Layered Photo Pop-Up

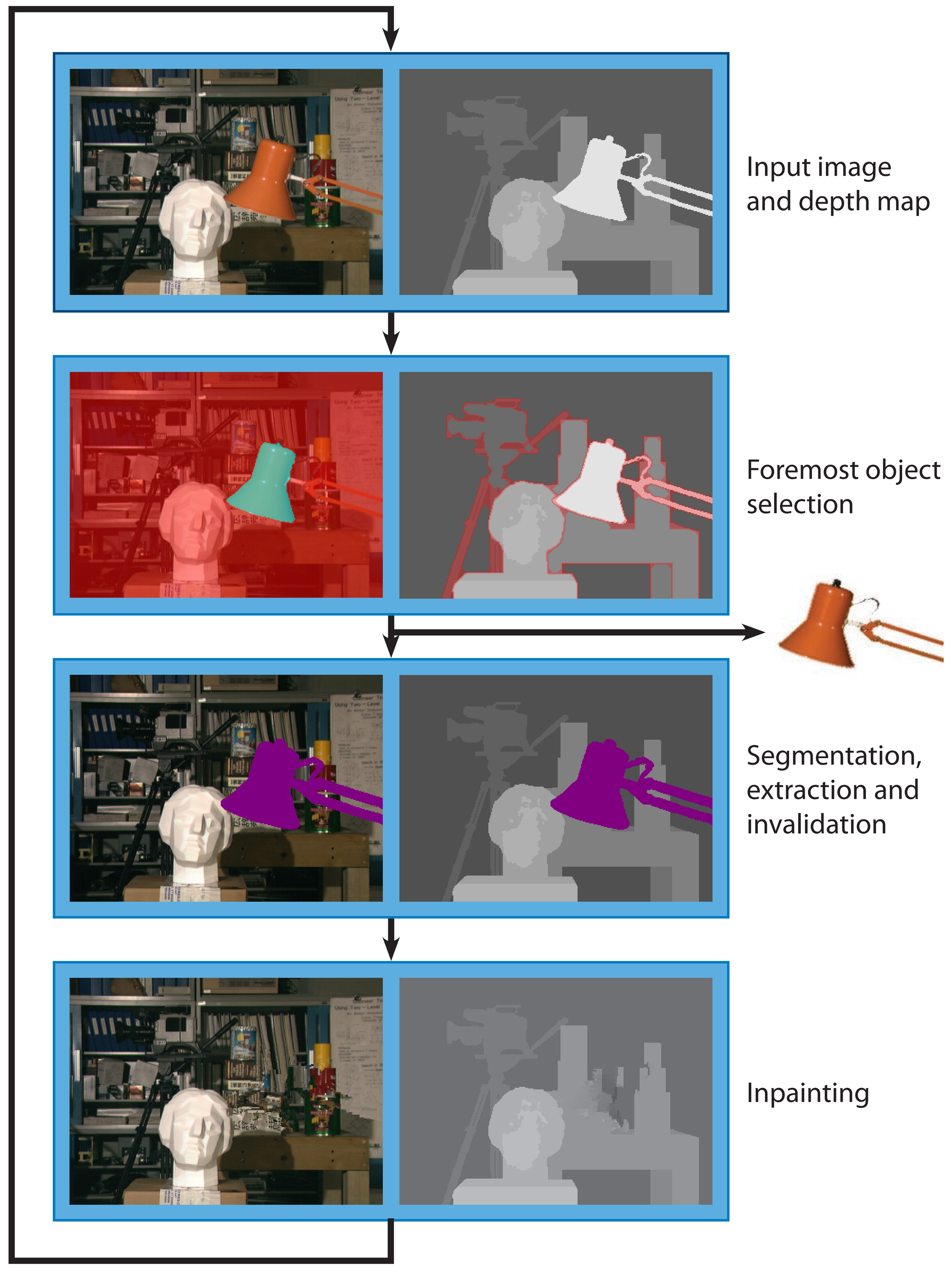
Lech Świrski, Christian Richardt & Neil A. Dodgson

Motivation

Documentaries present static images in an engaging manner by panning across them in 3D. This is normally a laborious process involving hours of manual rotoscoping. Our system automates the process, if given an image with depth.

