OBJECT DISTANCE DETERMINATION APPLICATION USING IMAGE PROCESSING TECHNIQUE

LIM KIM KHIAM

A thesis submitted in partially fulfillment of the requirements for the award of degree of Bachelor of Computer Science (Software Engineering)

Faculty of Computer Systems & Software Engineering
Universiti Malaysia Pahang (UMP)

MAY 2011

ABSTRACT

Object distance determination application using image processing is develop to determine the distance of an object in a pair of 2D images using pair stereo vision. This application is develop based on the human vision feature that not commonly applied in automotive and mobile robotic fields nowadays. Therefore this project is carried out to build a prototype to determine the distance of an object in 2D image and identify the image processing technique for determine object distance. Rapid application development is applied as a guidance of development and the flow design of this application consist of four steps which is image acquisition, image filtering and enhancement, feature extraction and object detection and the last step is output calculation. The formula of output is generating base on the linear graph with linear equation $y = mx + c$. The tools that use to develop this application are visual c++ with opencv library as engine and MFC as application interface. The testing of the application is using 100 samples with different object and environment. The average error rate is about ±2.93 % after testing phases. The application is giving benefit to car manufacturer and mobile robotic manufacturer to enhance their automation function and increase the reliability.

ABSTRAK

Aplikasi Penentuan Jarak Objek mengguna teknik *image processing* dibangunkan untuk menentukan jarak objek dalam sepasang gambar 2D dengan visi stereo. Aplikasi ini telah dibangunkan berdasarkan ciri-ciri visi manusia yang tidak umum diguna dalam bidang automotif dan *mobile robot.* Oleh demikian, projek ini dibina berdasarkan objektif membangunkan prototaip untuk menentukan jarak suatu objek dalam gambar 2D dan memgenalpastikan teknik *image processing* untuk menentukan jarak sesuatu objek dalam 2D gambar. Metodologi *Rapid Application Development* telah diguna sebagai panduan pembangunan aplikasi dan aplikasi ini mangandungi empat langkah iaitu perolehan gambar, proses peningkatan kualiti gambar and penapisan gambar, ekstraksi ciri-ciri dan menentu objek dan akhirnya pengiraan jarak. Fomula pengiraan jarak dihasilkan melalui graf garis lurus dengan fomula $y = mx+c$. Alat-alat yang digunaan semasa pembangunan aplikasi adalah visual c++ dengan openCV *library.* Antaramuka aplikasi dibangunkan dengan MFC. Ujian aplikasi dijalankan dengan 100 sampel yang mangandungi objek dan persekitaran yang berbeza.Purata tingkat kesalahan dalam lingkungan ±2.93% selepas ujian dengan 100 sampel. Aplikasi ini memberi manfaat untuk pengeluar kereta dan pengeluar *mobile robot* untuk meningkatkan fungsi automasi dan meningkat kebolehpercayaan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter is present about the Object Distance Determination Application using Image Processing Technique. Contents of this chapter include introduction, objective, problem statement and scope of the application and thesis organization.

1.1 Introduction Object Distance Determination Application (Using Image Processing)

Object Distance Determination Application is an application that applies image processing technique to determine distance of objects base on two 2D images. This application is using pair stereo vision which mean will function together with 2 web cameras with same specification to capture the image and process the image to determine the objects distance. This application is a real time application with simulating one of the human vision functions. This application can be applied in mobile robotic field and automotive field which able to help in avoid crashing with other object.

1.2 Problem Statement

Nowadays, automotive and mobile robotic has become widely in use by human in their daily activities. Most of these tools having some problem such as:

1) People unable to view blind spot of car side mirror
2) Crashing with object on road without notice or careless driving
3) Sonar and light sensor of a mobile robotic has limited detecting area

Generally, a normal human vision can simply determine the distance between he/she and the object. This can help them in avoiding collide with the object. Therefore, by applying this human vision feature on automotive and mobile robotic, it able to help in detect the distance of the object when they are closing with some object. Thus collision can be avoided. For an example, a riding automotive can detect the object on the road such as plastic traffic cone and measure the distance to remind driver to avoid crashing with it. A mobile robotic can detect the obstacle and avoid crashing with it.

1.3 Objectives

1) To identify image processing technique for determine the distance of objects
2) To develop a prototype for determine the distance of an object in 2D image

1.4 Scope

1) The target user of the application is mobile robotic user and automotive user.
2) The tools that use to develop the application is openCV and Visual C++.
3) The device that use in the application is 2 web cameras with same specification.

