MONOCULAR CAMERA BASED MACHINE VISION MODEL TO ESTIMATE DISTANCES

MOHAMMED ASAD REHMAN

School of Interdisciplinary Science & Technology, IGNOU-I2IT Centre of Excellence for Advanced Education & Research, Pune, India

Abstract- This paper presents a solution to the problem of measuring distances in known surrounding. Following text portrays an algorithm to estimate the actual distance of a symmetrical object of known dimensions with the use of a single camera. The distance calculation is performed using Digital Image processing without need of any additional sensors or reflective mirrors. Camera takes picture of object at exact location from where distance is to be estimated. Algorithm provides accurate estimations under acceptable error percentages. The algorithm can be implemented on live videos too.

Keywords- *Distance Measurement, Monocular Camera, Spatial Image Processing, Thin-lens formula.*

I. INTRODUCTION

Presented here is a Digital Image Processing technique to estimate distances, by applying thin-lens equation on pixel information of target object.

Other methods of distance estimation and measurement uses stereo vision, ultra-sound, laser projector, mirror etc. as additional equipments with camera to work [1][2].

Each of the aforementioned methods uses either a pair or additional equipments with camera to estimate distance.

This paper presents a method algorithm to estimate the distance by using single camera, which directly leads to lesser hardware and simpler application.

This work devices a machine vision model to estimate distances of objects in known surrounding using a single(monocular) camera. The work is done in spatial domain of image processing by direct manipulation of pixels in an image. This method is the one in which input is an image and the output is based on attributes extracted from the image.

The current paper is divided into five parts namely: I. Introduction, II. Algorithm Equation, III. Implementation, IV. Experimental Results, and V. Conclusions.

II. ALGORITHM EQUATION

In image analysis, it becomes extremely convenient to represent, interpret and analyze the three-dimensional world using co-ordinate systems often represented by X; Y and Z. Of the many possible notations, the widely used and easier one to remember is shown in figure 1. As can be seen from the figure, width is represented along the X-axis and height along the Y-axis. The positive going Z-axis extends into the world, away from the viewer.

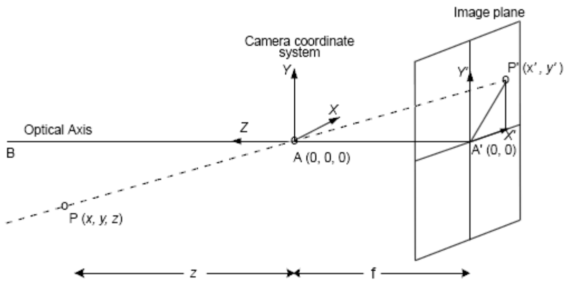

Figure 1. Image plane and co-ordinate geometry.

Thus, if an object is in view (or in front of the observer), then its Z value is positive.

Farther the object is from the viewer, higher is the value of Z. In certain cases, the object may move and be located behind the observer in which case Z is negative. Although various other conditions are possible, they are treated as special cases and hence shall not bother.

In the above figure 1, point A represents the origin which will be the aperture of the camera. Similarly the image plane will be the retina of the camera. Usually the focal length ' f ' can be varied for a clear view of the object or scene. Each point in the three-dimensional scene corresponds to a point (x', y') on the two-dimensional image plane. Using the relation of similar triangles, we can write the relation [3]:

$$\frac{x'}{x} = \frac{y'}{y} = \frac{f}{z} \qquad .1$$

or

$$x' = \frac{f}{z}x \qquad .$$

and

$$y' = \frac{f}{z}y \qquad .$$

This in simple words is thin-lens equation [4]:

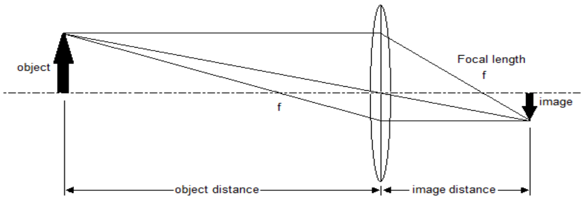

Figure 2. Thin Lens Equation.

Given by formula:

$$\frac{1}{object\ distance} + \frac{1}{image\ distance} = \frac{1}{focal\ length}$$

III. IMPLEMENTATION

The method is applied using a typical general-purpose system for digital image processing [5].

A. Hardware:

The aforementioned algorithm can be implemented on a image processing system having:

- A digital image sensing device;
- An image processing hardware/computer having CPU clock rate of 1.06GHz or higher;
- Image Processing software;
- Mass storage unit and
- Image display console.

B. Image Processing steps:

Following image processing ladder is followed from top to bottom every time an image is captured by image sensing device:

TABLE I. IMPLEMENTATION LADDER

<table>
<tr><td colspan="2" rowspan="2">Step Number</td><td colspan="2">Image Processing Ladder</td></tr>
<tr><td colspan="2">(Step name)</td></tr>
<tr><td colspan="2">1</td><td colspan="2">Acquire Image</td></tr>
<tr><td rowspan="3">2</td><td>A</td><td rowspan="3">Threshold Object</td><td>Extract image channels</td></tr>
<tr><td>B</td><td>Color condition filter</td></tr>
<tr><td>C</td><td>Extract foreground & background</td></tr>
<tr><td colspan="2">3</td><td colspan="2">Morphological Enhancement</td></tr>
<tr><td colspan="2">4</td><td colspan="2">Remove Background</td></tr>
<tr><td colspan="2">5</td><td colspan="2">Get area of object pixels</td></tr>
<tr><td colspan="2">6</td><td colspan="2">Algorithm Implementation</td></tr>
<tr><td colspan="2">7</td><td colspan="2">Display Result</td></tr>
</table>

1. Acquire Image

The image sensor part of work acquires the image of the object. The sensor device gives output in machine understandable image format (.jpeg) which is further processed to get desired results.

