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

Dario Comanducci
Dip. di Sistemi e Informatica, Univ. di Firenze
Via S.Marta 3, 50139 Firenze, Italy
comandu@dsi.unifi.it

Atsuto Maki
Toshiba Research Europe
Cambridge CB4 0GZ, UK
atsuto.maki@crl.toshiba.co.uk

Carlo Colombo
Dip. di Sistemi e Informatica, Univ. di Firenze
Via S.Marta 3, 50139 Firenze, Italy
colombo@dsi.unifi.it

Roberto Cipolla
Dept. of Engineering, Univ. of Cambridge
Cambridge CB2 1PZ, UK
cipolla@eng.cam.ac.uk

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. Geometric estimation includes camera self-calibration from epipolar geometry, and an original algorithm for the recovery of 3D visual parameters from the properties of planar homologies. Experimental results on real images show that, notwithstanding the simple "theatrical" model employed for the scene, the disparity maps generated with our approach are accurate enough to provide users with a stunning 3D impression of the displayed scene, and fast enough to be extended to video sequences.

1. 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. The simplest stereoscopic system consists of two cameras with parallel axes. Such a camera system produces images with only horizontal disparity, thus avoiding the vertical image disparity arising in stereoscopic systems that verge the camera axes. Studies about viewer comfort for stereoscopic displays agree about the fact that the amount of disparity should vary in a limited range [8]. This is because, for humans, eye convergence and accommodation (focusing) are tightly related. When we watch a 3D TV we are focusing the screen, and our eyes converge according to the distance from the screen. Hence, limiting disparities ensures that the viewer's perceived depth is controlled without stressing the convergence-accommodation bond. Points with zero disparity are located on the surface of the TV screen . Camera separation is indeed the most important parameter to produce realistic 3D images: as reported in [8], the length of the baseline between the two cameras is based on viewing arrangement, disparity range, scene depth and camera features. In practice, the proper value of camera separation is usually chosen manually.

Given a single stream of uncalibrated images as input, the goal of 2D-3D conversion is essentially to generate, *for each input image, a pair of stereo images* having disparities that directly reflect the actual depth of the scene. In order to deal with this highly ill-posed problem, it is a prerequisite to make certain assumptions [12], and therefore existing techniques either adapt models of scene geometry, or use manual procedures such as user-scribbles for guiding depth estimation (see [4] for a recent example). An alternative strategy for disparity map generation is to perform dense depth search [16]. Although such an approach appears most general and appropriate, finding dense correspondences can quite hard in the presence of textureless regions and/or occlusions. Even when a powerful bundle

optimization framework is employed [18], it is nevertheless difficult to obtain a clean segmentation between objects throughout an entire sequence, which typically results in a blurred perception of boundaries in the 3D scene. Another difficulty of dense stereo methods is that they are very time consuming, and therefore hardly usable for the stereoscopic rendering of long video sequences. To avoid the visual artifact due to the inaccurately recovered 3D information, in [17] the stereoscopic video frames are generated by selecting the most suitable frames within the input video. Stereoscopic effect, frame similarity and temporal smoothness are taken into account. This strategy is useful only in videos with a consistent panning motion.

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 [7, 14], 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 disparities of all scene elements are computed starting from the equation describing the actual position and orientation of the ground plane in the scene. To compute the ground plane equation, an original method based on scene homographies and homologies is employed, requiring as input only an estimate of the epipolar geometry of the two views. Experimental results on real images show that the disparity maps generated with the proposed method are effective in providing the users with a dramatic and vivid 3D impression of the displayed scene. This perceptual success is obtained notwithstanding the deliberate simplicity of the scene model, and is due in part to a proper rendering of texture as a dominant visual cue [1, 2]. Results also demonstrate that, for the purpose of successful 3D rendering and visualization, the correct ordering of layers in terms of their distance together with a neat segmentation of their boundaries is more important than a high accuracy of disparity estimates [9]. The overall process of stereoscopic rendering is fast enough to be fully automated and extended to 2D-3D video conversion.