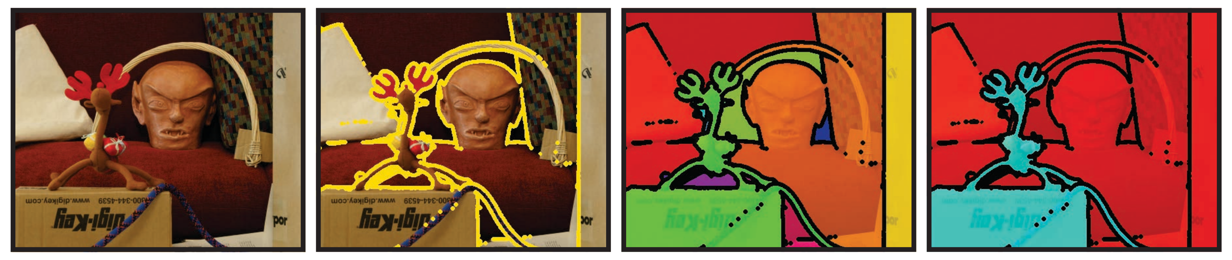
Overview

Our system takes an image plus depth information. We segment this into layers, which are used to synthesise new views of the scene and create effects that rely on occluded image and depth data, such as camera pans and depth-of-field.



Foremost object selection

The foremost object is automatically detected by finding object boundaries at depth discontinuities. Our system labels continuous regions inside these boundaries, and selects the foremost region as that which is in front of all its neighbouring regions.



Input — Discontinuities — Labelled regions — Foremost region

Hard Segmentation

We use a hard segmentation technique based on GrabCut[1]. Where the original GrabCut uses colour pixels, that is pixels in 3D RGB space, we extend it to depth by considering pixels in 4D RGBZ space; this gives a cleaner segmentation where objects at different depths have similar colour.

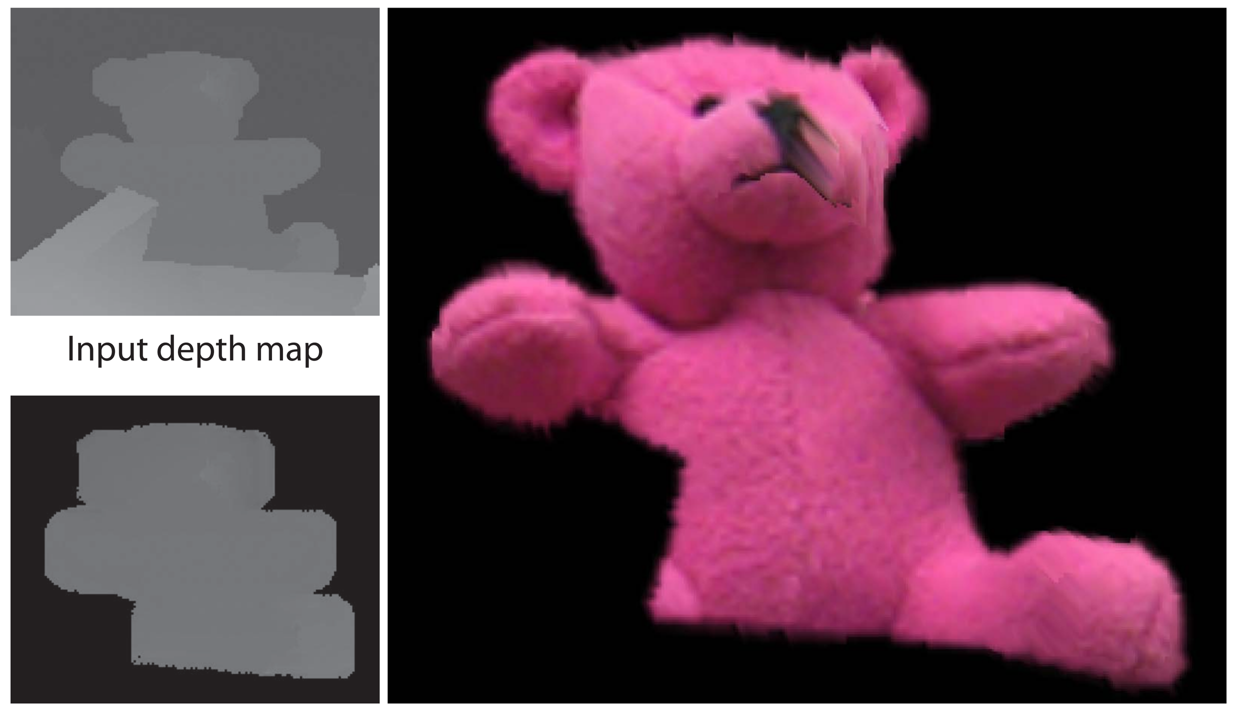
GrabCut RGB segmentation — Our RGBZ segmentation