2. Threshold Object

Image threshold is the step which gives out the object's image and distinguishes object pixels from the complete acquired image.

2.A. Extract Image channels

This step of threshold extracts the constituent channels from image format (Red, Green & Blue in present case) thus providing three binary images for processing.

2.B. Color condition filter

The color levels of three image channels are individually set to get the exact levels of desired object using logical operators and then the three channels are merged together to form back threshold object image.

2.C. Extract Foreground and Background

All the remaining pixels which are not the desired object but are there because they qualify the color filter are removed by further filtering out the pixel groups of size other than object pixel group size. Thus yielding a pixel group left over, this is the required object [6].

3. Morphological Enhancement

The image thus obtained is eroded from sides and over exposed areas. To extract the exact dimension of object, the image is dilated using appropriate structuring elements. Structuring elements are predefined pixel groups.

The operation of dilation [7] is defined as:

$$A \oplus B = \{z | [(B')_z \cap A] \subseteq A\}$$

Where,

A = object pixel group.
B = Structuring Element pixel group.
$(B')_z$ = Reflection of B about its origin and shifting of this reflection by z .
z = an integer.
$A \oplus B$ = Dilation of image A by B.

4. Remove Background

The dilated image is cleaned of background noise and only object pixels in the foreground are '1' rest of the image is '0'.

5. Get area of object pixels

Count the object pixels (i.e. pixels with value '1'), and using the relation between pixel resolution of screen & size of screen (in metric units). The area of object is obtained.

6. Algorithm Implementation

The area of image thus obtained is fed to the modified thin lens formula (discussed in Section II)

$$x' = \frac{f}{z} x$$

from

where,

x' = width of image
= square root of object pixel area.
f = focal length of camera.
z = distance of object to camera.
x = actual size of image

= square root of object's size (face area).

7. Display Result

The result of above calculation is displayed as title of image showing the test object as highlight.

IV. EXPERIMENTAL RESULTS

The above mentioned process was implemented on a set of pictures taken in standard lighting conditions. The object was a black box of size 41.5cm (width) and 41cm (height).To verify the algorithm 16 images were alternately picked from a series of 32 images taken at regular distance ranging from 140cm to 290cm.

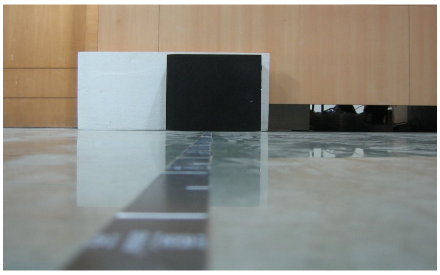

Figure 3. Image captured by Imaging Device at 230cm distance.

All the resultant estimated distances were in the ±9% error range. On plotting product of estimated distance (in cm) & square root of object pixel area calculated (in cm) against image series number [8] a constant graph of variation ±0.043 is obtained.

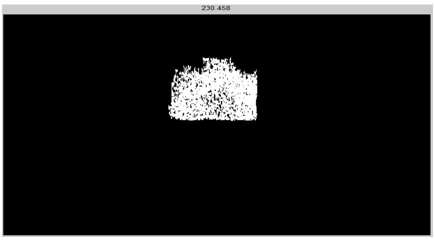

Figure 4. Output image shown in console for object at 230cm distance. The object pixels are highlighted as white and the distance is maked as 230.458 (in cms) which has +0.199% error.

V. CONCLUSION

This paper presents a distance measuring method for thin lenses. Width of lens is a key role player in this application. Object plane should face the sensing device in a perpendicular manner resulting in maximum exposure. The method is light exposure dependent a change in exposure may require recalibration of color levels; however use of HSV color model can uproot this limitation.

This method is not suitable for microscopic distances as the thin-lens equation works on the assumption that the object distance is much larger than the lens thickness.

REFERENCES

[1] Hyongsuk Kim, Chun-Shin Lin, Jaehong Song, and Heesung Chae, "Distance Measurement Using a Single Camera with a Rotating Mirror", IJCAS, vol. 3, no. 4, pp. 542-551, December 2005.

[2] Christian Plagemann, Felix Endres, J□urgen Hess, Cyrill Stachniss, Wolfram Burgard, "Monocular Range Sensing:A Non-Parametric Learning Approach", Department of Computer Science, University of Freiburg.

[3] R. Jain, R. Kasturi, and B. G. Schunck, "Machine Vision". McGrawHill, 1995.

[4] Eugene Hecht, "Optics", Fourth Edition, Sec.5.2.3.

[5] Rafael C. Gonzalez & Richard E. Woods, "Digital Image Processing: Second Edition," Pearson Education, Inc., Chap.1, sec. 5, pg. 29, 2002.

[6] The MathWorks, Inc.,"Image Processing Toolbox™ User's Guide" Version 8.1 (Release 2012b) page 1-17.

[7] Rafael C. Gonzalez & Richard E. Woods, "Digital Image Processing: Second Edition," Pearson Education, Inc., Chap.9, sec. 2, pg. 523, 2002.

[8] M. Keerthi Teja, "Distance Estimation of an object using image sequences for vehicle parking assistance", Dept. of Embd. Sys. Design, Int'l. Inst. Of Info. Tech., Pune, Chap. 4, Pg. 14-19, 34.