Scene Segmentation using Stereo Image with Color and Depth Information: A Retrospective

[1]Pratik Patel, [2]SandipPanchal
[1]PG Student, [2]Assistant Professor
[1,2]Department of EC, SVIT , Vasad , India
[1]pratikec39@gmail.com, [2]er sandip79@yahoo.com

***Abstract*— Scene segmentation is the well-known problem of identifying the different elements of a scene in image processing. Traditionally either color or depth information tracked from single scene view but it turns out to be poorly conditioned operation which remains very challenging. So fusing of color information and depth information is required for better scene distribution. It is very difficult to find depth information that allows recognizing the object which has same color with different depth, for more reliability purpose stereo image is used. This paper investigates stereo vision algorithms best suited to analysis depth information.**

Keywords— **Scene segmentation, Color information, Depth information,Stereo vision.**

I. INTRODUCTION

Scene segmentation is a crucial problem computer and image processing. Traditionally image tackled by exploiting only color information from single scene view [1]. Recent hardware and software developments allow to estimate in real-time scene geometry (depth) and open the way for new scene segmentation approaches based on the fusion of both color and depth data. Numerous applications can benefit from an effective scene segmentation method, e.g. 3D video, free viewpoint video, game controlling, forensic and in cryptography.

Image segmentation is the process of divide an image into multiplesets of pixelscalled segments. Purpose of segmentation is to simplify and change the representation of an image in more efficient and easier way to analyze. Scene segmentation is the widely known problem of identifying the different elements of a scene. This task is for identifying the image regions corresponding to the different scene elements or segments. Stereo vision systems provide estimates of depth of a framed scene from two or more views of it [1].

Many segmentation techniques have been developed, such as methods based on graph theory [2], other methods based on clustering algorithms [3].There are also some methods of region merging, watershed transforms, level sets and many other techniques [4]. The main drawback is that the information contains in image is not always sufficient to recognize the object.

II. NEED OF FUSION

Human eye can differentiate object and background with same color of scene, but same thing computer cannot do easily [5]. As shown in Fig. 1 back ground and hand of baby with same color. Here due to very similar color the part of baby skin and some map region associated with same segment (Fig. 2) by a classical segmentation algorithm based on color only information [5]. So the color information contain image is not always sufficient to recognize the object efficiently it can overcome with depth information [6].

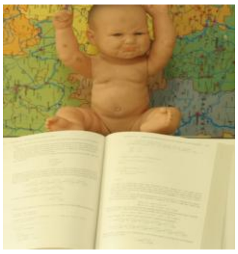


Fig. 1. Original image [5]

Fig. 2. Segmentation on the basis of color Data [5]

Fig. 3. Segmentation on the basis of depth Data [5]

Many different solutions have been proposed for the extraction of depth information relative to a real world scene each one with pros and cones. Now a day's market offers more expensive and precise active methods such as laser scanners, structured light systems, time-of-flight range cameras [7] (e.g. Mesa Imaging) and structured-light cameras (e.g. Microsoft Kinect [8]). Microsoft Photosynth [9] are gaining popularity. Some limits of color segmentation can be overcome but the difficulty is the regions that have similar depth but different colors. As shown in Fig. 3 the book and the baby feet were associated to the same segment due to their similar depth [5]. The use of depth information allows good segmentation performance but is not always effective. It can't solve situations where the objects of different colors placed close one to the other ach other [10].

III. DISPARITY BASED ON EPIPOLAR GEOMETRY

PatrikKamencay, et al [11] describes a simple way to find a depth from stereo image. Firstinput images are segmented and then the same matching points in the left view and right view images are found [12]. This idea is illustrated for an arbitrarily located 3D point P in Fig. 4. Consider a distant object is viewed by two cameras positioned in the same plane but separated by a distance 2l, known as the baseline. The object will appear in a similar position in both stereo images [11].

