

Visualization 2D-to-3D Photo Rendering for 3D Displays

Sumit K Chauhan¹, Divyesh R Bajpai², Vatsal H Shah³

¹*Information Technology, Birla Vishvakarma mahavidhyalaya, sumitskc51@gmail.com*

²*Information Technology, Birla Vishvakarma mahavidhyalaya, divyeshbajpai@yahoo.com*

³*Assistant Professor, Birla Vishvakarma mahavidhyalaya, vatsal.shah@bvmengineering.com*

ABSTRACT- We describe a computationally fast and effective approach to 2D-3D conversion of an image pair for the three dimensional rendering on stereoscopic displays of scenes including a ground plane. The stereo disparities of all the other scene elements (background, foreground objects) are computed after statistical segmentation and geometric localization of the ground plane. The disparity maps generated with our approach are accurate enough to provide users with a stunning 3D impression.

I. INTRODUCTION

The recent advent of commercial 3D screens and visualization devices has renewed the interest in computer vision techniques for 2D-to-3D conversion. The appeal of 2D-to-3D conversion is two-fold. First, direct production of 3D media contents through specialised capturing equipment such as a stereoscopic video camera is still quite expensive. Second, a facility for converting monocular videos to stereo format would support the full 3D visualization of already existing contents, such as vintage movies.

A stereoscopic camera system consists of a pair of cameras producing a stereo (left and right) pair of images. Different disparities (i.e., shifts of corresponding scene points in the left and right visual channels) are interpreted by the human brain as corresponding variations of scene depth.

In this paper, we describe a practical and effective approach to 2D-3D conversion of an image pair, under the basic assumption that the scene contains a ground plane. Once such a plane is first segmented in the images by a statistical algorithm [4], the rest of the scene elements (background and foreground objects, the latter segmented in a semi-automatic way) can be acquired and rendered. In particular, the background is modelled as a vertical ruled surface following the ground boundary deformation, whereas foreground objects are flat, vertical layers, standing upon the ground plane.

The paper is organized as follows. The next section discusses all the theoretical aspects of the approach, from geometric modelling and estimation (subsect. A) to image segmentation and stereoscopic rendering (subsect. B). In third sec. experimental results are presented and discussed. Finally, conclusions and directions for future work are provided in fourth sec.

II. THE APPROACH

Given an uncalibrated monocular view I of a 3D scene, our goal is to synthesize the corresponding stereo pair (I_l, I_r) for a virtual stereoscopic system by exploiting a second view J of the same scene. The cameras corresponding to the actual views are placed in general position and are therefore not in a stereoscopic system configuration. I and J are referred to as reference and support images, respectively.

The role of I and J can be swapped, so each of them can be rendered on a 3D TV screen. By exploiting the support image, epipolar geometry estimation and camera self-calibration are first carried out. Automatic ground segmentation then allows recovering the homography induced by the ground plane on the two actual views.

A. Geometric Estimation

We now develop the theory related to warping the image of the ground plane onto the stereoscopic pair I_l and I_r . The theory being actually general, in the following we will refer to any given planar region π in the scene. Explicitly, the homography warping $I_l\pi$ onto the right view I_r is

$$H_r = K_i (I - sn^t / d_\pi) K_i^{-1}, \quad (1)$$

Where K_i is the calibration matrix for view I , and $s = [\delta/2 \ 0 \ 0]$, δ being the baseline between the virtual cameras. The homography H_l for the left view has the same form, but $s = [-\delta/2 \ 0 \ 0]$. The plane orientation n_π can be computed as

$$n_\pi = K_i^{-t} l_\pi, \quad (2)$$

Where l_π is the vanishing line of π in image I . The signed distance d_π can be obtained by triangulating any two corresponding points under the homography H_π and imposing the passage of π through the triangulated 3D point. The infinite homography $H_\infty = K_j R K_i^{-1}$. The homography

$$H_p = H_\pi^{-1} H_\infty \quad (3)$$

mapping I onto itself is actually a planar homology, i.e., a special planar transformation having a line of fixed points (the axis) and a distinct fixed point (the vertex), not on the line. Given the fundamental matrix F between two views, the three-parameter family of homographies induced by a world plane π is

$$H_\pi = A - j_e v^t, \quad (4)$$

Where $[j_e] \times A = F$ is any decomposition (up to scale) of the fundamental matrix, and j_e is the epipole of view I in image J (in other words, $j_e^t F = 0$). The matrix A can be chosen as

$$A = [j_e] \times F. \quad (5)$$

Both the fundamental matrix F and the ground plane homography H_π are robustly computed by running the RANSAC algorithm [10] on SIFT correspondences [5].

A.1. Camera Self-Calibration

Camera self-calibration follows the approach of [6], where the fundamental matrix F between I and J is exploited. In our notation, F is defined as

$$j_x^t F i_x = 0, \quad (6)$$

For any two corresponding points $i_x \in I$ and $j_x \in J$. In [6], the internal camera matrices K_i and K_j are estimated by forcing the matrix $E = K_j^{-t} F K_i$ to have the same properties of the essential matrix. This is achieved by minimizing the difference between the two non zero singular values of E , since they must be equal. The Levenberg-Marquardt algorithm is used, so an initial guess for K_i and K_j is required.



