Depth Estimation and View Synthesis using Vanishing Point from Single View Image

Eu-Tteum Baek, Yo-sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro (Oryong-dong), Buk-gu
Gwangju 500-712, Republic of Korea
Email:{eutteum, hoyo}@gist.ac.kr

***Abstract*— Depth estimation from a single view image is probably the most difficult and challenging task. An algorithm performance for finding the accurate vanishing point in an image is affected by a different kind of image. In this paper. we proposed a depth map estimation from a single outdoor image and synthesis using the depth map which is estimated. First, we remove noise and find lines associated with vanishing point. Second, we segment an image into the sky and earth, then we extract depth map. Finally, we synthesize image using depth map estimated. Experimental results show that the proposed method provides accurate vanishing point, a depth map and a view synthesize which take into account a distance and a size of an object.**

Keywords—Vanishing point; Depth estimation; Hough transform; Canny edge detection; Image synthesis.

I. INTRODUCTION

Recently, depth estimation using stereoscopic images has been common. Previous algorithms from stereoscopic images provide robust and accurate depth map. Depth cameras such as a TOF camera and a kinect camera have become widely used in many applications. Those cameras have low cost real-time depth sensor. However, it is a hard work to estimate depth map from a single view image, and estimating a depth map has geometric constraints and its limitations. When a depth map is estimated with a single view depth map, a vanishing point is used. In order to detect a vanishing point, we extract lines from an image. If there are a few lines or no ideal lines in the image, it is difficult to find a vanishing point. The vanishing point is a point in the image plane that is the intersection of the projections of a set of parallel line in space in graphical perspective. Previous approaches of vanishing point detection rely on line segments detected in the image. A different approach is to utilize the intensity gradients of the pixel [1]. Barnard proposed the Gaussian sphere centered on the optical center of the camera as an accumulator space [2]. Van den Heuvel developed a method for detecting the three mutually orthogonal directions in the scene. The orthogonality criterion was explicitly used, which means that all combinations of three possible vanishing points have to be considered [3].

The single view image has no depth information, so we cannot apply 3D warping and synthesis using depth information. Several methods have been proposed to estimate the depth information from a single view. S. Battiato et al. utilize the color information to estimate the relative depth map in a 2D image [4]. S.A. Valencia *et al.* propose that measurement of the degree of focus, which is a spatial frequency measurement, can be achieved by a block-based wavelet transform, and then estimated the depth map by edge focus analysis [5].

In this paper, we proposed depth map estimation from a single outdoor image and synthesis using the depth map which is estimated. The outline of the algorithm is as follows: First, we extract accurate lines from an edge image, which is applied by canny edge detection. Second, we detect the vanishing point using an accumulation intersecting points. Third, we estimate depth map by using vanishing point and k-means segmentation. Finally, we synthesize an object and the image with the estimated depth map.

The rest of the paper is arranged as follows. Section 2 described system setup and system overview, and Section 3 described our algorithm in detail. In Section 4, experiment results and discussion is presented. Finally, the conclusion is described in Section 5.

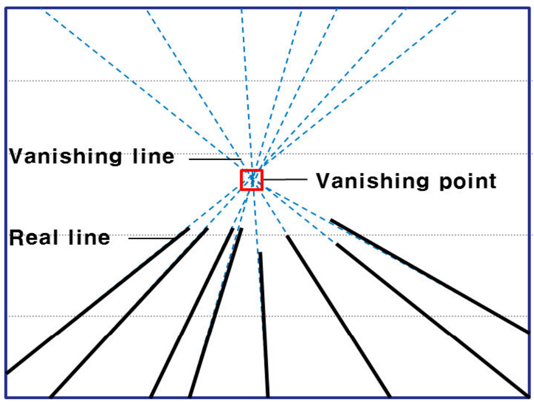

Fig. 1. Vanishing point

II. VANISHING POINT ESTIMATION

A. Preprocessing for extracting vanishing lines

In the single view image, distance information is very important to estimate depth. Figure 1 illustrates the vanishing point, real lines and vanishing lines. Real lines are extended infinitely. The intersection of extended line is a vanishing point. However, every line in the image does not intersect at a same point. Even though the line relates to the vanishing point, the line cannot intersect with the vanishing point in the image. Therefore, it is very important to find the accurate vanishing lines which intersect with vanishing point and to remove the noise and some lines which do not relate to vanishing point.

In order to extract edges, we apply canny edge detection [6]. In order to handle this problem, we used Erosion and Dilation. It is important to follow the order of the algorithm. If Erosion and Dilation are applied in the wrong order, incorrect results will come out.