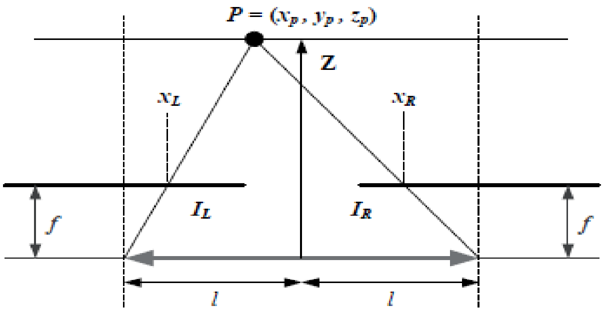


Figure. 4. epipolar geometry

$$d = x_l - x_R = f\left(\frac{x_p + l}{z_p} - \frac{x_p - l}{z_p}\right) \quad (1)$$

$$z_p = \frac{2fl}{d} = \frac{fb}{d} \quad (2)$$

d = disparity, x_L = horizontal co-ordinate left view of image, x_R = horizontal co-ordinate of Right view image, y_L,y_R =verticalco-ordinate of stereo image, f = focal length of camera, b=base line (2l).

IV. STEREO MATCHING ALGORITHM

Correlation based matching produces depth maps by calculating the disparity at each pixel with neighbor pixel [11]. This is achieved by taking a square window of certain size around the pixel of interest in the left view and finding the homologous pixel within the window in the right view. Some of stereo vision algorithms are as follows:

- Sum of Absolute Differences
- Sum of Squared Differences
- Normalized Cross Correlation

A. Sum of Absolute Differences

It computes the differences of intensity for each center pixel (i, j) in a window W as follows:

$$\sum_{i,j\in W} |(I_1(i,j) - I_2\,(X+i, Y+j)| \quad (3)$$

$I_1$and $I_2$are pixel intensity functions of the left and right image respectively. The disparity calculation is repeated within the x and y coordinates of frame image. The minimum difference value of the frame indicates the best matching pixel and position of the optimum defines the disparity of the actual pixel. After calculating aggregation of absolute differences within square, optimization is done with the winner-take-all (WTA) strategy [11].

Disparity map depends on square window size because wide window size corresponds to a greater probability of all correct pixel disparity calculated from matched points, although the calculation gets slower [13].

B. Sum of Squared Differences

The SSD algorithm is similar to the previously described SAD algorithm but SSD algorithm involves higher computational complexity and numerous multiplication operations. Instead of computing the absolute value this SSD algorithm computes squares of the intensity differences as follows [11]:

$$\sum_{i,j\epsilon W} \left((I_1(X,Y) - I_2(X+i,Y+j) \right)^2 \tag{4}$$

After calculating aggregation of square differences optimization is done with the winner-take-all (WTA) strategy [13].

C. Normalized Cross Correlation

Normalized Cross Correlation is complex then SAD and SSD algorithms. It involves numerous multiplication, division and square root operations [13].

$$\frac{\sum_{i,j\epsilon W}(\, (I_1(X,Y), I_2(X+i,Y+j))}{2\sqrt{\sum_{i,j\epsilon W}(I_1^2(X,Y)I_2^2(X+i,Y+j))}} \tag{5}$$

After calculating aggregation of Normalized Cross Correlation optimization is done with the winner-take-all (WTA) strategy [13].

Table I and Table II shows a SAD and SSD algorithm results of Cones and Tsukuba for window3, window5, window9. As shown in Table I and Table II with change in the size of window disparity map also change.

V. RESULTS OF SAD AND SSD ALGORITHM

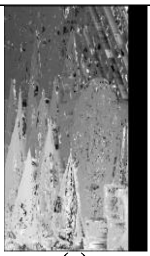
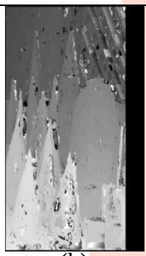
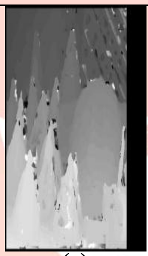
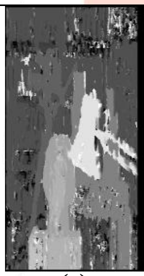
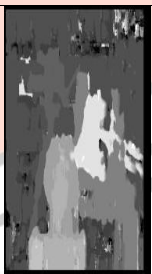
Image No.	Window 3	Window 5	Window 9
Cones	 (a)	 (b)	 (c)
Tsukuba	 (d)	 (e)	 (f)