4) The image is still image that imported from a video stream with single rectangle object.
5) The application will test by 100 samples
6) The range of distance that able to detect is 25cm-75cm (with 4 cm distance between cameras).
7) The image resolution is 640*480
8) The prototype application is assume to run in ideal environment.

1.5 Thesis Organization

This thesis consists of 6 chapters ranging from Chapter 1 until Chapter 6. Chapter 1 gives an overview of the study conducted. It also supply with the problem statement, objective and the scope of the study. Meanwhile, Chapter 2 reviews the previous research works that was conducted by other researches. All the relevant technical paper, journals, and books taken from those researches will be discussed in detail. Chapter 3 discuss on methodology that will applied through the whole development and also reveals the techniques and the algorithms that will be used in performing this study. It will discuss about the process flow in detail of this research. Details of the implementation of the study will be discussed in Chapter 4. Results of the testing are to be expounding in Chapter 5. Lastly, Chapter 6 concludes the entire thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Image Overview

Commonly, an image can be express as a 2D representation of 3 dimensional scenes. A digital image is express as numerical representative of an object. [1]

In term of technical interpretation, an image can be defined as a two-dimensional function f(x. y),where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y and the intensity values of f are all finite, discrete quantities, we call the image as a digital image. To be easier in understanding, an image can be defined as a digital image when it meets 2 criteria, which are:

- It must in 2 dimensional model and each of the point (can call as pixel) of the image will have unique spatial coordinates
- Each point have intensity or gray level (can also define as pixel value) must be a finite value [2]

Nowadays, images are widely and commonly in use of human daily life application such as satellite, television, computer tomography as well as in area of research and technology such as astronomy. [1] They use image to describe data or result, and also obtaining information from image to produce some result.

The type of image will be discuss is grayscale image in part 2.1.1 and color image in part 2.1.2.

2.1.1 Grey-Scale Image

The basic two-dimensional grey-scale image is a monochrome image which has been digitized.

A digitized grey-scale image is one where contain three characteristics which are:

- Spatial and grayscale values have been made discrete.
- intensity measured across a regularly spaced grid in x and y directions
- Intensities sampled to 8 bits (256 values). [3]

Another definition that define by R. Fisher, S. Perkins, A. Walker and E. Wolfart says that grayscale (or graylevel) image is simply one in which the only colors are shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel. In fact a `gray' color is one in which the red, green and blue components all have equal intensity in RGB space, and so it is only necessary to specify a single intensity value for each pixel, as opposed to the three intensities needed to specify each pixel in a full color image.

Often, the grayscale intensity is stored as an 8-bit integer giving 256 possible different shades of gray from black to white. If the levels are evenly spaced then the difference between successive graylevels is significantly better than the graylevel resolving power of the human eye.

Grayscale images are very common, in part because much of today's display and image capture hardware can only support 8-bit images. In addition, grayscale images are entirely sufficient for many tasks and so there is no need to use more complicated and harder-to-process color images. [4]

An example of grayscale image was shown in Figure 2.1.

Figure 2.1: Grayscale image [12]

2.1.2 Color Image

When talk about color image, it can define as images represented in the RGB color model which consist of three components image, one for each primary color. When fed into an RGB monitor, these three images combine on the screen to produce a composite color image. The number of bits used to represent each pixel in RGB space is called the pixel depth. Consider an RGB image in which each of the red, green, and blue images is an 8-bit image. Under these conditions each RGB color pixel [that is a triplet of values (R, G, B)] is said to have a depth of 24 hits (3 image planes times the number of bits per plane). The term full-color image is used often to denote a 24-hit RGB color image. The total number of color in a 24-bit RGB image is (2^8) ^3 = 16,777,216.

