EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2021)
A. Chalmers and V. Hulusic (Editors)

Short Paper

Reconstructing Dura-Europos From Sparse Photo Collections Using Deep Contour Extraction

Yifei Shen[1], Zeyu Wang[1], Qinying Sun[1], Anne Chen[1] and Holly Rushmeier[1]

[1]Yale University

Abstract

In this short paper we present work in progress on creating tools to facilitate 3D reconstruction of cultural heritage. We propose three new types of tools to make reconstruction easier – first we fetch linked open data to help organize source materials, next we extract key contours from photographs to speed up reconstruction, and finally we generate video tours of positioned photos and sketches. We also introduce a new, expanded 3D software system to support these tasks. The system is developed based on previous work on 3D sketching in the context of cultural heritage documentation, in particular CHER-ish. We demonstrate the potential of these tools by describing results obtained from the Dura-Europos data set.

CCS Concepts

• ***Human-centered computing*** → ***Visualization;*** • ***Applied computing*** → *Architecture (buildings);*

1. Introduction

A major challenge in the documentation and virtual reconstruction of cultural heritage sites comes from the sparse and highly heterogeneous nature of data. These data usually span a wide variety of sources: satellite photos, archaeological field photos, plans, hand drawings, illustrations, and metadata including labels and textual descriptions of photos. In many cases, the sparsity of source materials is compounded by spatial and temporal inconsistencies, making conventional computer vision techniques such as Structure from Motion (SfM) ineffective. With limited availability for new data and a diverse set of existing data of varying quality, how to make sense of and create an effective reconstruction out of the source material using a software system still remains a difficult problem.

Previous work in this area has proposed tools that allow the organization of historic photographs in 3D and modeling structural details by sketching [RLT*17]. However, such systems still present a few tedious aspects during the workflow:

- **Collecting and compiling image sets.** To correctly place historic photos in 3D, we must know which photos represent which structures. The Dura-Europos data set, for example, contains photos of major architectural structures as well as other details such as inscriptions and graffiti. Even among field photos, it is not immediately clear which images belong together because sites often feature more than one building with each structure photographed from multiple viewpoints. Often the metadata labels provide crucial information, but going through thousands of labels manually is a prohibitively lengthy and tedious process.
- **Inferring structural outlines from old images.** After placing historic photos in 3D, it is solely the user's responsibility to infer structural details and sketch them onto appropriate 3D scaffolds. Drawing all the details from scratch is both difficult and time-consuming. A semi-automatic process that can directly extract contours from the images is much more preferable.
- **Camera navigation and virtual touring.** To allow better visualization of the reconstruction result, a novel way of camera authoring is needed. Ideally this should include ways to create virtual video tours of the reconstructed structure in order for viewers to better comprehend the relationship between the already placed historic photos and the 3D structures.

In this paper we address these concerns with a new pipeline. First we pull data from open sources to facilitate organizing images into coherent sets. These image sets are subsequently fed into a machine learning model to generate predictions of structural outlines. Finally the predicted outlines are vectorized and imported into an integrated 3D software for the final reconstruction and camera authoring. We describe these steps in subsequent sections and present reconstruction results from the Dura-Europos data set in the end.

2. Related Work

Our work is mainly inspired by previous sketch- and image-based systems. Chen et al. [CMH*10] proposed a system that supports interactive sketching in 3D and locating, aligning, and organizing historic photos. Rudakova et al. [RLT*17] developed CHER-ish, whose main design involves placing virtual canvases in 3D, to which other 3D entities may be attached, including hand-drawn strokes, polygon fillers, and textured plane geometries (images). Their camera-work system allows the creation of "bookmarks", which encode camera poses that the user may return to at any time.

DOI: 10.2312/gch.20211408

Similar systems have been developed in the field of Architecture. Mental Canvas [DXS*07] provides a comprehensive 3D sketching toolkit and is primarily used in architectural prototyping and conceptual design. Its main idea has been extended in other systems such as Insitu [PKM*11] and Nested Explorative Maps [OCR*19]. Sinha et al. [SSS*08] also developed an architectural modeling system to reconstruct photorealistic, textured 3D models of urban scenes from photographs using techniques such as SfM.

Our idea of creating virtual tours out of placed historic photos is inspired by CHER-Ob [WSA*18]. This system includes an integrated video generator to produce introductory videos for cultural heritage projects, where the animation schemes are used to smoothly transition between annotations. Another source of inspiration is Photo Tourism [SSS06], which fetches publicly available images of tourist sites and sparsely reconstructs scene geometry using SfM techniques. Virtual tours can be generated by interpolating different views, evoking a powerful sense of immersion.

