Slow Glass: Visualizing History in 3D

Xuan Luo[1] Yanmeng Kong[1] Jason Lawrence[2] Ricardo Martin-Brualla[2] Steven M. Seitz[1,2]
[1]University of Washington [2]Google

Abstract

We introduce new techniques for reconstructing and viewing antique stereographs in 3D. Leveraging the Keystone-Mast image collection from the California Museum of Photography, we apply multiple processing steps to produce clean stereo pairs, complete with calibration data, rectification transforms, and disparity maps. We describe an approach for synthesizing novel views from these scenes that runs at real-time rates on a mobile device, simulating the experience of looking through an open window into these historical scenes.

1. Introduction

Wouldn't it be fascinating to be in the same room as Abraham Lincoln, visit Thomas Edison in his laboratory, or step onto the streets of New York a hundred years ago? In his 1966 science fiction story *Light of Other Days*, Bob Shaw [14] describes a transparent glass-like material so dense that it takes light many years to pass through. If a 150-year-pane of this *slow glass* were placed in the White House in the 1860s, looking through it today would be just like peering through a window at Abraham Lincoln himself. While slow glass does not exist, *simulating* such an experience may be possible, through the combination of advances in 3D modeling technology and virtual and augmented reality. Key to this possibility are antique *stereographs*, i.e., photos of historic scenes captured in stereo. Stereo cameras and viewers were invented in the mid 1800s, and quickly became very popular. Many of the world's most important people, events, and places over the following century were captured in stereo, and thousands of these stereographs survive to this day (Fig. 1).

This imagery opens up the fascinating possibility of reconstructing and visualizing historical scenes in 3D, and represents a potentially valuable resource for the computer vision community. However, these antique stereographs are not in a form that facilitates analysis and research. They are uncalibrated, unaligned, and contain many artifacts (scratches, damage, dirt, exposure differences, contrast loss, scanning errors, etc.). Another limitation of these stereographs is that they require special viewer hardware to experience the stereo effect.

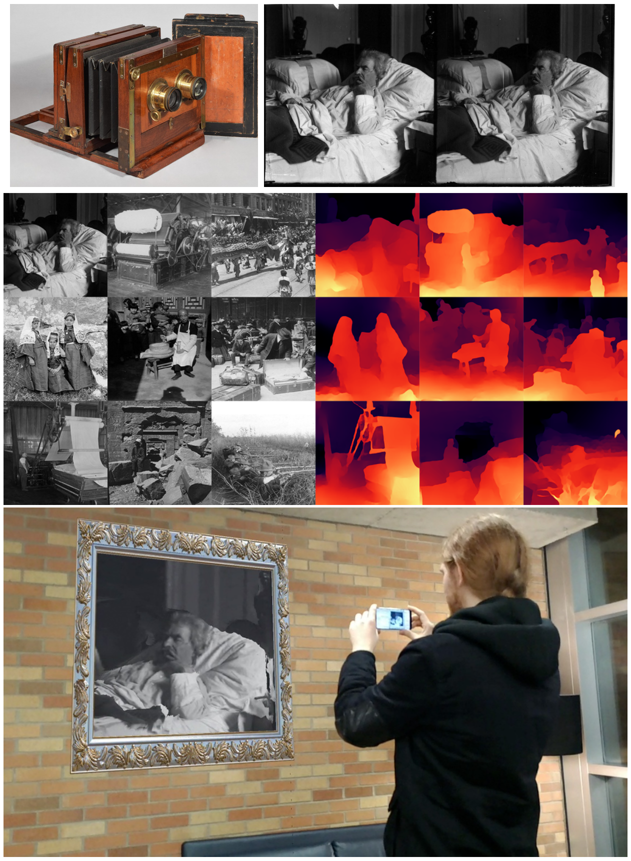


Figure 1: *Top:* Stereo cameras were invented in the 1850s and hundreds of thousands of antique stereographs are available today, like this one of Mark Twain. *Middle:* We describe a pipeline for estimating disparity maps from this imagery. *Bottom:* We also describe a real-time view synthesis technique that powers an AR application that creates the sensation of looking through an open window onto these historical scenes.

We describe a novel processing pipeline for these antique stereographs that produces rectified stereo pairs, along with a view synthesis technique for synthesizing smooth camera paths. This powers our *Slow Glass* Augmented Reality (AR) application that creates the experience of looking through a window onto these lost historical scenes (Fig. 1).