It is of interest to note that acquiring a color image is basically the process shown in Fig. 6.9 in reverse. A color image can be acquired by using three filters, sensitive to red, green, and blue respectively. When we view a color scene with a monochrome camera equipped with one of these filters. The result is a monochrome image whose intensity is proportional to the response of that filter. Repeating this process with each filter produces three monochrome images that are the RGB component images of the color- scene (In practice, RGB color image sensor usually integrate this process into a single device [2]

An example of color image was shown in Figure 2.2

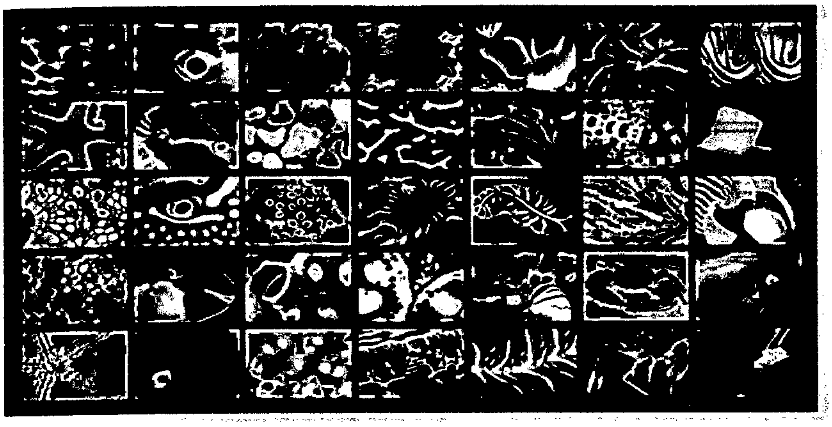

Figure 2.2: Color image [13]

2.2 Image Processing Technique

The field of digital image processing refers to processing digital images by means of a digital computer. Note that a digital image is composing of a finite number of elements, each of which has a particular location and value. These elements are called picture elements, image elements, pels, and pixels. Pixel is the term used most widely to denote the elements of a digital image. [1]

The image processing technique that will be making use in the application will be discussed in part 2.2.1, 2.2.2, 2.2.3 2.2.4 and 2.2.5.

2.2.1 Threshold

When going to understand an image, the most important prerequisite is to selecting the features within the image. Traditionally, it could be done by define a range *of brightness values in the original image,* select the pixels within this range as belonging to the foreground, and reject all of the other pixels to the background. Such a image is then usually display as a binary or two-level image, using black and white or other colors to distinguish the regions. [5]

There is a lot of threshold algorithm that available nowadays and according to the research, the threshold algorithm can basically categorize into six group:

1. histogram shape-based methods, where, for example, the peaks, valleys and curvatures of the smoothed histogram are analyzed
2. Clustering-based methods, where the gray-level samples are clustered in two parts as background and foreground (object), or alternately are modeled as a mixture of two Gaussians
3. Entropy-based methods result in algorithms that use the entropy of the foreground and background regions, the cross-entropy between the original and binaries image, etc.
4. Object attribute-based methods search a measure of similarity between the gray-level and the binaries images, such as fuzzy shape similarity, edge coincidence, etc.
5. The spatial methods use higher-order probability distribution and/or correlation between pixels
6. Local methods adapt the threshold value on each pixel to the local image characteristics. [6]

For most of the threshold function, users are allowed to choose the threshold value (mostly from 0-255) or the threshold algorithm can compute the value automatically which knew as automatic threshold. [7]

From all the statement above, it can be said that threshold technique can be apply to differentiate the background and foreground in an image (Segmentation) base on some calculation of pixel value and convert the image into binary image. Even there was a lot of threshold algorithm available nowadays but most of the threshold algorithm is depends to the threshold value. Figure 2.3 show then image before and after threshold is applied.

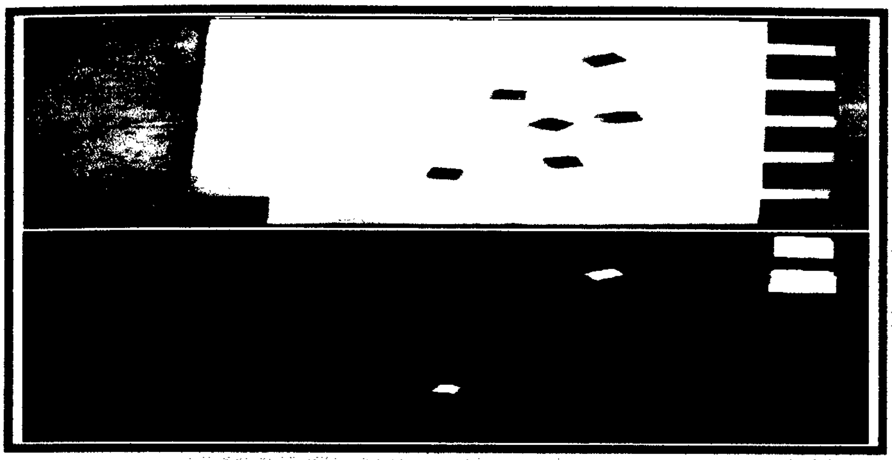

Figure 2.3: Example of threshold image [14]

