GENERATION OF IMAGES FROM MONOCULAR IMAGE FOR STEREOVISION

Priyanka K.B[1], Deepu R[2], Honnaraju B[3]
[1]M.Tech, [2]Asso Prof & HOD, [3]Assi Prof
Department of computer Science & Engineering
Maharaja Institute of Technology Mysore, India

Abstract: Stereovision is the process of finding depth when two or more images of a scene are known from different positions. It is the basic phenomena in the construction of 3D from multiple images. Stereo vision finds many applications in automated systems such as robotics, tracking object in 3D space and constructing a 3D spatial model of a scene. Most of the cameras are monocular in nature .One has to take multiple images from same camera from different viewpoints. In our work, we have developed a generic model wherein, second image is automatically generated and can be used for 3D reconstruction from a single image. Relationship between depth and disparity for any focal length has been established. Experiments are conducted to validate the proposed model and the results are compared to the conventional approach to confirm its accuracy and effectiveness. The model has been tested with middlebury stereo dataset. With this model, the existing monocular cameras are sufficient to build 3D view of any scene using a single 2D image.

I. INTRODUCTION

Stereopsis is the impression of depth that is perceived when a scene is viewed with both eyes by someone with normal binocular vision. Binocular viewing of a scene creates two slightly different images of the scene in the two eyes due to the eyes' different positions on the head. These differences, referred to as binocular disparity, provide information to the brain can be used to calculate depth in the visual scene. The impression of depth associated with stereopsis can also be obtained under other conditions, such as when an observer views a scene with only one eye while moving. Observer movement creates differences in the single retinal image over time similar to binocular disparity; this is referred to as motion parallax. This is despite the fact that in all these cases humans can still perceive depth relation because of experience. In stereo camera, if two calibrated cameras observe the same scene point p (refer to figure 1), its 3D coordinates can be computed as the intersection of two such rays. This is the basic principle of stereo vision that typically consists of three steps:

- Camera calibration.
- Establishing point correspondences between pairs of points from the left and the right images.
- Reconstruction of 3D coordinates of the points in the scene.

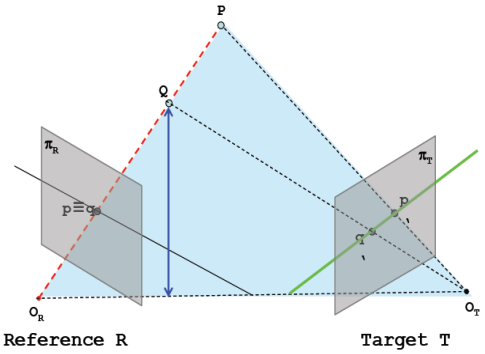

Figure 1: Stereo Camera

If the two cameras are calibrated such that they will be perfectly aligned and with the same focal length, then the depth can be easily calculated as shown in the equations below (refer to figure 2):

By considering similar triangles ($P^{O_R O_T}$ and P_{pp}'):

$$\frac{B}{Z} = \frac{(B+x_\tau)-x_R}{Z-f} \quad (1)$$

$$B*(z-f) = Z*(B+x_\tau - x_R) \quad (2)$$

$$B*f = Z*(x_R - x_\tau) \quad (3)$$

$$Z = \frac{B*f}{x_R - x_\tau} \quad (4)$$

Let $d = x_R - x_\tau$ is the disparity And B * f = Constant for the pair of cameras

Then

$$Z = \frac{Constant}{d} \quad (5)$$

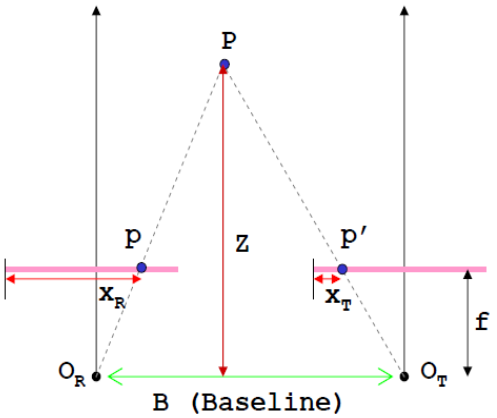

Figure 2: Aligned Stereo Cameras with the same focal length

Since the disparity is inversely proportional to the depth of the point (maximum disparity ≡ minimum depth), then if the point is near then its disparity is high and if the point is far then its disparity is low (refer to figure 3)