Soft Segmentation

After the hard segmentation, we perform a soft segmentation to create an alpha matte for the foremost object. We trace along the object's contour, and compute the colour of each foreground pixel using Bayesian Matting[2].

Hard segmentation — Soft segmentation

To compute the corresponding depth map, we recalculate the depth of all semi-transparent pixels along the object's contour. This is done by fitting a plane to the local neighbourhood.



Input depth map — Output depth map — Combined colour and depth

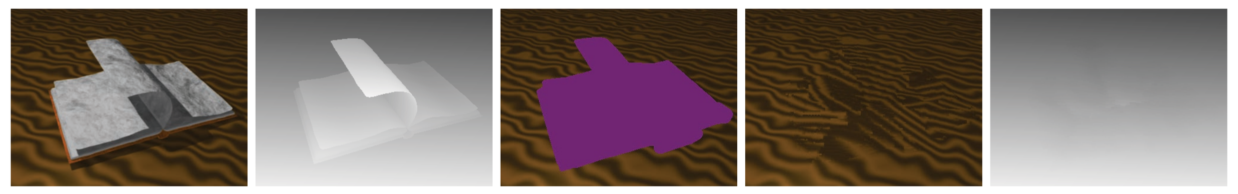
Inpainting

Criminisi's exemplar-based inpainting

Once the foreground object is extracted, the area behind it is invalidated and filled in using exemplar-based inpainting[3], extended to use and fill depth information. We select source patches by finding patches which are:

- Similar in colour — this favours patches with similar colour content and image structure, as in the original algorithm
- Similar in position — this favours patches which are nearer to the area being filled.
- Similar in 3D shape — this is the depth map equivalent of similarity in colour.
- Valid in depth ordering — this favours patches which will be behind the object being removed.

