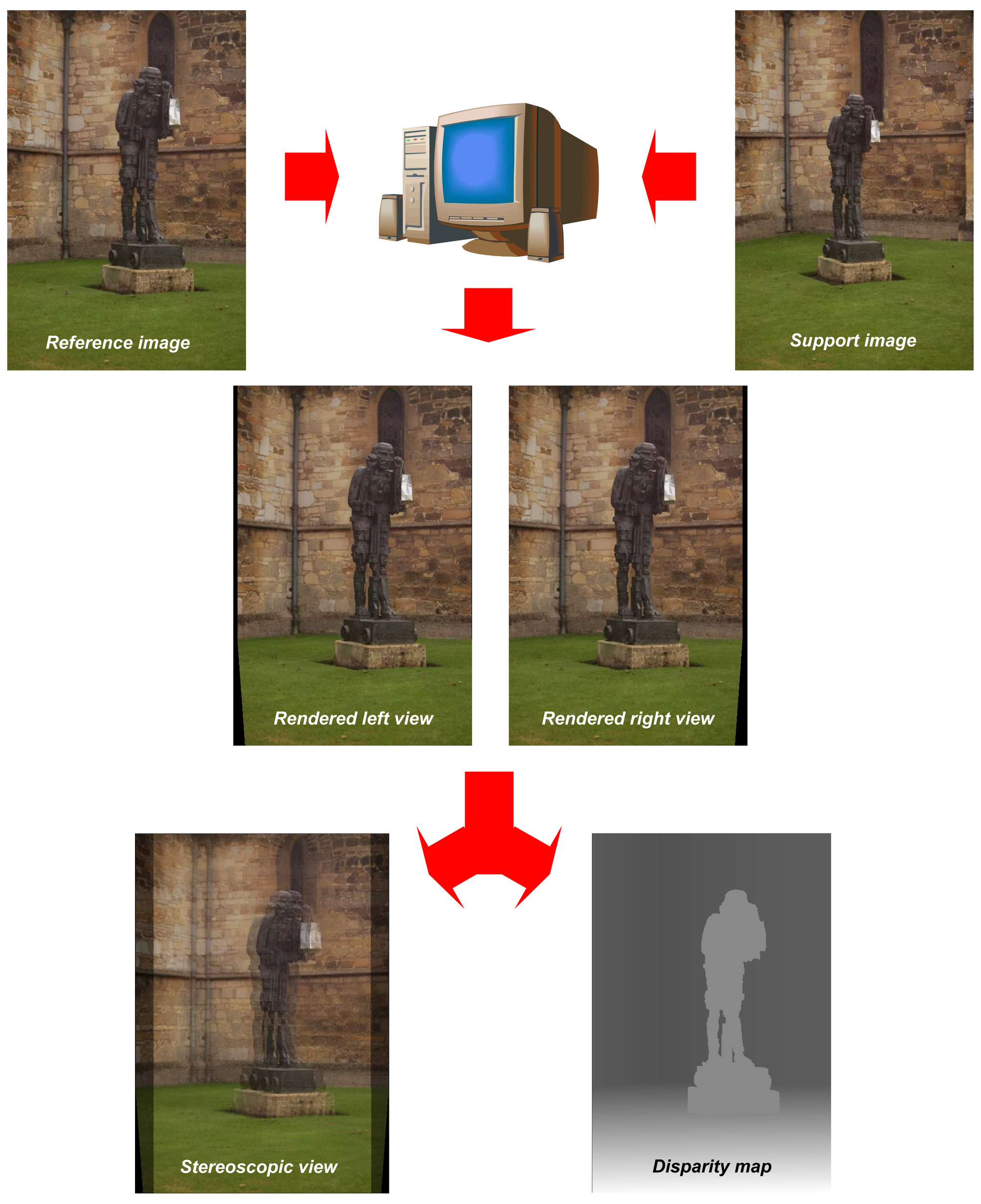
2D-to-3D Photo Rendering for 3D Displays

Dario Comanducci[1] **Atsuto Maki**[2] **Carlo Colombo**[3] **Roberto Cipolla**[4]

[1,3] Dipartimento di Sistemi e Informatica, University of Florence, Italy — {comandu, colombo}@dsi.unifi.it
[2] Toshiba Research Europe Ltd, Cambridge UK — atsuto.maki@crl.toshiba.co.uk
[4] Dept. of Engineering, University of Cambridge, Cambridge UK — cipolla@eng.cam.ac.uk