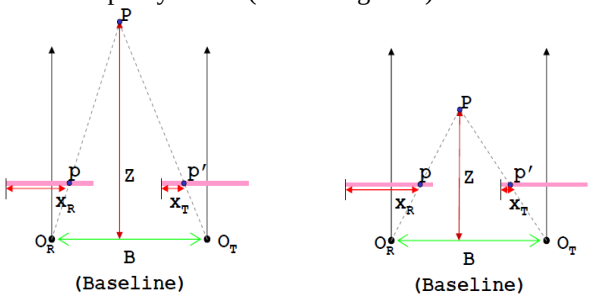

Figure 3: Disparity and Depth Relationship

As we have seen in figure 3 that the depth of a pixel in a reference image (left image) can be determined knowing its disparity from its corresponding pixel in a target image (right image), so in the next section we will introduce the methodology of calculation of the disparity of a given pixel in a reference image from its corresponding pixel a target image.

II. METHODOLOGY

Calibration approach uses vanishing points present in perspective view of an object in an image. Object used here is a cube placed at 45 degree to viewing plane. Algorithm is as follow.

Step 1 Bring the image co-ordinate system of indexing to Cartesian co-ordinate system of indexing for the image under consideration for camera calibration. For this, all pixels of the image represented by (r,c) of size (h, w) should have (h-r, c) in the new space.

Step 2 To each set of parallel edges, all the vanishing points are determined. We obtain two vanishing points as we have considered such images only. Name them with their co-ordinate values VHL(x1,y1) and VHR(x2,y2).

Step 3 A line is joined for these two vanishing points and is called 'Horizon Line'. A line is drawn parallel to the horizon line, and this is called as the 'Picture Plane'. This plane appears as a line because it is viewed from the top.

Step 4 VHL and VHR are projected onto the picture plane and are called VPL and VPR with co-ordinate values VPL (u1, v1) and VPR (u2, v2).

Step 5 The station point ST (xt, yt) is fixed up with equations

$$x_t = \frac{[(v_2 + m_1.u_1) - (v_1 + m_2.u_2)]}{(m_1 - m_2)} \quad (6)$$

$$y_t = m_1 \left\{ \frac{[(v_2 + m_1.u_1) - (v_1 + m_2.u_2)]}{(m_1 - m_2)} \right\} - m_1.u_1 + v_1 \quad (7)$$

Step 6 The focal length is the perpendicular dropped to the picture plane from the station point. This is calculated with equations

$$\left|\overrightarrow{PS_T}\right| = \sqrt{(x_t - X)^2 + (y_t - Y)^{2(8)}} \quad (8)$$

Step 7 The focal length line when extended on either ends, cuts the horizon line at the point 'CV(xv,yv)'. This point is the center of focus or the optical center of the camera. This is calculated with equations

$$x_v = \frac{[(y_t + m_h.x_1) - (y_1 + m_f.x_t)]}{(m_h - m_f)} \quad (9)$$

$$y_v = m_h.\left\{ \frac{[(y_t + m_h.x_1) - (y_1 + m_f.x_t)]}{(m_h - m_f)} \right\} - m_h.x_1 + y_1 \quad (10)$$