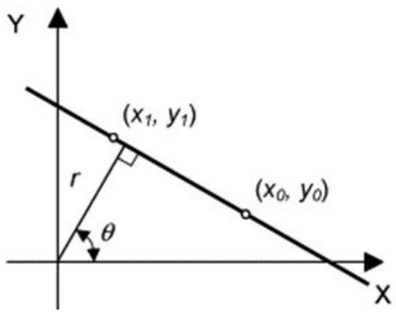

Fig. 2. The polar representation of straight line

B. Extracting lines

We use the edge image which noise is removed to apply Hough transform. The Hough transform is a feature extraction technique used in image analysis and digital image processing, and it is concerned with the identification of straight lines in the image. For each pixel at (x, y) and its neighborhood, the Hough transform algorithm determines if there is enough evidence of a line at that pixel (x, y). The Hough transform algorithm uses the polar representation of the straight line as Eq (1).

$$r = x\cos\theta + y\sin\theta \qquad (1)$$

r is the perpendicular distance from the line to the origin, and θ is the angle this perpendicular makes with the x axis. We now consider x and y as parameters r and θ as variables. In the polar coordinate(r, θ), we find the intersection points of curves and construct an accumulation array of the intersection points. The way of finding lines is by applying some form of threshold, which element of the accumulation array is more than the number that user set. However, because Hough transform algorithm simply detects arrayed pixels, it tends to find wrong edges or to extract edges which do not relate to the vanishing lines. Therefore, we are concerned with the minimum length of the line, the gap of lines which are the same r and θ, and the minimum element of the accumulation array. Figure 2 illustrates the polar representation of straight line.

```
P[x][y]; // set the entire array to 0
n // the number of equations

for i=0 to n-1;
  for j=i+1 to n;
    if ( an infinite number of solutions)
    then break;
    if ( no solution)
    then break;
    if  (one solution)
    then (find the solution (x, y) for the i-th and j-th
         equations)
         P[x][y]++;
```

Fig. 3. Algorithm for solution accumulation

C. Vanishing point estimation

We estimate the vanishing point by using lines from Hough transform. The extracted lines can be represented as Eq (2).

$$\begin{cases} r_0 = x\cos\theta_0 + y\sin\theta_0 \\ \vdots \\ r_{n-1} = x\cos\theta_{n-1} + y\sin\theta_{n-1} \\ r_n = x\cos\theta_n + y\sin\theta_n \end{cases} \qquad (2)$$

The equation (2) consists of two parameters r, θ and two variables x, y. To find a solution for a system of equations, we must find a value that is true for all equations in the system and construct an accumulation array of the intersection points. A system of equations has one solution, no solution, or infinite solutions. We do not accumulate the solution when the system of equations has no solution or infinite solutions. Figure 3 shows the algorithm for accumulating solutions. After accumulating solutions in the array P[x][y], we estimate the vanishing point using the distribution of the intersection points. However, in the image every vanishing point does not intersect each other. Because of this reason, we assume that there is the vanishing point where the accumulation values are high. In order to find the vanishing point, we calculate the Euclidean distance among candidates, and find the minimum sum of the distances. Equation 3 is the equation for Euclidean distance from (x_0, y_0) to (x_k, y_k).

$$W(P_0, P_k) = \sqrt{(x_0 - x_k)^2 + (y_0 - y_k)^2} \quad (3)$$

Equation 4 is the equation for the sum of Euclidean distance. n is the number of intersection points, p is an intersection point, and q is an the rest of the p.

$$\sum_{k=q}^{n} W(P(x_p, y_p), P(x_k, y_k)) \quad (4)$$

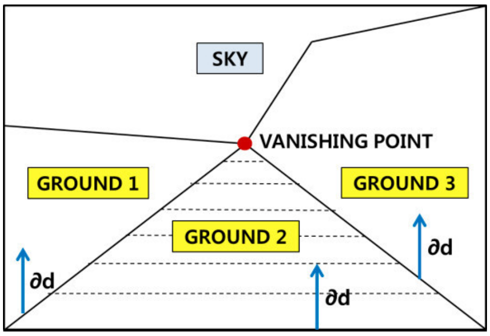

Fig. 4. Segmentation for depth value assignment

III. DEPTH MAP ESTIMATION AND SYNTHESISE

A. Depth map estimation

In a common picture in outdoor environments, the upside region of the image contain the sky, the bottom region of the image contains the ground. We assume that the distance between a camera and the sky is infinite, and the distance between the camera and the ground is close. We use the assumption and vanishing point to estimate the depth of the background. In order to estimate depth information, we use k-means segmentation to segment into the region of the sky and the region of the ground as Fig. 4. We fill the region of the sky with value 0 in grey scale. The ground is separated with three parts, and assigns the depth values. Each separated region of ground is assigned difference depth value sequentially. First, we use Eq (5) to assign the depth at the region of ground 2. y_{max} is a number of the height of the image, y_{van} is a number of y at the vanishing point, and y_0 is a current point of y.

$$I(x, y_n) = \frac{255}{y_{max} - y_{van}} * (y_n - y_{van}) \quad (5)$$

At the region of ground 1 and 3, from the edge between the corner of the bottom and the vanishing point to the edge between the vanishing point and upside of the ground we assign the depth.

