Proceedings of the International Conference on Man-Machine Systems (ICoMMS)
11 – 13 October 2009, Batu Ferringhi, Penang, MALAYSIA

Developing Stereo Vision System for Object Recognition and Positioning of AMAD-R Mobile Robot

Mohd Saifizi Saidon, Hazry Desa, Paulraj Murugesa, and Mohd Sofian M. Rosbi

Abstract **– Machine vision can be described as a useful robotic sensor since it mimics the human sense of vision and allows for non-contact measurement of the environment. The infinite number of possible poses relative to the viewer will cause the3-D object gives rise to an infinite variety of 2D images or views. Two things must be known—"what" the object is and "where" it is located i.e. the object has to be recognized and its coordinates must be known, In order for a vision system to be effective in assisting a robot to approach an object autonomously.**
This research presents a type of stable target acquisition system for mobile robot. A combination of vision sensors and mobile robot will be use in order to measure accurately the location of a target in real world coordinates. Then, a robust target extraction algorithm that able to pick out a target in an image will develop. After this is done we will develop a stereo vision algorithm to find the distance from the robot to the target.
Finally we will integrate all of these algorithms together. This is important in order to develop a target extraction technique that will be able to find the distance accurately to the target. With the distance we can then locate the real world position of the target.

I. INTRODUCTION

On a robot, stereo vision can be used to locate an object in 3D space. It can also give valuable information about that object (such as color, texture, and patterns that can be used by intelligent machines for classification). A visual system, or light sensor retrieves a great deal of information that other sensors cannot.
A camera head that connects the camera with a robot enables software-steered panning and tilting of the camera. Figure 1.1 shows a mobile robot, which is equipped with a camera. The image data are interpreted from an image-processing module that have to try to detect an object in the image using image-processing operators. The object will be analyze and reconstruct after detection is successful in order to determine its world coordinates.
A stereo vision system, consisting of twin camera, is available for the position estimation of the target in the workspace.[1-2] Stereo camera triangulation from figure 1 exploits geometrical realities to determine the distance of the object point from the focal point. This paper is focusing in study of developing stereo vision mobile robot for indoor environment navigation system.

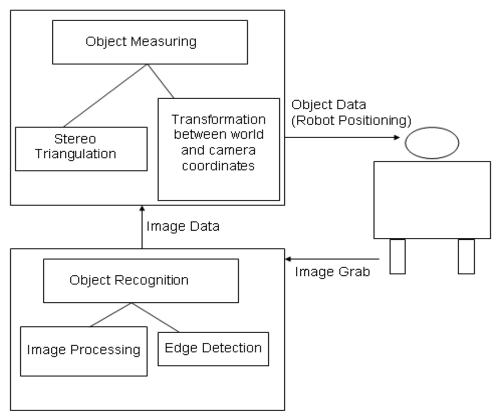

Fig.1. The architecture of a vision-based robot

II. SOFTWARE DEVELOPMENT

All software will be write using National Instruments Labview version 7.1, which creates programs using the G programming language. Labview was design to interface with data acquisition systems and perform analysis. This software led to the creation of a programming environment that simplifies communication with external devices. It also allows easy creation of graphical user interfaces with mechanical shapes and knobs. The inputs and outputs of the block diagram in Labview are linked to the controls and indicators on the user interface, called the front panel. The block diagram and front panel are shown in figure 2 and 3 respectively.
Labview with its graphical nature gives lot of advantages to the users by allowing them to see the state of any variables while the program is running. A rich library of indicators allows quick implementation of charts and visualization of images or arrays. Using these powerful indicators, debugging programs becomes easier and reducing development time.