(a)

(b)

Figure 1. (a) Reference image I. (b) Support image J.

B. Stereo pair generation and rendering

This section specializes the use of Eq. 1 to the case of a scene including a planar ground, and then expounds how to warp the background and foreground objects properly, given the image of the ground plane. Figure 1 shows the images I and J that will be used to illustrate the various rendering phases.

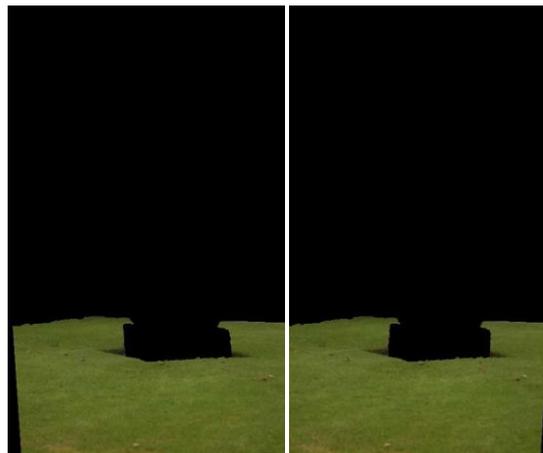
B.1. Ground Plane Virtual View Generation

The ground plane is segmented in the images I and J by exploiting the classification algorithm proposed in [4].



(a) (b)
Figure 2

(a) Ground classification for image I: The brighter the color, the more probable the ground region. (b) Recovery of the ground plane vanishing line (dashed line in the picture), after camera self-calibration and ground plane homography estimation.



(a) (b)
Figure 3. The two virtual views for the ground plane of image I. (a) Il. (b) Ir.

Fig. 2(a) shows the ground plane classification for the reference image of Figure 1(a). Figure 2(b) shows the computed vanishing line for the ground plane in the reference image I, after camera self-calibration and the computation of the ground plane homography $H\pi$ have been performed.



(a)

(b)

Figure 4

Background generation for Ir. (a) Top border of the background not occluded. (b) Recovery of the background for the occluded part of the ground top border. The resulting two virtual views (Il, Ir) of the ground plane are shown in Figure. 4.

B.2. Background Generation

Given the warped ground plane, the background of the scene is generated in a column-wise way. This is a direct consequence of modelling the background as a ruled surface perpendicular to the ground plane. For each point p of the top border of the ground in I, the corresponding point in Ir and Il is recovered. Figure 4(a) shows an example of occlusion by a foreground object (the statue). Figure 4(b) shows that background data have been correctly filled in.

B.3. Displacement Of The Foreground Objects

Foreground objects are segmented in a semi-automatic way with the GrabCut tool [7]. They are rendered in the images as flat and frontal objects, since the texture of the object is usually sufficient to provide the user with the impression of local depth variation due to the object's shape. Figure 5 shows the final stereo pair ((a) and (b)), the two images superimposed (c) and the disparity map (d).

B.4. Stereoscopic Display On A 3d Screen

For a parallel camera stereoscopic system, points at infinity have zero disparity, and appear to the user to be on the screen surface when the images are displayed on a 3D TV without modification. When a limited range $[-a, b]$ for disparity is introduced, the nearest and furthest points are associated with the extreme values of that range. Hence the zero disparity plane is not located anymore at infinity, but is frontal to the cameras, in a region between the nearest and furthest points.



(a)



(b)



(c)



(d)

Figure 5



(a)



(b)



(c)



(d)