$$m_h = \left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (11)$$

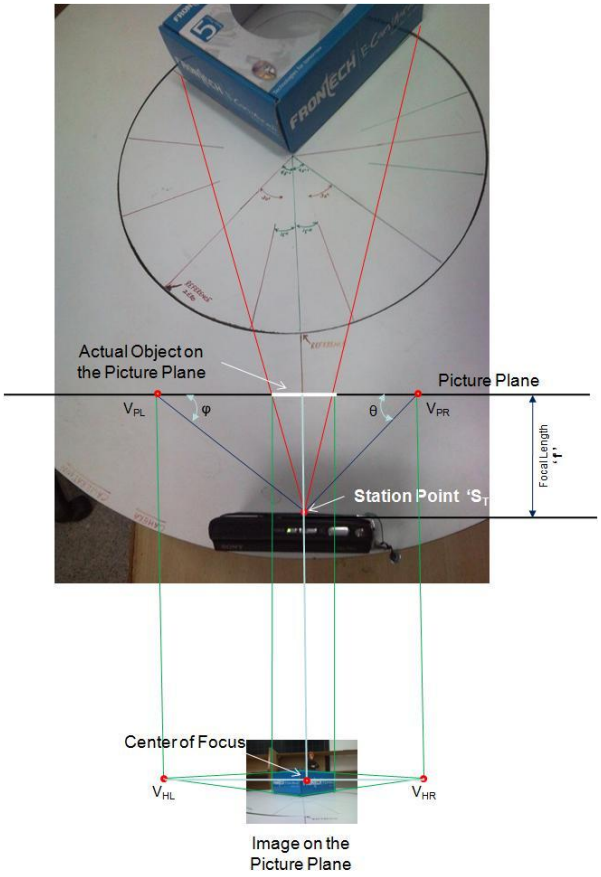

Figure 4: Construction of two point perspective: An image overview. (Construction is to demonstration purpose only and projections are not to scale)

The determined vanishing points are labeled 'V_{HL}' and 'V_{HR}' as shown in figure 4 .Thus the horizon line is drawn to join these two vanishing points.

The determination of station point is the crucial stage of this camera calibration algorithm. This is determined by following the procedure mentioned below:

- Drawing a parallel line to the horizon line called the picture plane.
- Projecting the 'VHL' and 'VHR' on to the picture plane to name them 'VPL' and 'VPR' respectively.
- With the known angles of the faces of the rectangular prism making with the picture plane, i.e. 'ABB1A1' making 'θ' with picture plane and 'ADD1A1' making 'φ' with picture plane, draw parallel lines from 'VPR' and 'VPL' with inclination of equal to 'φ' and 'θ' respectively and extend the line. Label the point of intersection of these two lines as 'ST'. This represents the station point.

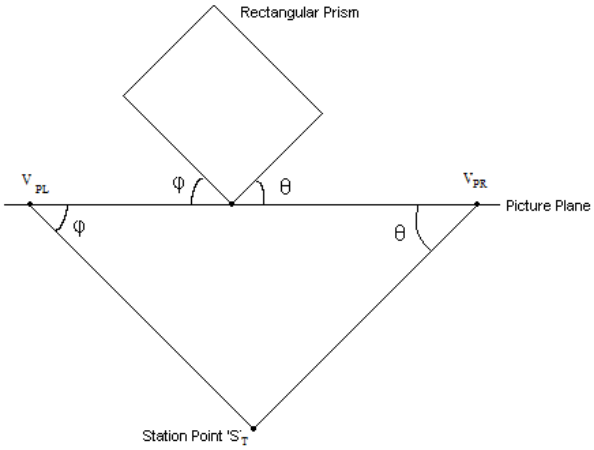

Figure 5: Depiction of Station point from the top view.

Station point is the eye / camera position point. Objects beyond the picture plane are oriented in some angles. I.e. the faces of the rectangular prism close to the picture plane makes angle 'θ' and 'φ' as shown in figure 5 and 6

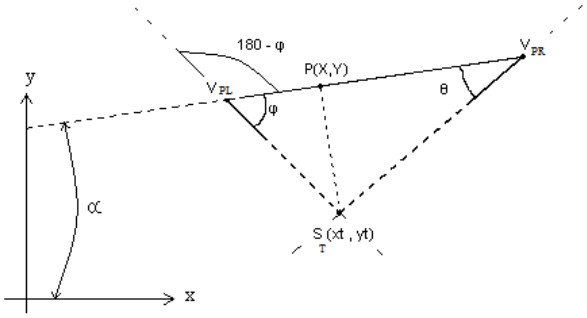

Figure 6: Orientation of the Picture plane with a possible inclination to the abscissa.

We used an Otsu algorithm which consists of eight main steps as follows:

- Read the input image I.
- Calculate the histogram for the input image I.
- Find the probability distribution of the image.
- Find the mean for the C1 and C2

$$\omega_0 = Pr(C_0) = \sum_{i=1}^{k} p_i = \omega(k)$$

$$\omega_1 = Pr(C_1) = \sum_{i=k+1}^{L} p_i = 1 - \omega(k)$$

- Find the zero-order and first order cumulative mean of the histogram up to kth level. Find the class variance.
- Measure the class separability, Find the within class variance between class variance and the total variance
- The optimal threshold is $\sigma_B^2(k\ ) = \max_{1 \le k < L} \sigma_B^2(K)$
- If k<256 go to step 5. Otherwise Exit.