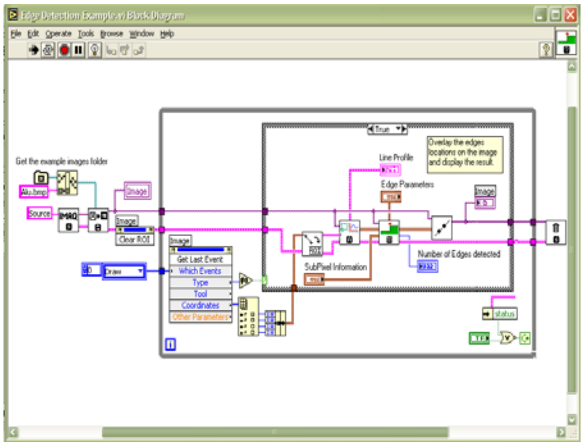

Fig.2. Labview Block Diagram

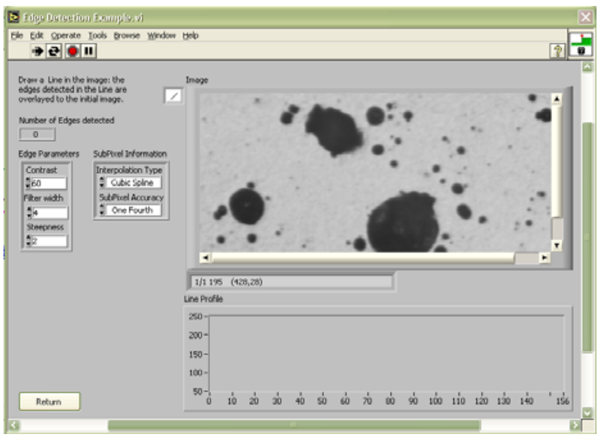

Fig.3. Labview Front Panel

III. STEREO VISION SYSTEM EQUIPMENT

A. Lenses

The camera lens is the interface between the environment and the sensor. A properly chosen lens will improve the quality and range of the results.

A large issue with the use of computer vision in an environment is variable lighting. Whether monocular or stereo, if the cameras being used create images from the visible light spectrum this will be an issue. Image processing will yield different qualities of results based on the lighting situation.

In conditions where the camera is gathering too much light, the image becomes over-exposed and will appear washed out or even completely white. Conversely, if the camera does not gather enough light, the image is under-exposed and large areas will appear black. In an indoor testing environment, the amount of light in the room can be fixed.

A camera's iris acts much like the iris of a human eye. The iris is an adjustable aperture, which can be made larger or smaller. With a larger aperture, more light is allowed to enter the camera. A smaller aperture allows less light. Camera lenses can have a manual iris or an auto-iris. The manual iris is adjusted by the user while the auto-iris uses feedback from the camera to make adjustments.

The lens's f-stop is a measurement of the size of its iris aperture. The number represents the relationship between the diameter of the opening and the focal length.

$$F - stop = \frac{f}{d} \quad (1)$$

In fixed lighting conditions, the user may set the iris of the camera to gather an optimal level of light prior to performing the task. An optimal level for stereo vision is one in which the images show features with the greatest texture.

Camera lenses may have variable focal lengths or fixed focal lengths. A variable focal length lens can zoom in and out. For this application, fixed focal length lenses were desired as variable focal lengths would add great complexity to the system.

Lenses with larger focal lengths create images that are zoomed in farther. It was desirable to choose a focal length that would allow the system to detect objects far enough away to provide adequate time for obstacle avoidance. The trade off is that the greater the focal length, the narrower the field of view. The field of view (FOV) of a lens can be computed by;

$$FOV_{horizontal} = 2\tan^{-1}\left(\frac{x}{2f}\right) \quad (2)$$

$$FOV_{vertical} = 2\tan^{-1}\left(\frac{y}{2f}\right) \quad (3)$$

where x is the horizontal width of the sensor, y is the vertical height of the sensor and f is the lens focal length.

B. Cameras

A stereo camera pair must have two identical cameras rigidly mounted so that they will not move with respect to each other. Cameras are available with a multitude of options. Some of the most important questions to consider are what kinds of outputs are required for the task, lens compatibility, shutter speeds, resolution, and ruggedness.

The decision of which camera to use is depend on the choice of method for image transfer from camera to computer. There are several formats for the signal that the camera sends containing the images. The format influences the speed of data transfer, image quality and resolution. Cameras that send analog signals must use a frame grabber (also called a capture card) to digitize the images.

C. Image Transfer

Three common analog video formats are described below.

One video signal format is s-video (separated video), also known as Y/C. In this S-video format (sometimes called S-VHS), the chrominance (color or C) information is kept separate from the luminance (intensity or Y) and sync information, to reduce the possibility of interaction. It has a resolution of 480 interlaced lines in NTSC format and 576 interlaced lines in PAL format.

