune

We cannot perceive depth of an image from its two dimensional representation. Depth here represents the distance of objects in the image from the camera. To extract depth we use a pair of images called stereo images. These images are of the same scene captured by two cameras separated by a horizontal distance. The correlation between object positions in the two images gives us the depth information that we seek. This is the concept of stereo vision and triangulation.

II. System Block Diagram

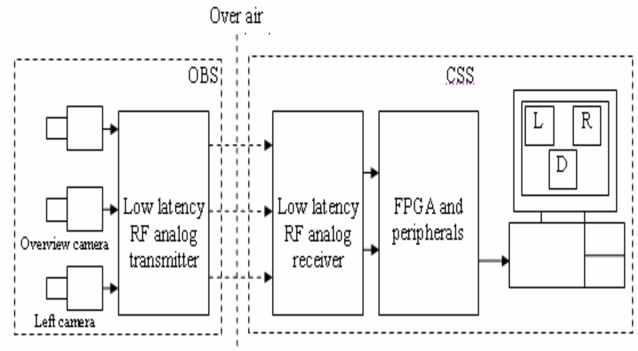

Fig. 1 Block Diagram of Overall System

Description

Inputs received from two cameras (left and right) placed in the field are transmitted through an analog transmitter-receiver pair. These images are received at the processing station. Images are processed to analyze the disparity between the left and right image and the depth of each pixel is calculated. The output is displayed on a TV screen in the form of a left image, right image and a depth image (optionally using pseudo colors). Also, the output in analog format is shown on an analog monitor in full frame size as well as full frame rate. An overview camera may also be provided to give the user a better perspective of the situation.

III. Stereo Vision Technology

Stereo means having 3 dimensions. It comes from the Greek word 'Stereos' which means firm or solid. Stereo vision technology is where 3-D images are used to judge the depth, contrast and distance of objects in our surroundings. 3-D image processing systems have reached the mainstream and are now embedded in a wide range of products including security and surveillance devices, industrial robotics, and autonomous vehicles.This concept of stereo vision is based on the human 'Eyes and Visual System'. This is 'Binocular Vision'. It is nothing but the ability to use two eyes to see an object as one. The human eye catches two images with the help of two eyes and integrates them into one. Thus we can achieve 3-D vision and depth perception.For instance, when you hold up an apple and look at it with just your right eye, the image will be different from the one you see with only your left. Not only that, you will notice that the position of the apple seems to have shifted, relative to the background. The apple will seem to be more to the left side when you look with your right eye, and conversely more to the right side when you look with your left. Yet, when the two images are sent from our eyes to the brain, it processes the two as one image. As the image of the apple from both eyes merged as one, it seems to us to be sharply-defined, detailed and depth-filled.

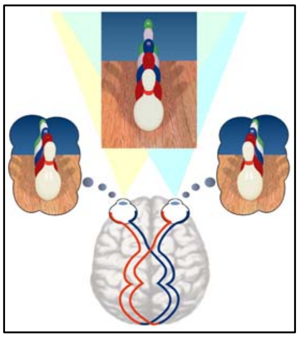

Fig. 2 Binocular Vision

This concept is implemented in the electronic world with two cameras, which mimic the way the human eye captures two images. The two cameras are placed in an Epipolar fashion i.e. they are displaced horizontally. The cameras are then modeled whereby they will see slightly different projections of the world view and thus capture the left and right images. The correlation between these two images gives us the depth information that we seek. We can also obtain information about the geometry of the objects in the images.