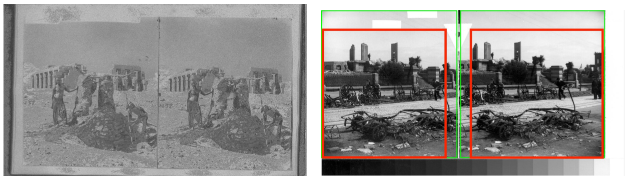


Figure 2: *Left:* Example stereographs that were culled due to excessive image artifacts. *Right:* Bounding boxes showing the full views and ”artifact free” regions in green and red, resp.

2. The *KeystoneDepth* Collection

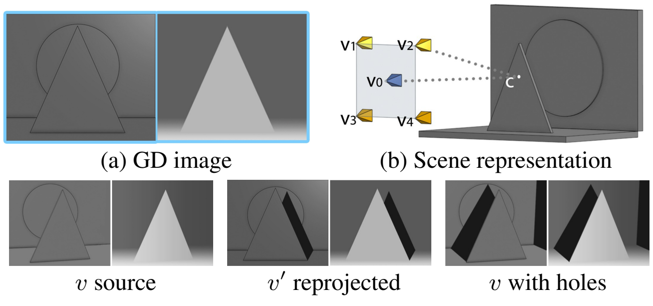
The Keystone-Mast Collection [16] is the largest known collection of antique stereographs.[1] It consists primarily of contact prints [17] made from the original negatives of the Keystone View Company, a major distributor of stereoscopic images. These images capture a wide variety of subjects ranging from popular tourist destinations and American presidents, to scenes of daily life on farms and factories during that time period.

Significant work is needed, however, to transform this raw imagery into a form that facilitates research and analysis. We hired a crowdsourcing company [12] to manually filter out images unsuitable for depth estimation, such as monocular images, backs of stereo-cards and poor quality images, and to correct upside-down images and inverted intensities. In a second step, we manually specify two pairs of axis-aligned bounding boxes marking 1) the *full* left and right views and 2) the largest artifact-free rectangular region on each side (Fig. 2). Finally, we rectify the full stereo views using the method proposed by Loop and Zhang [7], and estimate disparity maps using FlowNet2 [5] over the artifact-free regions.

Altogether, we have processed 37,244 stereo pairs to date, which we plan to make publicly available online as the *KeystoneDepth* collection. Each entry consists of the original stereoscopic image, metadata (description, date, subjects, original URL etc.), a rectified stereo pair, camera parameters, and a pair of disparity maps.

3. Visualizing Antique Stereographs

One limitation of stereographs is the need of special viewer hardware to experience the stereo effect. We wish to visualize stereographs in a way that conveys the depth of the scene from a regular 2D display, by synthesizing a continuous 3D camera path, similar to what you would see if you were present and moving your head around the scene itself. Synthesizing novel views can also enable novel AR experiences (Sec. 3.5). One option is *view interpolation* [2, 8, 11], which produces a narrow range of viewpoints in between the two inputs (limited to the space between the viewer’s eyes). While multiple input views have traditionally been needed to achieve larger head volumes [1, 4, 10, 19, 22, 20, 13], recent work has shown impressive view extrapolation from a single stereo view or even a single monocular view using deep neural networks [3, 9, 15, 18, 21].



(a) GD image (b) Scene representation

v source v' reprojected v with holes

(c) Double reprojection

Figure 3: Our scene representation for visualizing stereographs. (a) The input is a single grayscale intensity and depth image (“GD image”) computed at the left stereo view. (b) We compute four new GD images as seen from the corners of a quadrilateral surrounding the input image to expand the viewing volume. (c) We reproject a GD image at viewpoint v to a different viewpoint v' and then back to v, in order to generate a pattern of holes at v that are characteristic of those revealed by viewpoint changes.

One limitation of these recent deep net approaches [3, 9, 15, 18, 21] is the need to train on modern imagery – YouTube videos [21] or video game framegrabs [3] – which may not match the statistics of antique imagery. We describe an approach specifically designed for antique stereographs.