Another format is YUV. It is commonly used in video applications, where it is also referred to as component video. YUV is a color space in which the Y stands for the luminance component (the brightness) and U and V are chrominance (color) in one composite analog signal. A yellow RCA type connector is usually used to transmit composite video. Like s-video, it can also be used with NTSC or PAL format with the same resolution.

RGB video stands for red, green, and blue analog signal. Sometimes one or two more signals are sent with synchronization information. This format can send images with a resolution up to 1080 progressive scan lines and is better for tasks requiring very high resolution images.

Analog image signals must be converted to digital image for computer processing. Frame grabbers are used most often and typically consist of hardware that can be inserted into a PCI slot. For stereo vision, two images must be transferred to the computer at the same time. If the system uses multiple pairs of cameras, it may be desirable to transfer all of the images to the same computer. This with the input capabilities of the computer hardware used for processing should be taken into consideration when choosing a conversion method.

Because the images will be processed and not simply stored or displayed, it is necessary to choose a frame grabber that comes with a library for programming user applications rather than just commercial software.

D. Equipment Used

The decision was made to use manual-iris lenses because of the lighting is fixed. Vali-Focal Zoom lenses were chosen. The irises of these lenses range are from F1.6-close. Focal lengths of 8.5mm and 51.0 mm were tested to see which would provide the better data range.

This dictated the direction of the rest of the system hardware. Toshiba Teli CCD Cameras with analog NTSC output were the first choice. The camera allows for less data transfer and a faster system with 30 FPS.

Upon searching for a frame grabber that was suitable for the task, the RTV Series Multi-Channel Real-Time Video was selected. It accepted standard composite color (PAL, NSTC) or monochrome video format (CCIR, EIA). It also allows for a more compact system as the computer is only required to have one PCI slot. The card support VC++/VB/Delphi/C++ Builder programming under windows XP. This card also assist Angelo-LVIEW that fully compatible with LabView 6.0 and above.

A personal computer (Intel Core 2 Duo Processor, memory 4GB, Windows XP Professional and Speed Setup Graphics Card) was used to process images.

V. STUDY OF IMAGE PROCESSING

A. Image Processing

The design of the hallway following process in mobile robot is based on the following justification;

- In order for a robot to go straight along a hallway, the perspective projection in the camera image of either the left or the right hallway edge must be within a certain angular range.
- If the robot is not headed straight down the hallway- a condition that could be caused by an attempt at crash avoidance.

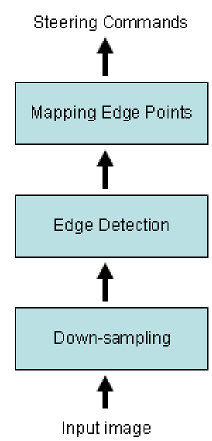

Fig.4.The Regulator Cycle

Fig.4 displays the flow of processing that takes a camera image for input and produces steering commands for the robot at the output. As shown in the flow diagram in Fig. 4, the camera image is first down-sampling. The down-sampling the image does not have abilities to approximate orientation and location of the hallway the corridor follower. A robust method for detecting edges that may be continuous or broken from an image consists of first applying an edge detector to an image.

In image processing step, we should know a lot of technique like;

- Segmentation- to separate the object from the background.
- Threshold- to remove parts of the image that fall within a specified intensity range.
- Edge detection- to analyze the pixels to detect the edges.

B. Segmentation

In the analysis of the objects in images it is essential that we can distinguish between the objects of interest and "the rest". The segmentation techniques that are used to find the objects of interest are usually referred to as segmentation techniques - segmenting the foreground from background.

In detecting object, segmentation of the image is to separate the object from the background. There are various approaches to this, with varying degrees of effectiveness. In this paper we study two of the most common techniques: threshold and edge finding.

C. Threshold

Threshold means pixel value set to discriminate dark and light intensities. This technique is based upon a simple concept. This technique can be adapted easily to the case where we have light objects on a dark, dominant background. Further, it can be used if the object peak dominates and we have reason to assume that the brightness distribution around the object peak is symmetric.