B. Image synthesis

Typically, when an image is synthesized, the distance is not concerned. The result of synthesis which does not concern with the distance is not natural. It is important to concern with proportion of the height and the distance. We assume that the size of the object is the size of the object is closest to the bottom of the background. We adjust the size of the object using the proportion. Figure 5 shows the proportion according to the height of the object and the distance. Y is the distance from vanishing point to the bottom of the image, H is the height of the image, Y' is the point to top of the object, and H is the size of the synthesized object. We use H : Y = H' : Y' to calculate H'.

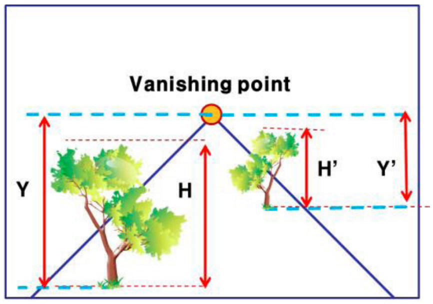

Fig. 5. Proportion according to the height of the object and the distance

IV. EXPERIMENTAL RESULTS

Fig 6(a) is the original image and Fig 6(b) shows the result of line extraction. The figure 6(b) and 6(c) include comparison result of proposed and Kong's algorithm [8]. The size of the image is 480*352. We use the edge image to segment into the lines and noises, and apply Hough transform to extract the strong lines. In order to extract strong lines, we set the minimum length. The strong line must have more than 40 pixels. The minimum length depends on the size of the image. Fig 6(c) shows the result of vanishing point estimation. In the proposed method, the vanishing point is (189, 167), the vanishing point of the Kong's method is (196, 175), and the vanishing point of manual selection is (192, 166). According to the result, we know that the proposed method more accurate than Kong's algorithm.

(a) Original image

(b) Line extraction: the left image using the Kong's algorithm, the right image using the proposed algorithm

(c) Vanishing point estimation: the left image using the Kong's algorithm, the right image using the proposed algorithm

Fig. 6. The result of vanishing point estimation

Figure 7 shows the result of the depth map estimation. We use the vanishing point and the segmented image to estimate the depth map. Figure 8 shows the result of the synthesis.

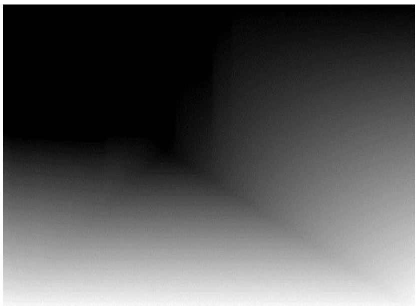

Fig. 7. The result of the depth map estimation

V. CONCLUSIONS

If we use a depth camera or a stereo camera, we can take an accurate depth map. However, there are many single view cameras, so the approach for depth map estimation from single view is important. In the paper we proposed depth map estimation from single view using vanishing point and synthesis using the depth map. We extracted strong lines, and estimate vanishing point. Using a depth map, we synthesized the image and the objects. The experiment results show that the vanishing point is almost accurate. However, the proposed depth map estimation has a weakness. Further research about the depth of the objects is needed.

Fig. 8. The result of the synthesis

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2013-067321)).

REFERENCES

[1] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. *The cascaded Hough transform as an aid in aerial image interpretation.* In Proc. ICCV, pp. 67–72, January 1998.

[2] S. T. Barnard. *Interpreting perspective images*, Artificial Intelligence 21, pp. 435-462, 1983.

[3] F. A. van den Heuvel. *Vanishing point detection for architectural photogrammetry*. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII part 5, pp. 652-659, 1998.

[4] S. Battiato, S. Curti, E. Scordato, M. Tortora, and M. La Cascia, *Depth map generation by image classification*, Three-Dimensional Image Capture and Applications VI, Vol. 5302, pp. 95-104, 2004.

[5] S. A. Valencia and R. M. Rodriguez-Dagnino, *Synthesizing Stereo 3D Views from Focus Cues in Monoscopic 2D Images*, in Proc of the SPIE, Vol. 5006, pp.377-388, 2003

[6] J. Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, Issue 6, pp. 679-698, Nov 1896.

[7] B. K. P. Horn. *Robot Vision*, pp. 66-69, 299-333. The MIT Press, 1986.

[8] H. Kong, J.-Y. Audibert, and J. Ponce, *Vanishing point detection for road detection,* in Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 96–103, 2009.