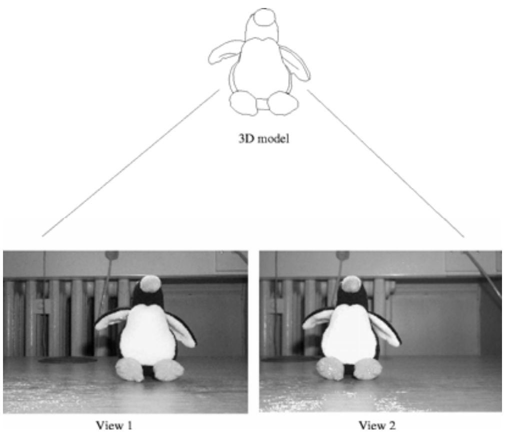

Fig. 3 Concept of Stereo Vision

IV. Algorithm Details

The calculation of dense disparity map in real time poses one of the biggest challenges in Stereovision. Even a fast stereovision algorithm like the pixel-to-pixel algorithm takes slightly less than 250 milliseconds per frame[1] on a PC whereas the requirement is to process each frame within 40 milliseconds. The algorithm suggested by Wang et al[3] , works at 43 fps for images of 320 x 240 and at 16 disparity levels. Keeping the same number of disparity levels, this translates to less than 8 fps for PAL full frame. There are other algorithms like the ones suggested by Tombari et al[2] and Yang et al[4] which work in near real-time. Hence, there was a need to develop a new algorithm that would work in the real time constraint.The basic concept of the algorithm is derived from the concept of stereo geometry:

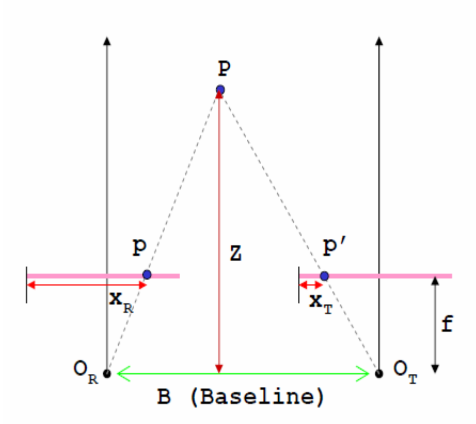

Fig. 4 Epipolar Geometry

Two images of the same scene are taken from different viewpoints. Here P is the point in real world. Z is its actual distance from the camera. p and p’ are the positions of that point in the stereo images. The distance between the viewpoints is called the baseline (B). The focal length of the lenses is f. The horizontal distance of point P taken from left border of image is X_R (For Left Image) and X_T (For Right Image). Under this condition, the difference between X_R and X_T is called the disparity. Using similarity of triangles we can prove that distance Z = B.f/d where d= X_R- X_T.The Epipolar arrangement of the camera system is exploited in the algorithm. Epipolar Arrangement means setting up the cameras such that their image planes are embedded within the same plane. Hence a 2D search domain is narrowed down to 1D search domain as the corresponding points are constrained on the same image scan line. This reduces the complexity of the algorithm. Hence the problem of stereo vision reduces to the problem of finding the correct correspondences in a given set of images. Once that is done, simple triangulation technique can yield the depth information.Suppose L is the length of the Left and Right arrays and N is the search window size in Right array to look for a given pixel in the left array, then the brute force pixel intensity difference algorithm, which is popularly used, requires O(L*N) operations per scan line where L is the length of the Left and Right arrays.

And N is the search window size in right array to look for a given pixel in the left array. In the proposed algorithm, the numbers of operations do not depend on the search window size, but it is constant for each pixel. It needs O (L) operations per scan line. Also, the proposed algorithm has simplicity of implementation that is essential to achieve real-time dense disparity calculations.The indexed histogram records the pixel locations along with the count of the number of pixels of the gray level under consideration along a scan line. This feature affords us quick retrieval of the matching pixel in the left scan line given the gray level of the pixel in the right scan line. Thus, a quick and easy recording of disparity between pixels followed by its inverse relation to distance gives us an almost instantaneous map of the relative distances between the objects within the camera sight.