D. Edge Detection

Segmentation subdivides an image into its constituent regions or objects. Segmentation algorithms for monochrome images generally are based on the basic properties of image intensity values. Edges are places in the image with strong intensity contrast. Since edges often occur at image locations representing object boundaries, edge detection is extensively used in image segmentation when we want to divide the image into areas corresponding to different objects. Representing an image by its edges has the further advantage that the amount of data is reduced significantly while retaining most of the image information. We can use Sobel, Robert, Prewitt, Canny edge detector algorithm to extract edge since it can detect edges with noise suppressed and be adaptable to complex environment.

VI. Study of Basic Stereo Vision Principle

Reconstruction of the world seen through stereo cameras can be divided in two steps:

- Correspondence problem. For every point in one image find to the corresponding point in the other image and compute the disparity (distance in pixels) of these points [3].
- Triangulation. Given the disparity map, the focal distance of the two cameras and the geometry of the stereo setting (relative position and orientation of the cameras) compute the (X, Y, Z) coordinates of all points in the images.

A number of textbooks give adequate descriptions of the theory and practice of stereo reconstruction and 3-D world perception. To generate accurate three-dimensional information from stereo analysis, we need to know the geometric parameters of the stereo cameras. These parameters can be classified in two groups:

- Internal (or intrinsic): geometric and optical characteristics of the lenses and the imaging device. Generally, we will try to find parameters that relate the actual cameras to ideal pinhole cameras. Correction of lens distortion is very important.
- External (or extrinsic): position and orientation of the camera in a world reference system.

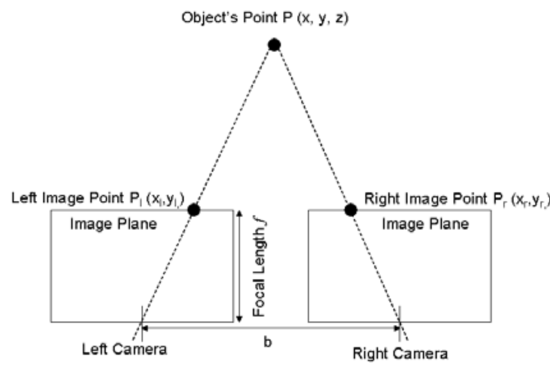

Fig.8. Illustration of Stereo Geometry

Fig. 8 illustrates the geometry of stereo vision. In this example, the optical axes of the cameras are aligned parallel and separated by a baseline of distance, *b*. A coordinate system is attached in which the x-axis is parallel to the baseline and the z-axis is parallel to the optical axes. The points labeled “Left Camera” and “Right Camera” are the focal points of two cameras. The distance *f* is the perpendicular distance from each focal point to its corresponding image plane. Point P is some point in space which appears in the images taken by these cameras. Point P has coordinates (x, y, z) measured with respect to a reference frame that is fixed to the two cameras and whose origin is at the midpoint of the line connecting the focal points. The projection of point P is shown as P_r in the right image and P_l in the left image and the pixel coordinates of these points are written as (x_r, y_r) and (x_l, y_l) in terms of the image plane coordinate systems shown in the figure. Note that the disparity defined above is x_l-x_r. Using simple geometry, the 3D word coordinates (x,y,z) are computed as:

$$x = \frac{x_l b}{x_l - x_r}, y = \frac{y_l b}{x_l - x_r}, z = \frac{fb}{x_l - x_r} \quad (4)$$

The accuracy of 3-D data reconstruction depends on the accuracy of the disparities, stereo system calibration, images rectification and overall stereo system construction.

The camera-based 3D coordinate acquire in the stereo vision system needs to be transform into the robot-based 3D coordinate, which is necessary for the motion planning of the mobile robot. After the coordinate transformation, the generated robot-based 3D coordinate will be sent to the motion control system for the automatically target positioning of the mobile robot.

REFERENCES

[1] S.Das, et.al. :, "Performance Analysis of Stereo", IEEE Trans. Patt. Anal. and Mach. Intell., Vol.17, No.12, pp.1213-1219, Dec.(1995)

[2] Y.Kuniyoshi, et.al. : "Deferred imitation of human head movements by an active stereo vision head", IEEE Proc., 6th Int. Work. Rob. Human Comm. (ROMAN97), pp.88-93, Sep. (1997)

[3] F.G. King, G.V. Puskorius and F. Yuan, "Vision guided robots for automated assembly", Proc. 1988 IEEE Int. Conf. Robotics and Automation, Vol. 3, April 1988, pp. 1611-1616.