3. System Overview

3.1. Linked Open Data

Since the early 2000s, the online gazetteer Pleiades has geo-located, disambiguated, and catalogued changing naming traditions of significant ancient Mediterranean sites. In partnership with Pleiades, and as a test case for methodological extension to other sites of significance, the Yale Digital Dura-Europos Archive (YDEA) has recently worked to increase the granularity of Pleiades data at Dura-Europos so that it geo-spatially situates and disambiguates buildings within the archaeological site itself. Before searching for relevant images in a large archive, we first identified the relevant Pleiades gazetteer entry (Palmyrene Gate), and from it determined a list of all the keywords (i.e. alternate naming traditions) we would likely need to search. Specifically, these include "main gate" and "city walls".

A simple search into the Artstor digital library with "Dura-Europos" as keywords yields more than 30,000 images. These include photos of plans, hand-drawn illustrations, excavated artifacts, inscriptions, graffiti, and field photos of various extant structures within the city. Although the digital archives are well-structured and support advanced search, manual search queries are often unwieldy because of varying naming schemes, over-specification and under-specification of keywords, and the somewhat inconsistent locations where important information appears among the metadata labels. Therefore, we have created a catalog that reorganizes the images more consistently by subject and location. This dramatically reduces the tedium of compiling an image set for the Palmyrene Gate. Note also that once a catalog is created of a particular site, it can be reused for the reconstruction of any structure within the site.

3.2. Contour Extraction

A contour extractor automatically infers main structural outlines from field photos. Conventional edge detection algorithms fail because they cannot tell the difference between texture and structure, and their output contains excessive noise that greatly confuses the goal of contour extraction (Figure 1). The problem is exacerbated with older historic photos due to poorer image quality.

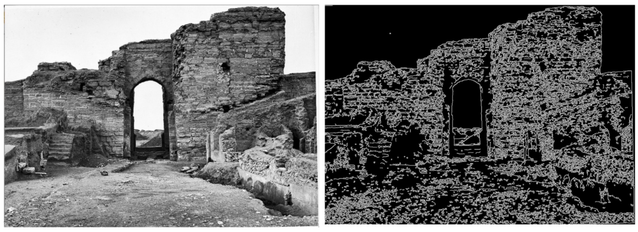

Figure 1: *An example input image of the Palmyrene Gate (left) and its Canny edge detection output (right).*

Figure 2: *From left to right: input from test set, ground truth, Photo-Sketch pretrained model output, refined model output.*

We have decided to solve this problem by machine learning and adapted the model from Photo-Sketching [LLMe*19]. We reused the generator and discriminator architectures but refined the weights by training the model on top of the pretrained Photo-Sketch model. Our own data set consists of 89 annotated images drawn from the Dura-Europos field photos, among which 63 are randomly chosen as the training set and the rest as the validation set. The subject matters of the 89 images are chosen to be diverse to ensure that the trained model is not biased in favor of any specific type of shape or structure. The annotations are tracings of the field photos where participants are explicitly instructed to ignore textural details and focus on structural outlines. Given a particular image, different participants generally produce slightly different tracings, which makes the Photo-Sketch model an especially appropriate choice because its loss function accommodates the presence of multiple diverse example outputs for a single input image.

After 600 epochs of training, we were able to refine the weights and obtain a new model that can generate plausible output given any field photo of Dura-Europos. Note that compared to the Photo-Sketch pretrained model, our refined model produces a lot more structural details while also managing to omit the textural details that plague outputs from conventional edge detectors (Figure 2).

Because the outputs of the contour extractor are bitmaps, we need to vectorize them before they can be imported directly into a stroke-based 3D sketching environment. To this end, we used an off-the-shelf line drawing vectorization algorithm [BS19] to convert the bitmaps into SVG format (Figure 3). After this conversion is complete, the contours are ready to be imported into 3D.

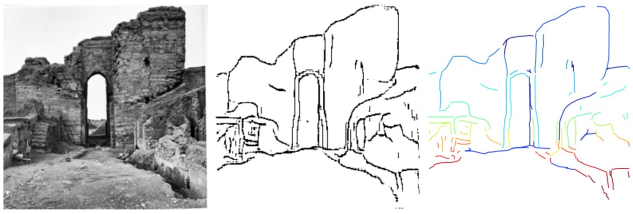

Figure 3: *Raw image (left) is fed into the contour extractor to produce a bitmap (middle), which is then vectorized (right).*

3.3. Reconstruction in Integrated 3D Environment

Based on the main ideas of CHER-ish, we created a new web-based software package. The 3D features are implemented using WebGL and three.js, and the interface is written with Vue (Figure 4). Our system also relies on strokes and images, which must live on 3D scaffolds. However, besides virtual canvases (infinite planes), our system also supports curved surfaces, a few geometric primitives, and custom OBJ shapes for convenient scene assembly.