TABLE I

SAD disparity map

(a) Window 3 for Cones (b) Window 5 for Cones (c) Window 9 for cones (d) Window 3 for Tsukuba (e) Window 5 for Tsukuba (f) Window 9 for Tsukuba

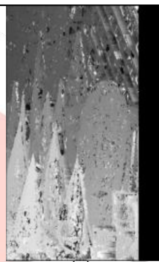
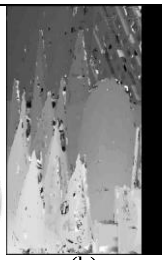
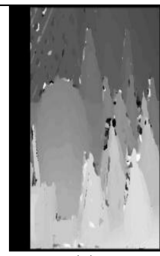
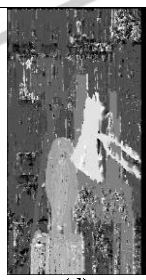
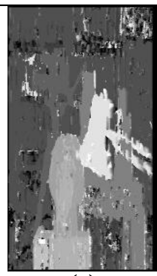
Image No.	Window 3	Window 5	Window 9
Cones	 (a)	 (b)	 (c)
Tsukuba	 (d)	 (e)	 (f)

TABLE II

SSD disparity map

(a)Window 3 for Cones (b) Window 5 for Cones (c) Window 9 for cones (d) Window 3 for Tsukuba (e) Window 5 for Tsukuba (f) Window 9 for Tsukuba

VI. REFERENCES

[1] Carlo Dal Mutto, Student Member,PietroZanuttigh, and Guido M. Cortelazzo, "Fusion of Geometry and Color Information for Scene Segmentation," IEEE Journal of Selected Topics In Signal Processing, Vol. 6, No. 5, September 2012

[2] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient graph-based image segmentation," International Journals of Computer Vision, 59(2):167–181, 2004.

[3] R. Szeliski, "Computer Vision Algorithms and Applications," New York, Springer, 2010.

[4] ChinkiChandhok, "Color Image Segmentation Using K-Means Clustering,"International Journal of VLSI & Signal Processing Applications, Vol 2, Issue 3, ISSN 2231-3133 , (241-245), June 2012.

[5] Carlo Dal Mutto, Fabio Dominio, PietroZanuttigh and Stefano Mattoccia, Stereo Vision and Scene Segmentation, 2012.

[6] Anil Z Chitade, "Colour Based Image Segmentation Using K-Means Clustering," International Journal of Engineering Science and Technology Vol. 2(10), 5319-5325, year 2010

[7] Mesa Imaging, [Online]. Available:http://www.mesa-imaging.ch

[8] MicrosftKinect, [Online]. Available: http://www.xbox.com/en-US/kinect

[9] Microsoft Photosynth, [Online]. http://photosynth.net/

[10] Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, "Scene segmentation assisted by stereo vision," IEEE computer society in Proc. 3DIMPVT '11, Hangzhou, China,May 2011.of Magdeburg Germany 2002.

[11] P. Kamencay, M. Breznan, R. Jarina, P. Lukac, M. Zachariasova, "Improved Depth Map Estimation from Stereo Images Based on Hybrid Method," radio egineering, vol. 21, no. 1, april 2012

[12] Diana Eva Bedolla Orozco, "Using Epipolar Geometry in Stereo Image Analysis," International Center of Theoretical Physics and ScuolaInternazionaleSuperiore di StudiAvanzati Trieste, Italy, September 2005.

[13] M. Tornow, M. Grasshoff, N. Nguyen, A. Al-Hamadi and B. Michaelis, "Fast Computation of Dense and Reliable Depth Maps from Stereo Images," Otto-von-Guericke University.