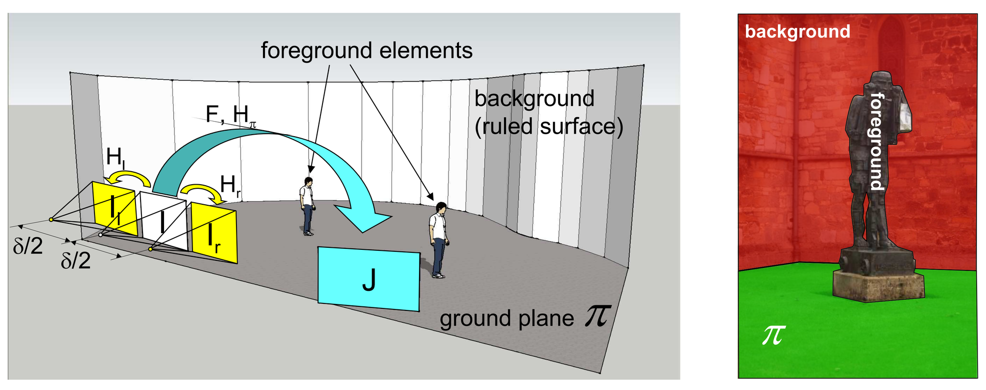
Problem

Semi-automatic 2D-3D conversion of an image pair for rendering on stereoscopic displays.



- ⋄ Inputs: *reference* image I, *support* image J
- ⋄ I is transformed into the *virtual* stereo pair (I_l, I_r)
- ⋄ J is employed to recover calibration and geometry
- ⋄ Roles of I and J can be swapped

Theatrical scene model



- ⋄ Ground plane π
- ⋄ Background: vertical ruled surface
- ⋄ Foreground elements: vertical and flat

Overview

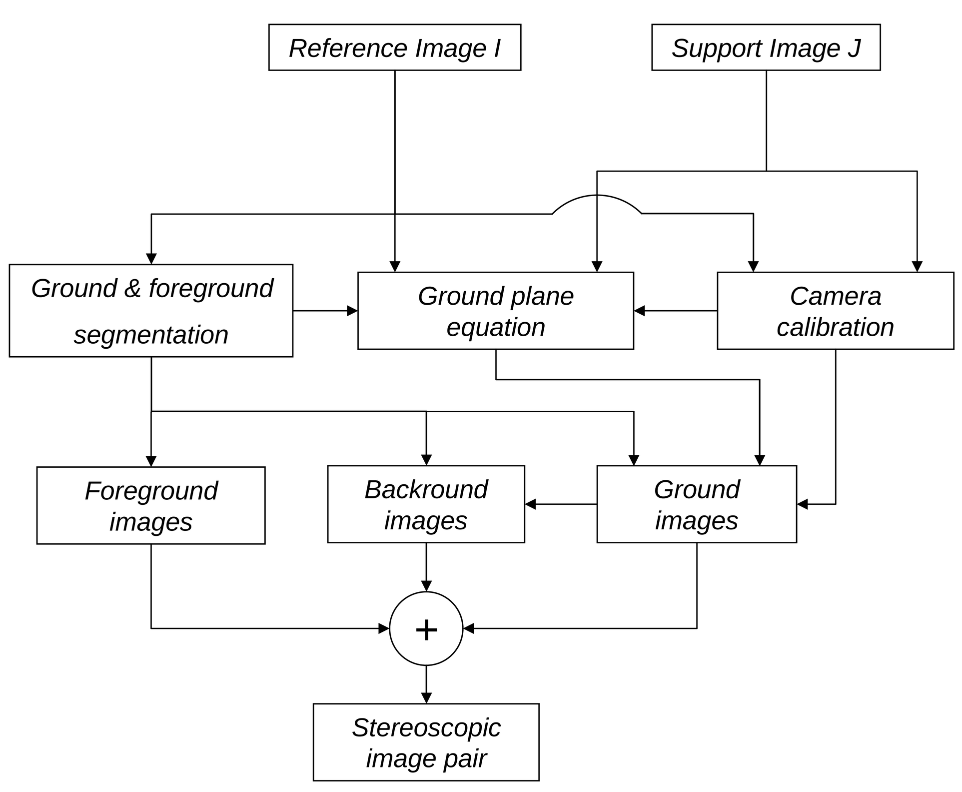


Image segmentation

- ⋄ Ground plane: supervised learning [1]
- ⋄ Foreground: interactive [2]
- ⋄ Background: the remaining regions