3.1. Scene Representation

We seek a lightweight scene representation that can be interactively viewed on smartphones and web browsers. Following Chen and Williams [2], we represent scene appearance using multiple depth images, each composed of a grayscale image (G) aligned to a depthmap (D), denoted *GD*. Because the Keystone stereographs have a small baseline, we support a larger viewing region by synthesizing reference GDs at the corners of a desired headbox, defined by a rectangle aligned with the image plane (Fig. 3). Reprojecting and compositing these four reference GDs and the GD for the left stereograph view produces high quality views within and just in front of this rectangle.

3.2. Intensity and Depth Inpainting

We aim to synthesize hole-free GD images at the corners of the rectangle. If we directly apply a state-of-the-art color inpainting technique [6] to fill in the holes from reprojection, the inpainted content tends to blend foreground and background content together as shown in Fig. 4. We observe, however, that holes in these reprojected images have a special structure: they follow object *boundaries* and correspond primarily to *background* regions in the scene that be-

[1] http://ucr.emuseum.com/collectionoverview/3631

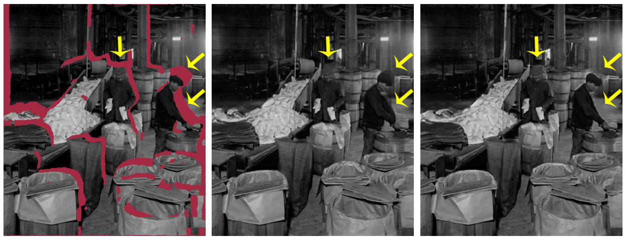


Figure 4: *L*eft: Sample input GD with holes. *M*iddle: Inpainted result using method of Liu et al. [6]. *R*ight: Result with our proposed double reprojection and boundary guidance, which produces sharper transitions between foreground and background.

come visible due to disocclusions. Therefore, we synthesize these *disocclusion holes* in our training data using a *double reprojection* technique and introduce a *boundary mask* input to help the network fixate on inpainting the background region of disocclusion holes in the corner GD images. We find these adaptations help to produce sharper transitions between the foreground and background (Fig. 4).

Double Reprojection (DR). We project the input GD image recorded at viewpoint v to a new viewpoint v' and then reproject it back to v (Fig. 3c). We use the original hole-free GD image at v to supervise content to be inpainted into the generated holes at v. We call this technique *double reprojection*, and it has the advantage of only requiring a single reference GD to produce a large number of training images with holes and hole-free ground truth.

Boundary Guidance (BG). To avoid blurry transitions between the foreground and background, we introduce a boundary mask as an additional input to the network, which indicates the location of depth discontinuities. Specifically, we detect pixels on the foreground side of the reprojection holes and store them in a binary mask image B.

Networks and Losses. We adapt the partial convolution network [6] to inpaint depths and intensities. Let I be the grayscale image, M be the binary mask of holes, and $\hat{D}$ be the normalized disparity map, i.e., disparities scaled to [0, 1]. We train two separate networks that take $< I, \hat{D}, B >$ masked by $1 - M$ as input and output inpainted intensities and normalized disparities $\hat{D}_p$, respectively. For intensity inpainting, we use the same objective as Liu et al. [6]. For depth inpainting, we seek to minimize the following loss

$$L(\hat{D}, \hat{D}_p; M) = L_{valid} + \lambda_{hole} L_{hole} + \lambda_{tv} TV(\hat{D}_{comp}),$$

$$\begin{aligned} L_{valid} &= ||(\hat{D} - \hat{D}_p) \odot (1 - M)||_1, \\ L_{hole} &= ||(\hat{D} - \hat{D}_p) \odot M||_1, \\ \hat{D}_{comp} &= M \odot \hat{D}_p + (1 - M) \odot \hat{D}, \end{aligned}$$

where $TV(x)$ is the total variation loss. We set $\lambda_{hole} = 6, \lambda_{tv} = 0.1$.

3.3. Implementation Details

We synthesize holes with double reprojection as well as by simulating random strokes. We quantitatively study the effect of DR and BG on 74 random samples from the collection. Compared with the baseline partial convolution network [6], adding DR or BG alone decreases the MSE of the inpainted normalized disparity map from 9.99×10^{-4} to 3.56×10^{-4} and 2.98×10^{-4}, respectively. Combined, the MAE is reduced to 2.15×10^{-4}.