The paper is organized as follows. The next section discusses all the theoretical aspects of the approach, from geometric modelling and estimation (subsect. 2.1) to image segmentation and stereoscopic rendering (subsect. 2.2). In sect. 3 experimental results are presented and discussed. Finally, conclusions and directions for future work are provided in sect. 4.

2. 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.

Fig. 1 provides a general overview of the approach. 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. By combining this homography with calibration data, the ground plane equation is estimated. Hence, ground plane and calibration data are exploited to compute the two homographies generating the stereoscopic images of the ground plane. After that, the rendered ground plane images are used to render the background and foreground images, are eventually all the rendered images are merged together to form the stereoscopic image pair to be displayed.

2.1. 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. The image $I_\pi \subset I$ of π can be warped onto I_l and I_r according to a pair of homographies $\mathtt{H}_l$ and $\mathtt{H}_r$ that depend on plane orientation $\mathbf{n}_\pi$ in space, signed distance d_π from the reference camera, and calibration data. Explicitly, the homography warping I_π onto the right view I_r is

$$\mathtt{H}_r = \mathtt{K}_i(\mathtt{I} - \mathbf{s}\mathbf{n}_\pi^\top / d_\pi)\mathtt{K}_i^{-1} \quad , \tag{1}$$

where $\mathtt{K}_i$ is the calibration matrix for view I, and $\mathbf{s} = [\delta/2\ 0\ 0]^\top$, δ being the baseline between the virtual cameras. The homography $\mathtt{H}_l$ for the left view has the same form, but $\mathbf{s} = [-\delta/2\ 0\ 0]^\top$. These formulas are the specialization of the general homography between two views of a plane for the case when the two cameras are only shifted of a quantity $\pm\delta/2$ along the horizontal camera axis.

We discuss hereafter the estimation of geometric entities related to the planar region π. Estimation of the epipolar geometry between the views I and J and camera self-calibration of both intrinsic and extrinsic parameters from the fundamental matrix $\mathtt{F}$ between views I and J will be addressed later on.

The plane orientation $\mathbf{n}_\pi$ can be computed as

$$\mathbf{n}_\pi = \mathtt{K}_i^{\top\, i}\mathbf{l}_\pi \quad , \tag{2}$$

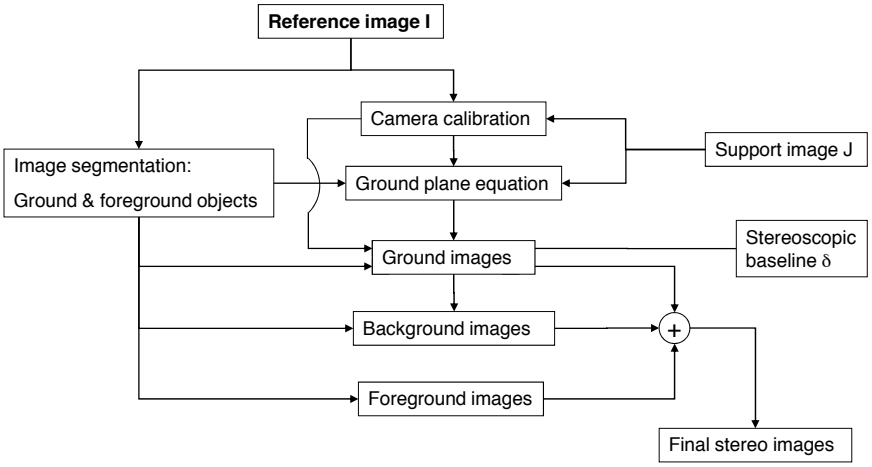


Figure 1. Overview of the approach.