V.Implementation Details

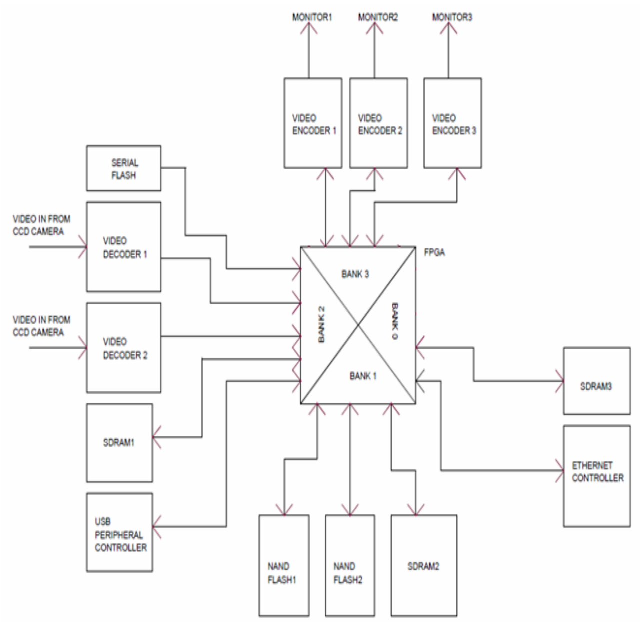

Fig. 5 Block Diagram of Hardware

The input images (left image and right image) are captured by two CCD cameras placed epipolarly. These images are processed through a decoder into video pixel 16 bit output format and given to the FPGA where the disparity is calculated. The FPGA then supplies a 16 bit output to each of the video encoders. The encoder supplies output to the TV in the form of a disparity graph/ false colored image displaying the relative distances of the objects involved in the image. For field applications an Ethernet controller can also be included.

Fig. 6 Sample Input

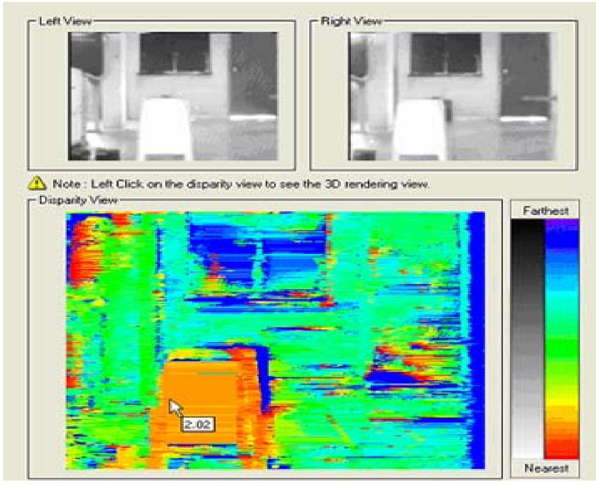

Fig.7 Sample Output

VI. Conclusions

In this paper we have proposed a novel concept of using stereo vision technology effectively to calculate relative distances of objects in front of the cameras. We have also discussed a new disparity calculation algorithm of indexed histogram for matching. This algorithm is very fast and requires fixed amount of computation time irrespective of the images. Implementation details of this system using FPGA have also been mentioned. Thus this system promises high speed, efficiency and reliability. We can find relative distances of objects up to 80 m from the camera.

Applications

1. Security and Surveillance Devices
2. Industrial Robotics
3. Autonomous Vehicles
4. Calculation of Contour Maps
5. Aerial Surveys
6. Tracking Systems
7. Velocity Measurement System

REFERENCES

[1] S. Birchfield, C. Tomasi, "Depth Discontinuties by Pixel-to- Pixel Stereo", International Journal of Computer Vision, Vol. 35(3), pp. 269-293, (1999).

[2] F. Tombari, S. Mattoccia, L. Stefano, E. Addimanda, "Near real-time stereo based on effective cost aggregation", 19th International Conference on Pattern Recognition, pp. 1-4, (2008).

[3] L. Wang, M. Liao, G. Minglun, Y. Ruigang, "High quality real-time Stereo using Adaptive Cost aggregation and Dynamic Programming", Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 798-805, (2006).

[4] Q. Yang, C. Engels, A. Akbarzadeh, "Near Real-time Stereo for Weakly-Textured Scenes", British Machine Vision Conference, (2008).