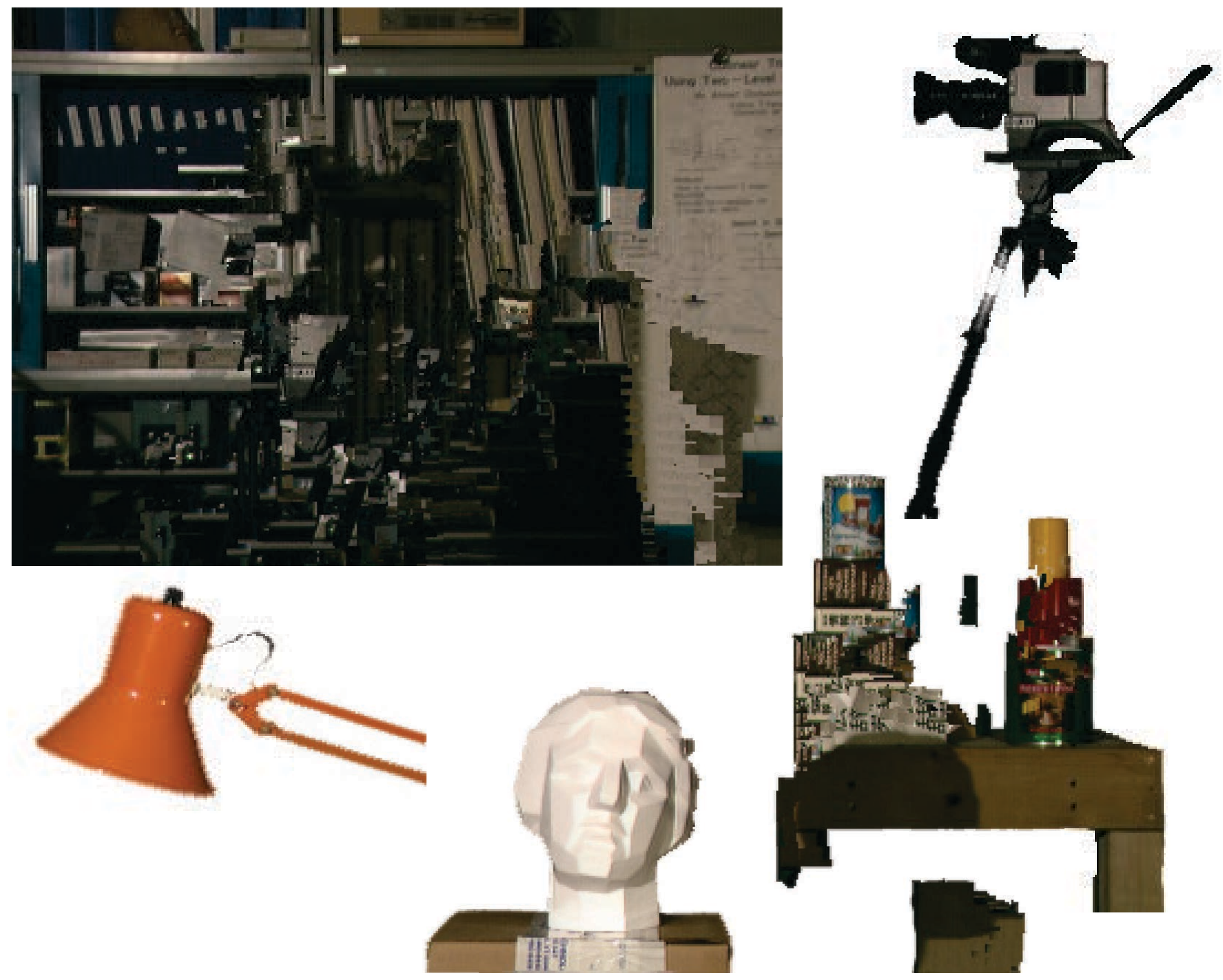
These terms are inspired by Stereoscopic Inpainting[4], which we improve on by using depth rather than disparity, and zero-normalising depths of patches. This allows fairer patch comparisons and improves inpainting of non-fronto-parallel planes.



Input colour and depth — Invalidated — Inpainted colour and depth

Result

The output of our system is a set of layers, with each layer being an image with depth. These layers can then be rendered easily, for example as meshes, using standard methods. Effects, such as depth-of-field, can be applied as they would be to normal 3D geometry, without occlusion issues or rubber sheet effects.



Layers automatically segmented by our system

References

Rother, C., Kolmogorov V., and Blake A. 2004. GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Computer Graphics*
Chuang, Y.-Y., Curless, B., Salesin, D. H., and Szeliski, R. 2001. A Bayesian approach to digital matting. In *CVPR*
Criminisi, A., Perez, P., and Toyama, K. 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*
Wang L., Jin H., Yang R., and Gong M. 2008. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *CVPR*

Without our system: Panning across a single image with depths produces undesirable rubber-sheet effects.

With our system: Panning across our layered representation trivially solves occlusion problems, has no rubber-sheet effects, and allows us to add a depth-of-field effect.

EPSRC
Engineering and Physical Sciences Research Council
UNIVERSITY OF CAMBRIDGE
Computer Laboratory
RAINBOW RESEARCH GROUP