where ${}^i\mathbf{l}_\pi$ 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 $\mathtt{H}_\pi$ (induced by π between I and J, and estimated as detailed in subsect. 2.2.1) and imposing the passage of π through the triangulated 3D point. The vanishing line ${}^i\mathbf{l}_\pi$ of the planar region π is composed of points that are mapped from I to J both by $\mathtt{H}_\pi$ and by the infinite homography $\mathtt{H}_\infty = \mathtt{K}_j\mathtt{R}\mathtt{K}_i^{-1}$. The homography

$$\mathtt{H}_p = \mathtt{H}_\pi^{-1}\mathtt{H}_\infty \tag{3}$$

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. In the case of $\mathtt{H}_p$, the vertex is the epipole ${}^i\mathbf{e}_j \in I$ of view J, and the axis is the vanishing line ${}^i\mathbf{l}_\pi$, since it is the intersection of π with the plane at infinity π_∞ [5]. Thus, thanks to the properties of homologies, ${}^i\mathbf{l}_\pi$ is obtained as ${}^i\mathbf{l}_\pi = \mathbf{w}_1 \times \mathbf{w}_2$, where $\mathbf{w}_1, \mathbf{w}_2$ are the two eigenvectors of $\mathtt{H}_p$ corresponding to the two equal eigenvalues.

In order to obtain robust warping results, it is required that the homography $\mathtt{H}_\pi$ be compatible with the fundamental matrix $\mathtt{F}$, i.e., $\mathtt{H}_\pi^\top\mathtt{F} + \mathtt{F}^\top\mathtt{H}_\pi = \mathtt{0}$. This is achieved by using a proper parametrization for $\mathtt{H}_\pi$ [5]. Given the fundamental matrix $\mathtt{F}$ between two views, the three-parameter family of homographies induced by a world plane π is

$$\mathtt{H}_\pi = \mathtt{A} - {}^j\mathbf{e}_i\mathbf{v}^\top \quad , \tag{4}$$

where $[{}^j\mathbf{e}_i]_\times\mathtt{A} = \mathtt{F}$ is any decomposition (up to scale) of the fundamental matrix, and ${}^j\mathbf{e}_i$ is the epipole of view I in image J (in other words, ${}^j\mathbf{e}_i^\top\mathtt{F} = \mathbf{0}^\top$). Since $[{}^j\mathbf{e}_i]_\times[{}^j\mathbf{e}_i]_\times\mathtt{F} = -\left\|{}^j\mathbf{e}_i\right\|^2\mathtt{F}$, the matrix $\mathtt{A}$ can be chosen as

$$\mathtt{A} = [{}^j\mathbf{e}_i]_\times\mathtt{F} \quad . \tag{5}$$

Both the fundamental matrix $\mathtt{F}$ and the ground plane homography $\mathtt{H}_\pi$ are robustly computed by running the RANSAC algorithm [3] on SIFT correspondences [10]. In particular, for the ground plane homography the parametrization of Eq. 4 is used, thus requiring only three point correspondences for its estimation.

2.1.1 Camera self-calibration

Camera self-calibration follows the approach of [11], where the fundamental matrix $\mathtt{F}$ between I and J is exploited. In our notation, $\mathtt{F}$ is defined as

$${}^j\mathbf{x}^\top\mathtt{F}\,{}^i\mathbf{x} = 0 \quad , \tag{6}$$

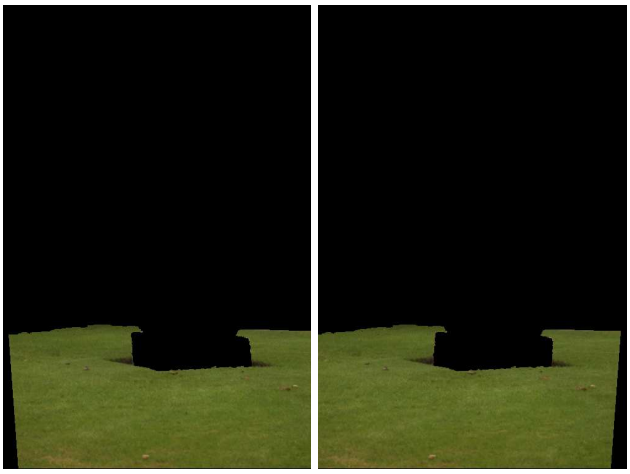
for any two corresponding points ${}^i\mathbf{x} \in I$ and ${}^j\mathbf{x} \in J$. In [11], the internal camera matrices $\mathtt{K}_i$ and $\mathtt{K}_j$ are estimated by forcing the matrix $\hat{\mathtt{E}} = \mathtt{K}_j^\top\mathtt{F}\mathtt{K}_i$ to have the same properties of the essential matrix. This is achieved by minimizing the difference between the two non zero singular values of $\hat{\mathtt{E}}$, since they must be equal. The Levenberg-Marquardt algorithm is used, so an initial guess for $\mathtt{K}_i$ and $\mathtt{K}_j$ is required. The most uncertain value among the entries of $\mathtt{K}_i$ and $\mathtt{K}_j$ is the focal length: as suggested in [6], this value is expected to fall in the interval $[1/3(w+h), 3(w+h)]$, where w and h are respectively the width and height of the image. In our approach, the first guess for the focal length is obtained with the method proposed in [15] if the solution falls in the above interval, otherwise it is set to $w + h$. The principal point is set in the center of the image, while pixels are assumed square (unit aspect ratio and zero skew). Extrinsic parameters (rotation matrix $\mathtt{R}$ and translation vector $\mathbf{t}$) of the support camera with respect to the reference camera are then recovered by factorizing the estimated essential matrix as $\hat{\mathtt{E}} = [\mathbf{t}]_\times\mathtt{R}$ [5].