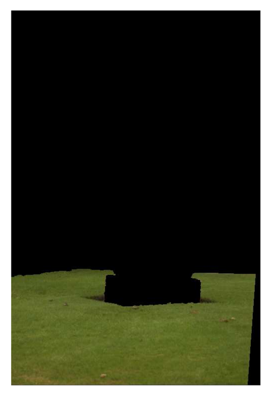
Rendering

Ground plane

L and R homographies induced by π:

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

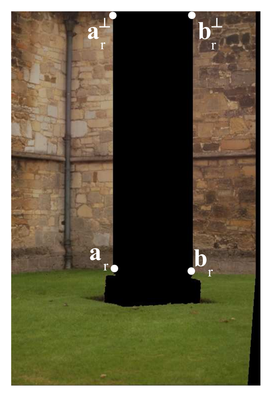
- ⋄ $[\mathbf{n}_\pi^\top \ d_\pi]^\top$ ground plane parameters
- ⋄ $\mathtt{K}_i$ calibration matrix for view I
- ⋄ $\mathbf{s} = [\delta/2\ 0\ 0]^\top$
- ⋄ δ baseline for the virtual cameras



Ground rendering for I_r.

Background

- ⋄ For each point $\mathbf{P}$ on the top border of π, its image $\mathbf{p} \in I$ is mapped onto I_l and I_r, respectively as $\mathbf{p}_l = \mathtt{H}_l\mathbf{p}$ and $\mathbf{p}_r = \mathtt{H}_r\mathbf{p}$. The image of the vertical segment above $\mathbf{P}$ is then rendered on either image as $\mathbf{q}_{l,r}(\lambda) = \mathbf{p}_{l,r} + \lambda\mathbf{v}^\perp$, where $\mathbf{v}^\perp$ is the vanishing point of the normal to π.



- ⋄ When (e.g. in I_r) the top border of the ground is occluded by the foreground between $\mathbf{a}_r$ and $\mathbf{b}_r$, the quadrilateral $\{\mathbf{a}_r, \mathbf{b}_r, \mathbf{a}_r^\perp, \mathbf{b}_r^\perp\}$ is computed, where $\mathbf{a}_r^\perp$ and $\mathbf{b}_r^\perp$ are respectively the intersections of the lines $\mathbf{a}_r \times \mathbf{v}^\perp$ and $\mathbf{b}_r \times \mathbf{v}^\perp$ with the border of I_r. Unoccluded background points within this quadrilateral are mapped from I by a suitable homography.

- ⋄ Background regions that are occluded in I by the foreground are filled in I_l and I_r by color interpolation.

Foreground

- ⋄ Foreground elements are rendered as flat and vertical silhouettes.
- ⋄ Depth is assigned as the value corresponding to the point of contact with the ground (the lowest silhouette point).

Geometry

Camera self-calibration

- ⋄ Robust estimation of fundamental matrix $\mathtt{F}$.
- ⋄ Estimation of internal camera matrices $\mathtt{K}_i$, $\mathtt{K}_j$ by forcing $\hat{\mathtt{E}} = \mathtt{K}_j^\top\mathtt{F}\mathtt{K}_i$ to have the same properties of the essential matrix [3].
- ⋄ Recovery of extrinsic camera parameters, i.e. rotation matrix $\mathtt{R}$ and translation vector $\hat{\mathbf{t}} = \mathbf{t}/||\mathbf{t}||$ by factorization of $\hat{\mathtt{E}} \sim [\mathbf{t}]_\times\mathtt{R}$.