A. Automatic selection of corresponding points and Disparity calculation

Sequence of images at different distances 'D' from the object for different focal lengths is taken. At every distance 'D', camera is moved along horizontal direction and we have used a triangle on board as shown in figure 4 and converted every image to greyscale. Later, they are scaled to 512 * 512 resolutions in order to standardise the values for fixed resolution.

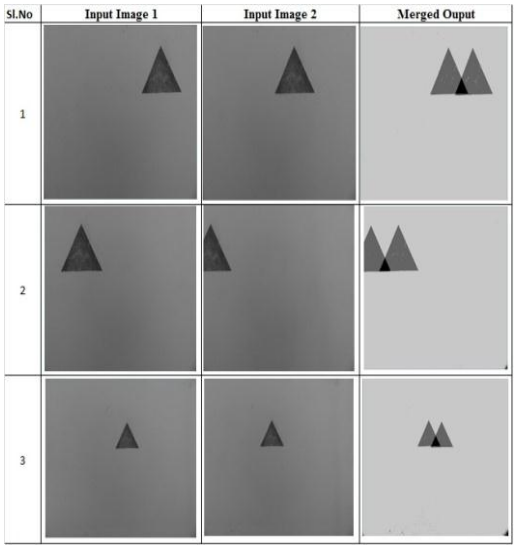

Figure 7: Data set for building model

Consecutive images on every row were selected for the purpose of identifying corresponding points. A specific point of an object is selected on left image and corresponding point on right images were initially obtained through manual selection on overlapped images as shown in figure 7. But, manual selection of corresponding points may leads to errors sometimes. Hence, automation of this was done where in every binary image is traced for specific corner point. This is carried for all the images in every row and average of difference in position of corner points in consecutive images is recorded. Later, the disparity is normalized for 512 X 512 images by dividing difference in pixel position by 512 and result is obtained as in Table 1. The entry for object which was occluded in the second image is dropped as shown in Table 1.

Depth	1MM	2MM	3MM	4MM	5MM
8	0.3457			0.4277	0.2324
12	0.2363	0.416	0.205	0.2714	0.1903
16	0.1746	0.3398	0.1542	0.1933	0.1479
20	0.1425	0.2792	0.125	0.1582	0.1196
24	0.121	0.2292	0.1093	0.1191	0.1019
28	0.1015	0.1914	0.0937	0.1074	0.0871
32	0.08	0.164	0.0839	0.0957	0.0797
36	0.0726	0.1562	0.0722	0.0839	0.071

Table 1: Depth versus disparity

B. Model Building

Though, there are many functions satisfying above values, exponential model gives best fit with 2 coefficients a and b as given in Figure 9. 'R' tool has been employed to construct the relationship between x='Depth' and y='Disparity' for various focal lengths. This model clearly shows the amount of distortion for different depths and when camera is moved away from the object, distortion decreases. The relationship is represented as

Distortion = a * e-b * Depth

R^2 value closer to 1 promise the accuracy in the obtained relationships. The above graph shows the depth versus disparity for different focal length by considering the distortion tabulated in table 1

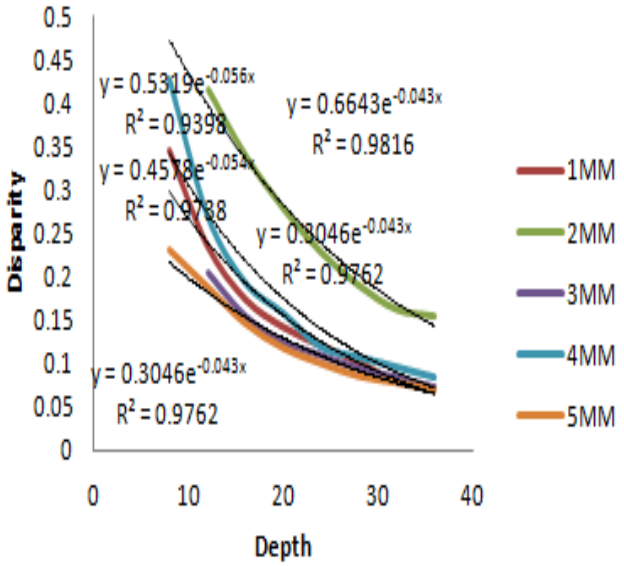

Figure 8: Depths versus Disparity

The relationship between each coefficient with focal length has been obtained in polynomial expression as shown in Figure 8 using R tool. The polynomial representation gives R^2 value as 1 which is not observed in other models. Hence, the following equations are considered as model equations for the generation of coefficients a and b for any focal lengths.