(a) (b)

Figure 2. (a): Reference image I. (b): Support image J.

(a) (b)

Figure 3. Ground plane recovery. (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 4. The two virtual views for the ground plane of image I. (a): I_l. (b): I_r.

2.2. Stereo pair generation and rendering

So far, we have described how to compute the pair of homographies mapping the image of a generic planar region onto the two translated virtual views forming the stereoscopic pair (I_l,I_r). 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. Fig. 2 shows the images I and J that will be used to illustrate the various rendering phases.

2.2.1 Ground plane virtual view generation

The ground plane is segmented in the images I and J by exploiting the classification algorithm proposed in [7]. Fig. 3(a) shows the ground plane classification for the reference image of Fig. 2(a). Fig. 3(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 $\mathtt{H}_\pi$ have been performed. The resulting two virtual views (I_l, I_r) of the ground plane are shown in Fig. 5.

2.2.2 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 $\mathbf{p}$ of the top border of the ground in I, the corresponding point in I_r and I_l is recovered, and the whole column of pixels above $\mathbf{p}$ is copied in I_r and I_l starting from $\mathtt{H}_r\mathbf{p}$ and $\mathtt{H}_l\mathbf{p}$ respectively. When the top border of the ground is not visible because it is occluded by a foreground object, the corresponding image column cannot be copied as described before. Instead, the missing background part is recovered by linearly interpolating the corresponding background column indexes in I. In particular, the missing background pixel columns are obtained by uniformly sampling the reference image I in the range $[y_l, y_r]$, where y_l and y_r denote the borders of the missing background in image I. If there are several connected missing parts, the procedure must be repeated for each of them. Fig. 5(a) shows an example of occlusion by a foreground object (the statue). Fig. 5(b) shows that background data have been correctly filled in. When the foreground object does not occlude the top border of the ground, but it occludes some pixels of the corre-

sponding background column, the foreground pixels are not copied. The remaining background portions, i.e., those occluded by the foreground objects, are filled in with smooth color interpolation.

(a) (b)

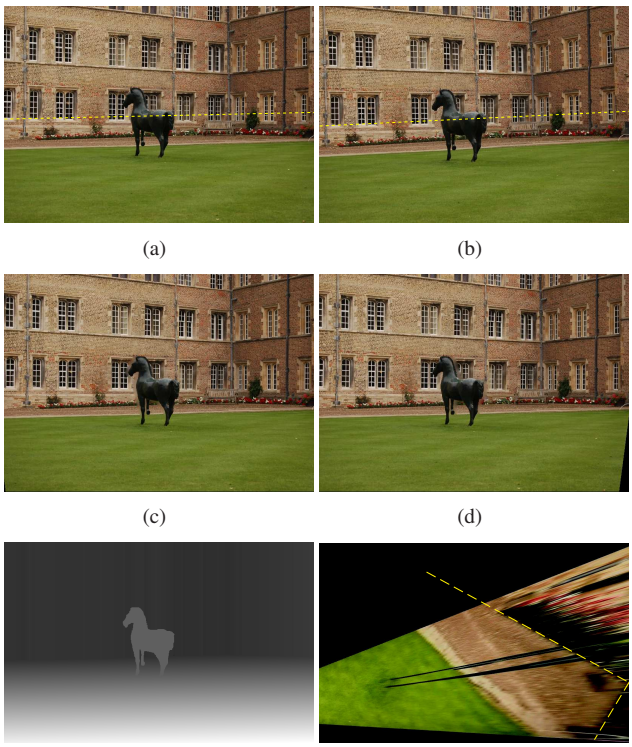
Figure 5. Background generation for I_r. (a): Top border of the background not occluded. (b): Recovery of the background for the occluded part of the ground top border.