Ground plane equation

- ⋄ $\mathbf{n}_\pi = \mathtt{K}_i^\top\mathbf{l}_\pi$, where $\mathbf{l}_\pi$ is the vanishing line of π in image I.
- ⋄ $\mathbf{l}_\pi$ is the fixed line of the planar homology $\mathtt{H}_p = \mathtt{H}_\pi^{-1}\mathtt{H}_\infty$, where $\mathtt{H}_\infty = \mathtt{K}_j\mathtt{R}\mathtt{K}_i^{-1}$.
- ⋄ $\mathtt{H}_\pi$ is the homography induced by π between I and J. It is parameterized as $\mathtt{H}_\pi(\mathbf{v}) = [\mathbf{e}']_\times\mathtt{F} - \mathbf{e}'\mathbf{v}^\top$, being $\mathbf{e}'^\top\mathtt{F} = 0$. The parameter vector $\mathbf{v}$ is computed from at least 3 point correspondences.

- ⋄ d_π is obtained by triangulating any two corresponding points of the ground plane.

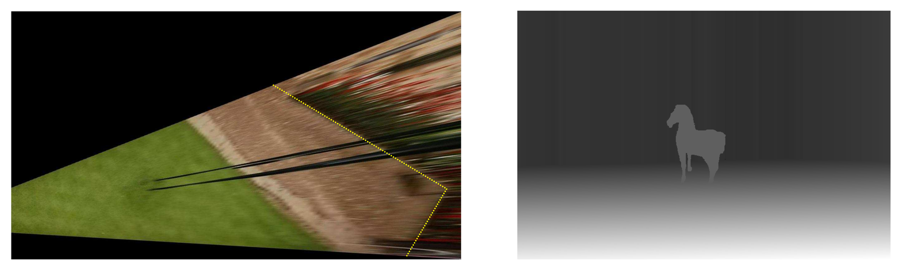
Vertical vanishing point

It is obtained from the normal to π as $\mathbf{v}^\perp = \mathtt{K}_i\mathbf{n}_\pi$ and is the same for I, I_l, I_r.

Results

1. Horse

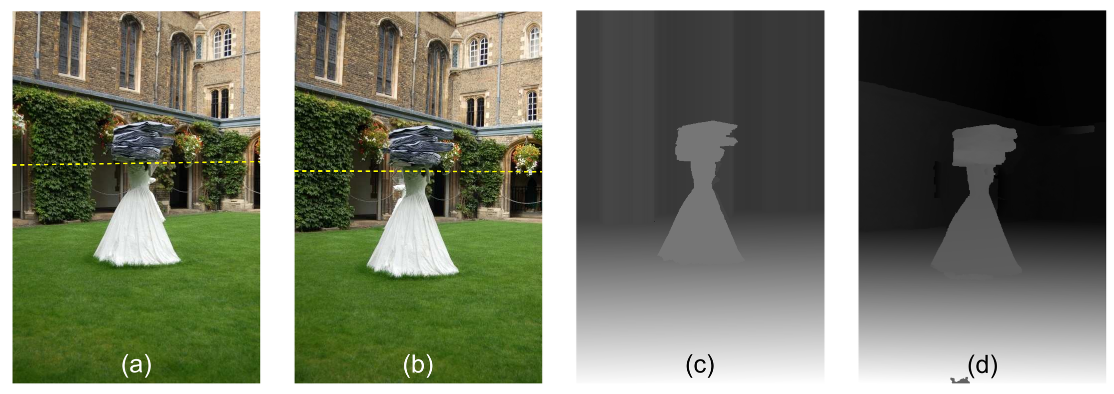
I and J and their estimated vanishing lines.



Left: Rectified ground plane. The right angle between the two walls is correctly recovered, proving that geometry was correctly estimated. Right: Estimated disparity map.

Some frames of a synthetic video sequence obtained from image I. The camera performs a virtual translation along the x-axis. Occluded background points are colored in black.

2. Bride



(a, b): I and J and their estimated vanishing lines. (c): Recovered disparity map: The effect of the ruled surface assumption id visible in the left picture; nevertheless the visualized stereoscopic image is perceived without artifacts. (d): Comparison with a state-of-the-art dense stereo algorithm [4].

3. Bushes

I and J and their estimated vanishing lines.

Recovered disparity map (left) compared with the result of [4] (right).

Conclusion

- ⋄ 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.
- ⋄ Comparing the disparity map with the result of [4], the two maps look very similar. However, dense stereo is ten times slower than our approach.

Selected references

1. D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. IJCV, 75(1), 2007.
2. C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. SIGGRAPH, 2004.
3. P. Mendonça and R. Cipolla. A simple technique for selfcalibration. CVPR, 1999.
4. O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. PAMI, 31(12), 2009.