3.4. Visualization Quality

To evaluate the quality of our view synthesis technique, we generated videos for stereographs in our collection that show a continuous camera path. We manually labeled 711 randomly chosen videos as: “very few artifacts,” “some artifacts,” or “failure.” We found 23% to have very few artifacts, 49% have some artifacts, and the remaining 28% are failures. The “some artifacts” results, while imperfect, are still useful in visualizing the 3D effect, which is otherwise difficult to achieve on a 2D display. As such we were able to generate visualization results for approximately 71% of the collection.

3.5. "Slow Glass" AR App

We use our view synthesis technique to power an Augmented Reality (AR) mobile application that simulates the experience of looking through a window onto a historical scene (Fig. 1). The application works by detecting the plane of a wall facing the user and then places a simulated window on the wall and uses our view synthesis technique to generate continuous viewpoint changes through this virtual window as the user moves their phone. Note viewing from a different perspective shows parallax as in Fig. 1.

4. Conclusion

We convert a large collection of antique stereographs into rectified stereo imagery with disparity maps. We introduced a novel view synthesis technique and “Slow Glass” AR App for viewing historical imagery as if looking through a window.

Acknowledgements

We thank Ying Wang for help on the dataset, Ayush Saraf and Brennan Stein for help on the “Slow Glass” AR App. We are grateful to Konstantinos Rematas, Alex Colburn, Kiana Ehsani and Aditya Sankar for many helpful comments, and Rick Szeliski for introducing us to Slow Glass. We also want to give special thanks to the staff at the California Museum of Photography, in particular Leigh Gleason, for their help accessing and understanding the Keystone-Mast collection.

References

[1] Gaurav Chaurasia, Sylvain Duchene, Olga Sorkine-Hornung, and George Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Transactions on Graphics (TOG)*, 32(3):30, 2013.

[2] Shenchang Eric Chen and Lance Williams. View interpolation for image synthesis. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 279–288, New York, NY, USA, 1993. ACM.

[3] Inchang Choi, Orazio Gallo, Alejandro Troccoli, Min H Kim, and Jan Kautz. Extreme view synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7781–7790, 2019.

[4] Peter Hedman and Johannes Kopf. Instant 3d photography. *ACM Transactions on Graphics*, 37:1–12, 07 2018.

[5] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *IEEE conference on computer vision and pattern recognition (CVPR)*, volume 2, page 6, 2017.

[6] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

[7] Charles Loop and Zhengyou Zhang. Computing rectifying homographies for stereo vision. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 125–131. IEEE, 1999.

[8] Dhruv Mahajan, Fu-Chung Huang, Wojciech Matusik, Ravi Ramamoorthi, and Peter N. Belhumeur. Moving gradients: a path-based method for plausible image interpolation. *ACM Trans. Graph.*, 28(3):42:1–42:11, 2009.

[9] Simon Niklaus, Long Mai, Jimei Yang, and Feng Liu. 3d ken burns effect from a single image. *ACM Transactions on Graphics (TOG)*, 38(6):1–15, 2019.

[10] Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. 36(6), 2017.

[11] Steven M. Seitz and Charles R. Dyer. View morphing. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 21–30, New York, NY, USA, 1996. ACM.

[12] InforSearch BPO Service. http://www.infosearchbpo.com/.

[13] Jonathan Shade, Steven Gortler, Li-wei He, and Richard Szeliski. Layered depth images. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 231–242, 1998.

[14] Bob Shaw. Light of other days. *Nebula Award Stories*, 2, 1970.

[15] Shubham Tulsiani, Richard Tucker, and Noah Snavely. Layer-structured 3d scene inference via view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 302–317, 2018.

[16] University of California Museum of Photography. Keystone-mast collection. http://ucr.emuseum.com/collectionoverview/3631.

[17] Wikipedia. Contact print. https://en.wikipedia.org/wiki/Contact_print.

[18] Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. SynSin: End-to-end view synthesis from a single image. In *CVPR*, 2020.

[19] Ke Colin Zheng, Alex Colburn, Aseem Agarwala, Maneesh Agrawala, David Salesin, Brian Curless, and Michael F. Cohen. Parallax photography: Creating 3d cinematic effects from stills. In *Proceedings of Graphics Interface 2009*, GI '09, pages 111–118, 2009.

[20] Ke Colin Zheng, Sing Bing Kang, Michael F Cohen, and Richard Szeliski. Layered depth panoramas. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[21] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. In *SIGGRAPH*, 2018.

[22] C. Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, 2004.