2.2.3 Displacement of the foreground objects

Foreground objects are segmented in a semi-automatic way with the GrabCut tool [13]. 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. Depth is assigned as the value corresponding to the point of contact with the ground, considered to be the bottom point of their silhouette. Users are allowed to change the position of the point of contact by clicking on the desired point in the image. Fig. 6 shows the final stereo pair ((a) and (b)), the two images superimposed (c) and the disparity map (d).

2.2.4 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. Since the scene is in front of the camera, in our approach an overall translation is applied to the two images I_r and I_l in order to have zero disparity in the bottom line of the ground. Doing so, the user has the impression that the 3D image is inside the TV, starting from the screen surface. Users are nonetheless free to change the overall shift and put on the screen surface other frontal regions, if required.



(a) (b)

(c) (d)

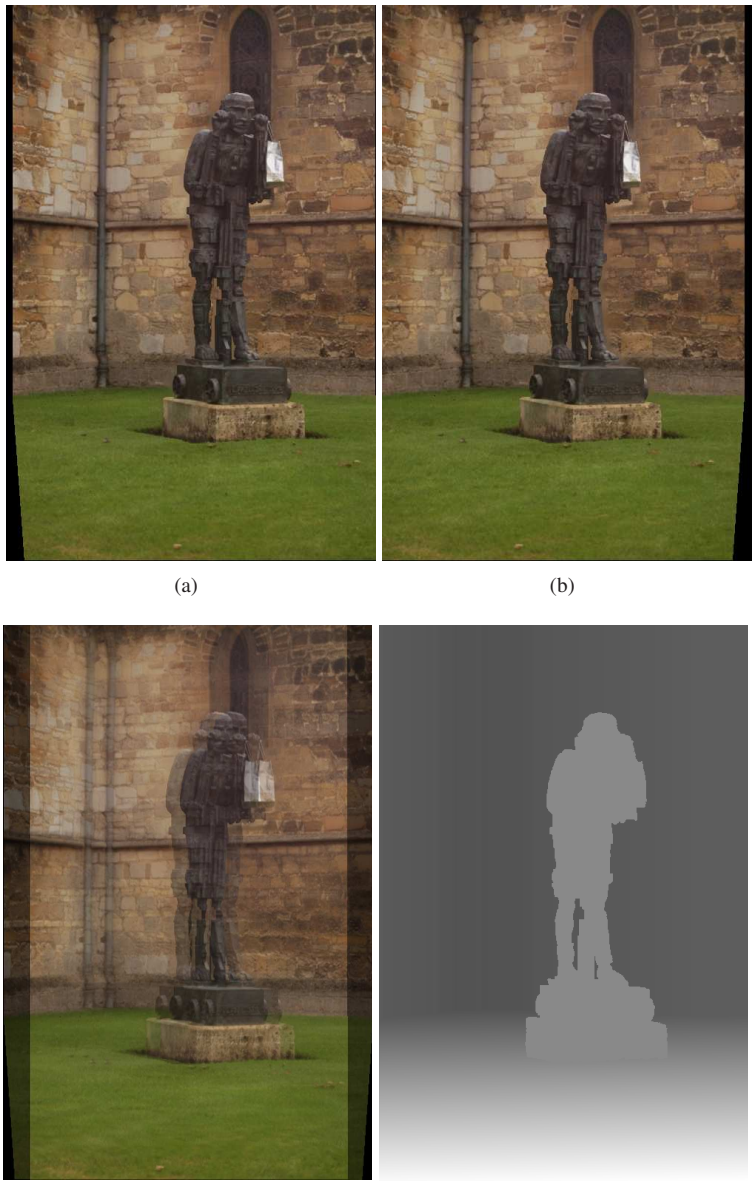
(e) (f)

