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, 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.

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.

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\(^2\).

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-norma- lising depths of patches. This allows fairer patch comparisons and improves inpainting of non-fronto-parallel planes.

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.

References

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.

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.