2.2.2 Mask

A mask in image processing technique usually represent the template the will stack to an image to keep the region of interest and remove other region. More specific, a mask usually in binary form (black and white color). Mask operation can be done in several physical ways. The binary mask can be use in as overlay, or alpha channel, in the display hardware to prevent pixels from being display. Mask can also be use to modify image storage by multiplying with a grayscale image by the mask. If the mask is white for the background and black for foreground pixels then the brighter pixel values at each location will erase all background pixels and keep the grey value for the foreground pixels. [5]

2.2.3 Convolution Matrix

Other than process image pixel by pixel, there is another very common operation on images performed in the spatial domain which is known as convolution. A convolution using represent in a matrix form which is multiplied by each pixel and its neighbors in a small region whether can be 3*3 or 5*5. This is applied to all of the pixels in the image and in all cases, the original pixel values are used to apply some mathematic operation for example: multiplication and addition, and the new pixel value will replace the old pixel values to produce a new image. A convolution is usually use to smoothing the image. [5]

2.2.4 Gaussian Blur

The Gaussian blur which also known as Gaussian smoothing operator is a 2-D convolution operator that used to `blur' images and remove detail and noise. It is similar to mean filter, but it uses a different kernel that represents the shape of a Gaussian (`bell-shaped') hump. This kernel has some special properties.

2-D distribution as a `point-spread' function is the basic idea that use in Gaussian blur, and this can be achieved by convolution. Since the image is stored as a collection of discrete pixels we need to produce a discrete approximation to the Gaussian function before it can perform the convolution. In theory, the Gaussian distribution cannot be zero, which will require an infinitely large convolution kernel, but in practice it is effectively zero more than about three standard deviations from the mean, and so that the kernel can be truncate at this point. [4]

2.2.5 Image Contour

Contour usually represents in a line which provides boundary information and the line is guaranteed to be continuous. In a constant brightness in the image, the lines cannot end, although the line may branch or loop back upon themselves. But in a continuous image or an actual topographic surface, the line is end up until certain point since there is always a point through which the line can pass.

The contour line can principle, be fit as a polygon throughout the points interpolated among pixel centers for all such pairs of pixels that bracket the contour value. It permits measuring the locations of these lines, and the boundaries that may represent, to less than the dimensions of the pixel, called sub-pixel sampling or measurement. This is seldom done for the whole image because of the computation amount involved and the complexity in representing the boundary by such a series of points, which must be put into a polygon.

Instead, most of the use of the contour lines is to mark the pixels the lie closest to, or closest to and above. These pixels near the contour line to the resolution of the pixels in the original image; form a continuous band of touching pixels (touching in an eight neighborhood sense), and can be used to delineate features in many cases. Creating the line from the image is simply a matter of scanning the pixels once, comparing each pixel and its neighborhood above and to the left contour value, and marking the pixel if the values bracket the test value. [5]

2.3 Existing System

A lot of researchers and investigators are working rapidly for estimation of the distance of an object from in image or video. That is more than one procedure has been discover nowadays [9]. Based on this statement, there is already proof that estimating of distance based on image is no longer a new issue and technique nowadays. Base on the existing case study, the technique that study can be categorized into 2 categories, which is single vision technique and pair visions technique.

2.3.1 Single Vision Technique

This technique was proposed by P. Gil, S. Lafuente and S. Maldonado and F.J. Acevedo. This technique was applied to video surveillance systems by using single vision with defocusing image to estimate the distance of a moving object.

This approach will start with estimating the background, means the structure of background will be learned by the system to make comparison in the result. After estimating background, this approach will continue with depth estimation. An object that situated in a distance will not render a perfect focused image, therefore some process are necessary to apply in order to get the result. In depth estimation, the first process that applies is convolution operation and the formula is given as below:

$$i(x,y) = i_f * h(x,y) \qquad (1)$$

Where *i(x,y)* define as defocus image and *if* (x,y) is define as focused image. For h(x,y), it was call as point spread function and it involve the amount of blur that apply to focus image to obtain the defocus image. PSF can also be model into pillbox function and pillbox function is defined as below:

$$h(x,y) = \frac{1}{\pi r_b^2} II\left(\frac{\sqrt{x^2+y^2}}{2r_b}\right) \qquad (2) \qquad (2)$$