Figure 7. The "horse" example. (a): Reference image I. (b): Support image J. (c): Left stereoscopic image I_l. (d): Right stereoscopic image I_r. (e): Resulting disparity map for (I_l, I_r). (f): Front-to-parallel view of the ground in the horse case. The ground plane corner forms a right angle as it is delimited by two perpendicular walls.

3. Experimental results

The approach was tested on several image pairs with perceptually pleasing results and a convincing 3D visual effect.

In Figs. 7(a) and (b) are shown the reference and support images of the "horse" pair together with their associated ground plane vanishing lines. The original pair does not form a parallel camera stereoscopic system, as the vanishing lines are not coincident. Figs. 7(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. Fig. 7(e) shows the resulting disparity map. Finally, Fig. 7(f) shows a front-to-parallel view of the ground plane. Such a view, obtained by metric rectification of the ground plane in image I based on the vanishing line and the camera calibration data, provides a clear visual proof of the



(a) (b)

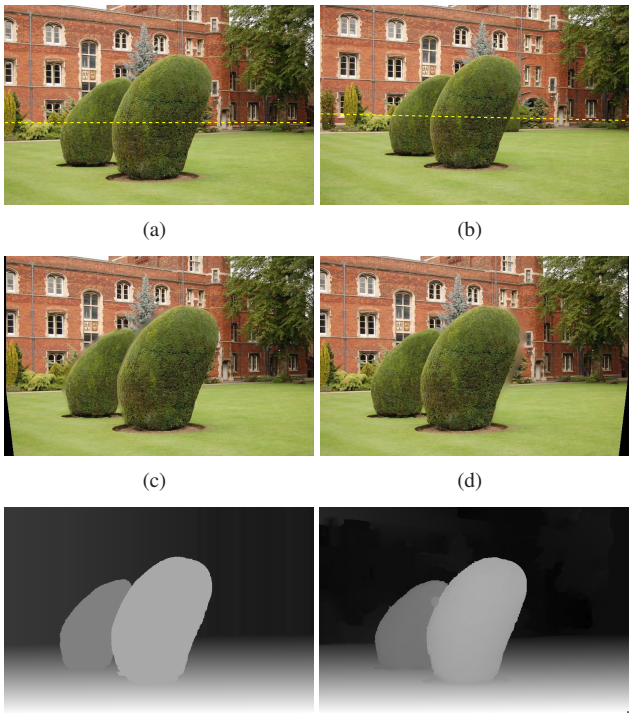
(c) (d)

Figure 6. Stereoscopic rendering for I of Fig. 2(a). (a): I_l. (b): I_r. (c): Superimposed stereoscopic images. (d): Disparity map.

good accuracy of geometric estimates. Indeed, the ground boundaries next to the walls (dashed lines) are almost perfectly orthogonal, as it should be, despite the very slanted view of the ground in the original image.

Figs. 8(a) and (b) illustrate the "bushes" pair, where two partially self-occluding foreground objects are present. Notice, from both Figs. 8(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. As evident from the disparity map of Fig. 8(e), the two bushes are correctly rendered as belonging to two distinct depth layers. The good quality of the disparity map obtained with our approach is confirmed by a visual comparison against the disparity map of Fig. 8(f), which was obtained with a state-of-the-art dense stereo approach [16]: The two maps look very similar. However, *dense stereo is much slower than our approach*, taking about 50 minutes for each image pair on a quad core Intel Xeon 2.5GHz PC. In the present MATLAB implementation of our approach, the overall processing time for an image pair is less than 5 minutes, also taking into account the semi-automatic foreground segmentation procedure.