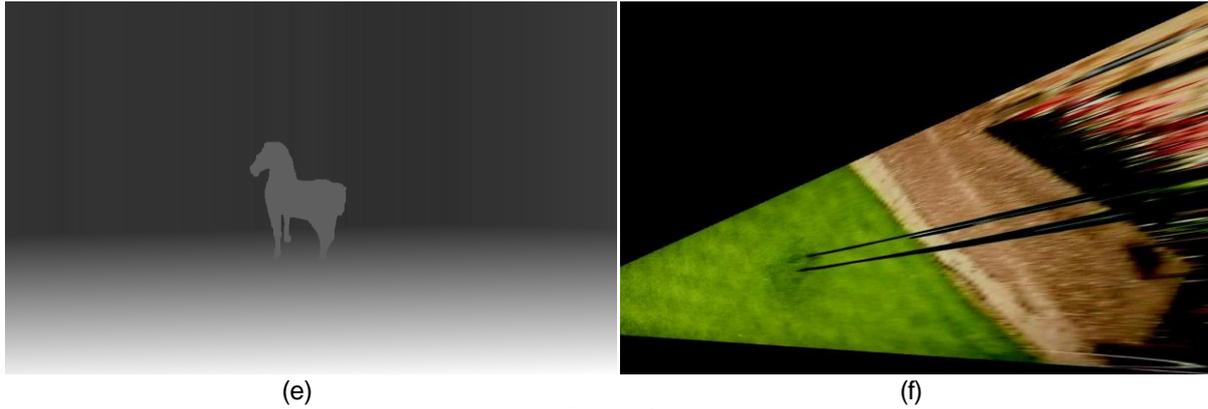
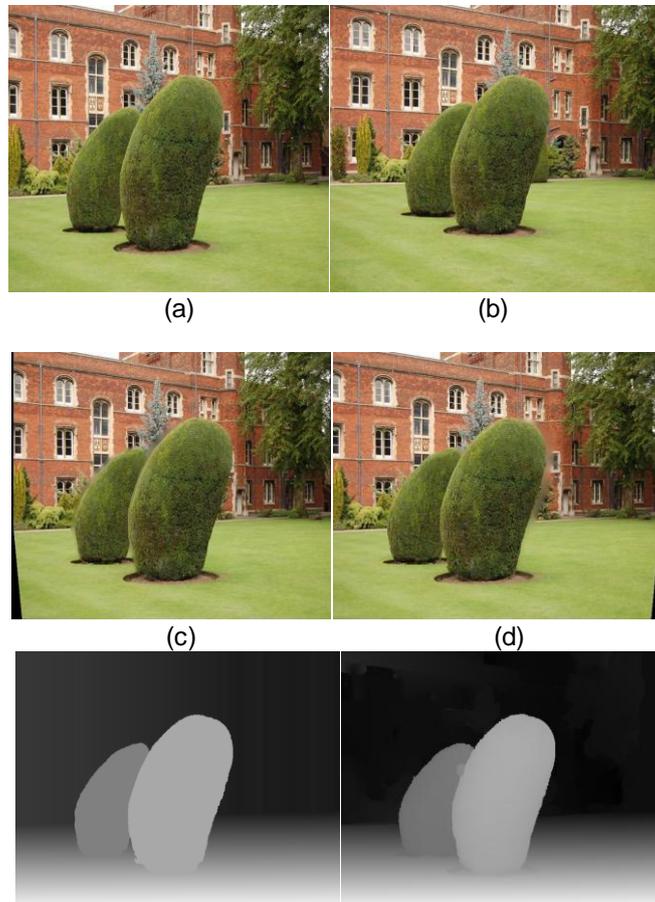


Figure 6

III. EXPERIMENTAL RESULTS

In figure. 6(a) and (b) are shown the reference and support images of the “horse” pair together with their associated ground plane vanishing lines. Figure 6(c) and (d) show the obtained stereoscopic pair, featuring coincident vanishing lines. Notice at the bottom left (c) and right (d) corners the black (i.e., empty) regions arising after warping the original images. Figure 6(e) shows the resulting disparity map.

Fig. 6(f) shows a front-to-parallel view of the ground plane. Figs. 7(a) and (b) illustrate the “bushes” pair, where two partially self-occluding foreground objects are present. Notice, from both Fig 7(c) and (d), the small blurred regions—especially evident to the left (c) and right (d) of the closer bush—due to color interpolation inside occluded background areas.



(e)

(f)

Figure 7

The “bushes” example. (a) Reference image I. (b) Support image J. (c) Left stereoscopic image Il. (d) Right stereoscopic image Ir. (e) Disparity map with our approach. (f) Disparity map with a dense stereo approach.

The good quality of disparity map obtained with our approach is confirmed by a visual comparison against the disparity map of Figure. 7(f), which was obtained with a state-of-the-art dense stereo approach [9]. As evident from the disparity map of Figure 7(e), the two bushes are correctly rendered as belonging to two distinct depth layers.

IV. CONCLUSIONS AND FUTURE WORK

We have described and discussed a simple yet fast and effective approach to 2D-3D conversion of an image pair for parallel stereoscopic displays, where the disparities of all scene elements are generated after statistical segmentation and geometric localization of the ground plane in the scene.

Future work will address (1) extending the approach to videos (which will lead to investigate the problem of temporal consistency among frames), (2) relaxing the ground plane assumption (3) implementing an automatic method to determine the optimal range of disparities for 3D perception.

V. REFERENCES

- [1] S. Coren, L. M. Ward, and J. T. Enns. *Sensation and Perception*. Harcourt Brace, 1993.
- [2] A. Criminisi, M. Kemp, and A. Zisserman. Bringing pictorial space to life: computer techniques for the analysis of paintings. In on-line Proc. Computers and the History of Art, 2002.
- [3] M. Guttman, L. Wolf, and D. Cohen-Or. Semi-automatic stereo extraction from video footage. In Proc. IEEE International Conference on Computer Vision, 2009.
- [4] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal on Computer Vision*, 75(1), 2007.
- [5] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2):91–110, 2004.
- [6] P. Mendonca and R. Cipolla. A simple technique for selfcalibration. In Proc. Conf. Computer Vision and Pattern Recognition, 1999.
- [7] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cut s. In *ACM Transactions on Graphics (SIGGRAPH)*, 2004.
- [8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [9] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(12):2115–2128, 2009.
- [10] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24(6):381–395, 1981.