$$f = -0.0908a^4 + 1.0995a^3 - 4.6076a^2 + 7.6957a - 3.6389 \quad (12)$$

$$f = -0.0022b^4 + 0.0234b^3 - 0.0808b^2 + 0.1006b + 0.013 \quad (13)$$

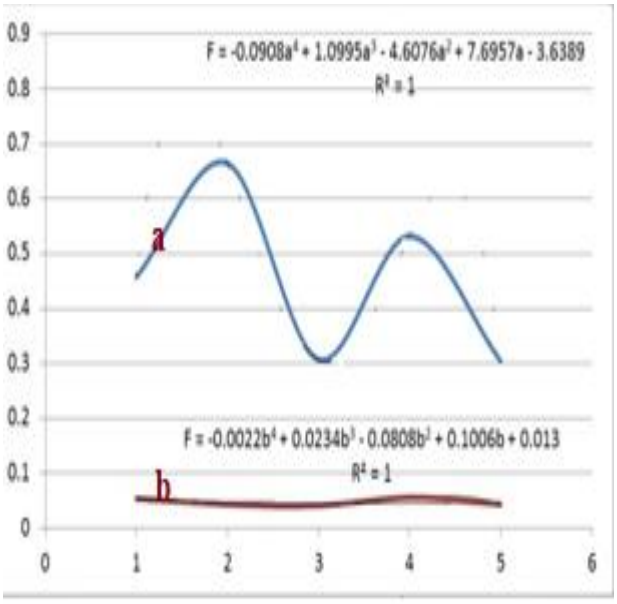

Figure 9: Functions relating Focal with Coefficients a and b

C. Generation of second Image for unknown images

During this phase, camera is calibrated using the zhang method discussed above and camera parameter were obtained. Once we calibrate the camera, we get focal length from which we are supposed to find coefficients 'a' and 'b' from above model. The image is automatically segmented using Otsu's algorithm, object is extracted for 512 X 512 images and its depth is obtained as per the method discussed in paper [1]. The depth which is in terms of pixels are converted to inches based on the conversion model constructed using the existing dataset. Once depth is known for this object, we get amount of distortion for different points on the selected object as per our exponential model given above and recorded in an array. Once we complete this estimation, we are supposed to move the pixels by distortion in order to get the second image. When model based results are analysed, the object movement was done with 95.3% accuracy on different test images. Hence, our model gives accurate second image from single image by providing the camera movement information along horizontal direction.

D. Other methods

Not all binocular stereo correspondence algorithms can be described in terms of our basic local algorithm. Here we briefly mention some additional algorithms that are not covered by our paper. A univalued representation of the disparity map is not essential. Multi-valued representations, which can represent several depth values along each line of sight, have been extensively studied recently, especially for large multi-view data set. Another way to represent a scene with more complexity is to use multiple layers, each of which can be represented by a plane plus residual parallax. Finally, deformable surfaces of various kinds have also been used to perform 3D shape reconstruction from multiple images.

III. EXPERIMENTAL RESULTS

In this section, we describe the experiments used to evaluate the stereo algorithms. Using the implementation framework we have found, we use the middlebury dataset for verification of the result

A. Resultant Stereo Images

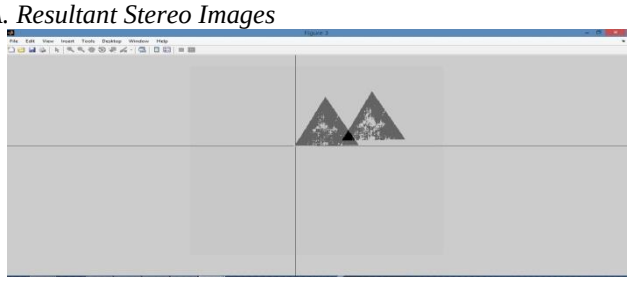

Figure 10: Stereo Images

The point is selected manually to find the disparity from left image to right image show in figure 18.