Fig. 9 illustrates the results obtained with the "bride statues" pair. This pair also includes two foreground objects,

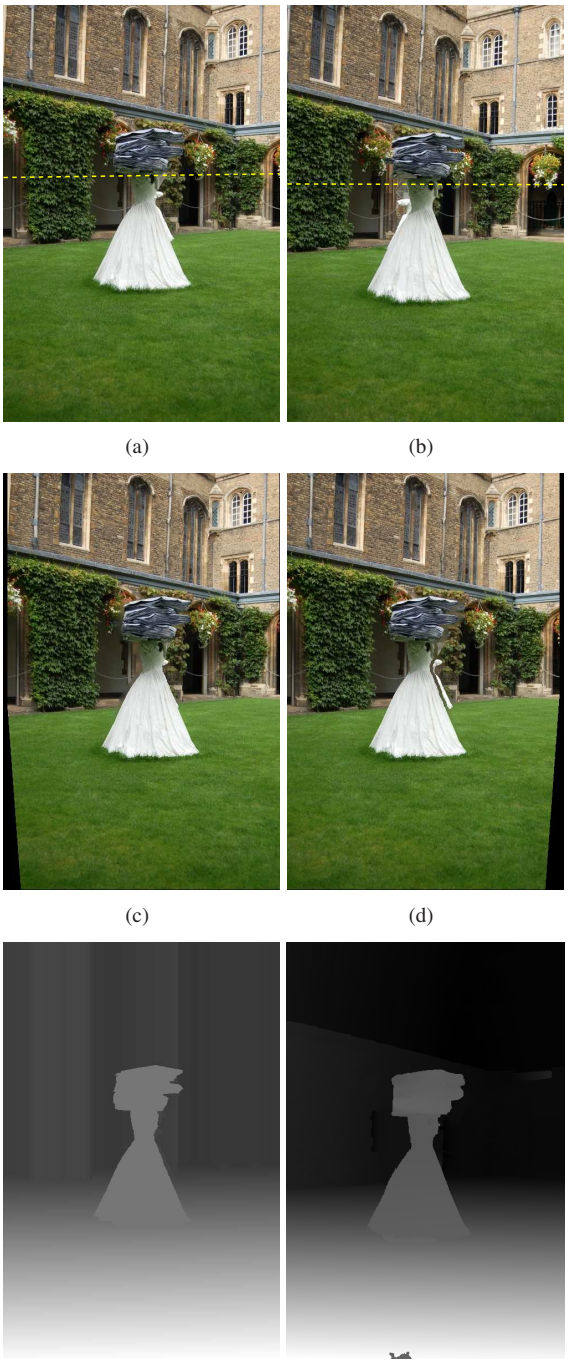


(a) (b) (c) (d) (e) (f)

Figure 8. The "bushes" example. (a): Reference image I. (b): Support image J. (c): Left stereoscopic image I_l. (d): Right stereoscopic image I_r. (e): Disparity map with our approach. (f): Disparity map with a dense stereo approach.

but, differently from the "bushes" pair, the second foreground object is almost completely occluded by the first. However, the disparity map of Fig. 9(e) clearly shows that the unoccluded part of the second foreground object was nevertheless correctly rendered in a depth layer between the first foreground object and the background. Also notice from the disparity map that, due to the ruled surface model, the background is rendered at different depths, thus reflecting the irregular shape of the ground plane upper border in the image. Although the ruled surface model is but an approximation of the real background (as evident from a comparison with the dense stereo disparity of Fig. 9(f), where the shape of the background building is nicely captured, while the second foreground object is totally missing), still it represents the visual scene accurately enough to produce an impressive 3D illusion.

Finally, some frames of a synthetic video generated from the stereo data extracted for the "horse" example (see again Fig. 7) are shown in Fig. 10. The camera performs a virtual translation along its $x-$axis, showing the parallax effect on the horse position w.r.t. the background.



(a) (b) (c) (d) (e) (f)

Figure 9. The "bride statues" example. (a): Reference image I. (b): Support image J. (c): Left stereoscopic image I_l. (d): Right stereoscopic image I_r. (e): Disparity map with our approach. (f): Disparity map with a dense stereo approach.

4. 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

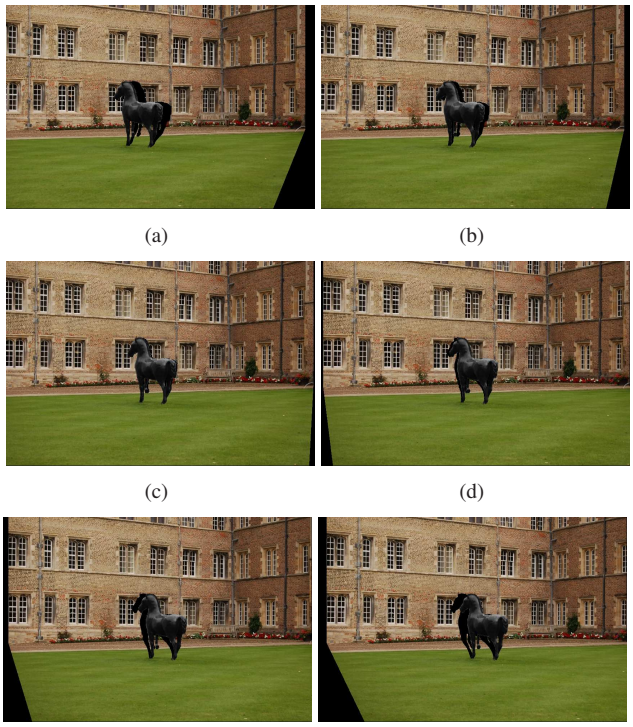


Figure 10. Some frames of a synthetic video sequence for the "horse" example of Fig. 7. The camera translates along its $x-$axis from right to left. Black pixels around the horse correspond to occluded background points.

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 of temporal consistency among frames), (2) relaxing the ground plane assumption, (3) performing a totally automatic image segmentation based on a multi-planar scene model, thus further speeding up computations (in the current implementation, more than 90% of the time is taken by the semi-automatic foreground object segmentation) while retaining the basic geometric structure of the approach expounded in subsect. 2.1, (4) implementing an automatic method to determine the optimal range of disparities for 3D perception.

Acknowledgements

We heartily thank Oliver Woodford for providing us with the experimental results used to compare our approach with his dense stereo method [16].

References

[1] S. Coren, L. M. Ward, and J. T. Enns. *Sensation and Perception*. Harcourt Brace, 1993. 2

[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. 2

[3] 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. 3

[4] M. Guttman, L. Wolf, and D. Cohen-Or. Semi-automatic stereo extraction from video footage. In *Proc. IEEE International Conference on Computer Vision*, 2009. 1

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 3

[6] A. Heyden and M. Pollefeys. Multiple view geometry. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*. Prentice Hall, 2005. 3

[7] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal on Computer Vision*, 75(1), 2007. 2, 4

[8] G. Jones, D. Lee, N. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems VIII*, volume 4297, 2001. 1

[9] J. Koenderink, A. van Doorn, A. M. L. Kappers, and J. T. Todd. Ambiguity and the 'mental eye' in pictorial relief. *Perception*, 30(4):431–448, 2001. 2

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2):91–110, 2004. 3

[11] P. Mendonça and R. Cipolla. A simple technique for self-calibration. In *Proc. Conf. Computer Vision and Pattern Recognition*, 1999. 3

[12] V. Nedovic, A. W. M. Smeulders, A. Redert, and J. M. Geusebroek. Stages as models of scene geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (in press), 2010. 1

[13] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (SIGGRAPH)*, 2004. 5

[14] A. Saxena, M. Sun, and A. Y. Ng. Learning 3-d scene structure from a single still image. In *Proc. IEEE International Conference on Computer Vision*, pages 1–8, 2007. 2

[15] P. Sturm. On focal length calibration from two views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2001. 3

[16] 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. 1, 6, 8

[17] G. Zhang, W. Hua, X. Qin, T. T. Wong, and H. Bao. Stereoscopic video synthesis from a monocular video. *IEEE Transactions on Visualization and Computer Graphics*, 13(4):686–696, 2007. 2

[18] G. Zhang, J. Jia, T.-T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(6):974–988, 2009. 2