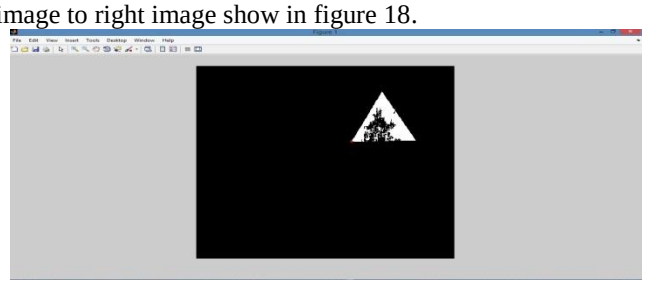

Figure 11: Automatic selection of any corner point in an image to find the disparity

The selection of point in other image is show in figure 12

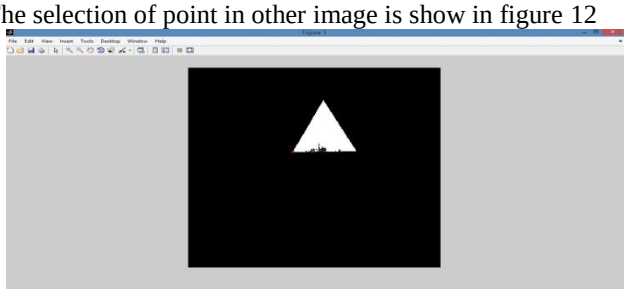

Figure 12: Automatic selection of any corner point in another image to find the disparity

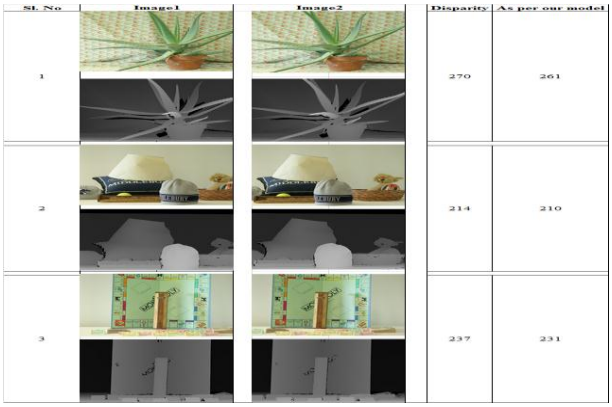

Figure 13: Verification of results for middlebury stereo dataset.

B. Application of our model

Our Model finds its application in any area where in 3D image becomes important. Our model gives an alternate to stereo cameras while generating 3D views. Using any mobile phone or portable device with single camera, we can generate 3D view equivalent to that of stereovision. It can be used to view 3D scenes in vehicle tracking, hump distance estimation using single image, 3D view from painting etc.

IV. CONCLUSION

In this paper we have introduced the methodology of camera calibration of novel approach, then we have used Otsu algorithm for segmentation. Over the last few decades, stereo and other triangulation cues have been successfully applied to many important problems, including robot navigation, building 3D models, and object recognition. But, they were confined to the use of at least two images. The model that we have developed is highly accurate in generating 3D views of any object from a single image as well as the entire image and literature shows that much attention was not given on this problem. The robust algorithms of automatic segmentation, camera calibration, and second image generation have promised accuracy in the results. The availability and low price of single camera makes this model an attractive work to develop more sophisticated applications in the field of vision system. We believe that the model finds its importance in many other applications of vision. Finally we would like to say that stereovision is one of the most active research areas in computer vision, due to its importance in real-time application, and the biggest challenge in this area of research from the past decade for its various practical applications in the field of machine vision and image analysis

REFERENCES

[1] Deepu R, Murali S, Vikram Raju, “A Mathematical model for the determination of distance of an object in a 2D image”, IPCV: vol.2, pp.585-592, 2013
[2] N. Otsu, “A threshold selection method from gray level histograms”, IEEE Trans. Syst. Man Cybern. SMC-9, 62-66 (1979).
[3] http://vision.middlebury.edu
[4] Georgios Vouzounaras , Petros Daras & Michael G. Strintzis, “Automatic generation of 3D outdoor and indoor building scenes from a single image”, Multimedia tools and applications, Springer Science + Business Media LLC , pp 1-18, 2011.
[5] Hansung Kim, Muhammad Sarim, Takeshi Takai. Dynamic 3D Scene Reconstruction in Outdoor Environments .IEEE
[6] Camera Center Estimation Using Vanishing Points, in proc. “IEEE International Conference on Signal and Image Processing”, Hubli, India, Dec. 2006.