After the necessary scaffolds are placed, the images and SVG files are imported. Polylines in SVG files are automatically converted into strokes on a plane. The user can then group the image with the strokes and use a 3D transformation widget to place them. This step of the process can be time-consuming because the user is essentially performing bundle adjustment. Although a robust SfM pipeline would provide considerable speedup, the input domain of our system is mostly historic photos, and our experiments with existing SfM algorithms failed to yield good results due to image quality, lack of feature points, and spatial/temporal inconsistency. On the other hand, the user possesses existing knowledge of the overall structure of the site and thus enjoys an inherent advantage over SfM techniques that only rely on the feature points. Therefore, we currently defer image placement to the user.

With images and their corresponding SVGs in place, the user can project the SVG strokes onto the 3D scaffolds to quickly generate a sketch-based reconstruction. Because relying exclusively on projection does not guarantee the most satisfactory results, the user can take advantage of our 3D drawing toolkit to manually edit the end result by erasing extraneous strokes or adding more strokes as they see fit. Finally, a polygon filling option is available to create occluding surfaces to give the reconstruction a more natural look.

3.4. Camera Navigation and Video Tours

We have implemented a new camera-work authoring system based on camera bookmarks. The user can create animation sequences with at least two (possibly non-unique) bookmarks. The order in which the bookmarks are selected determines in which order the animation will be played.

When an animation sequence is played, a camera path is generated by calculating a closed-loop centripetal Catmull-Rom spline that starts at the location of the first camera bookmark in the sequence, goes through each bookmark in order, and loops back to the first bookmark. For each segment of the spline curve between

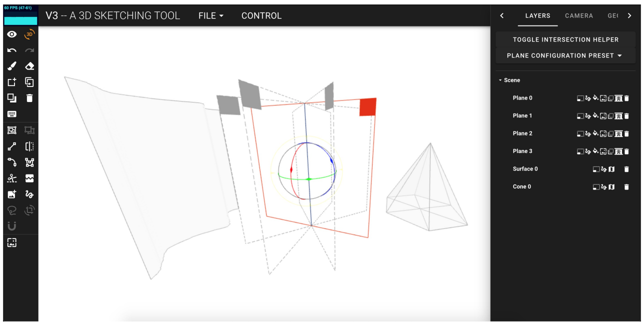

Figure 4: *Interface of our 3D software, showing a custom curved surface, a co-axial plane preset, and a cone.*

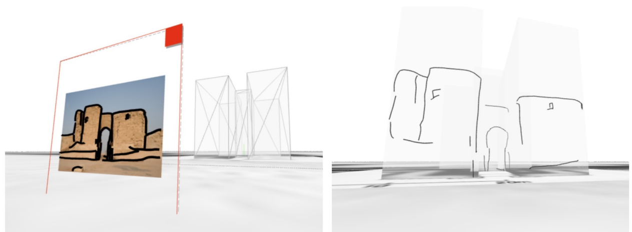

Figure 5: *Left: an image of the facade and its contours, with the 3D scaffolds placed. Right: contours are projected onto scaffolds.*

two successive bookmarks, an interpolation is calculated based on time. While the camera's position travels along the spline, the camera's viewing direction is interpolated from the two viewing direction vectors at both ends of the segment. To eliminate the extra degree of freedom in the camera's z-axis, we explicitly make sure that the cross-product of the "up" vector and the viewing direction is horizontal (i.e., camera pitch and yaw are allowed, but roll is forbidden). This is not an unreasonable constraint on camera movement because the camera controls in our 3D software do not allow rolling, so it is impossible for the user to create a bookmark whose camera parameters cannot be replicated without rolling the camera.

As a further enhancement to the virtual touring system, image planes can optionally be linked with camera bookmarks. An image plane can be linked with multiple camera bookmarks, but a camera bookmark can be linked with at most one image plane. When an animation sequence plays, an alpha fading effect is added to all linked images. When the camera leaves a bookmark, the image linked with that bookmark will begin to fade out, whereas the image linked with the next bookmark will begin to fade in. All other images along the animation path are fully transparent unless they belong to a bookmark that the camera has just left or is about to reach. Images not linked with any bookmark in the animation sequence are not affected by alpha fading.

4. Results

We demonstrate the potential of this new system through a case study on Dura-Europos with particular focus on the Palmyrene Gate. The asset catalog we developed quickly narrowed down the

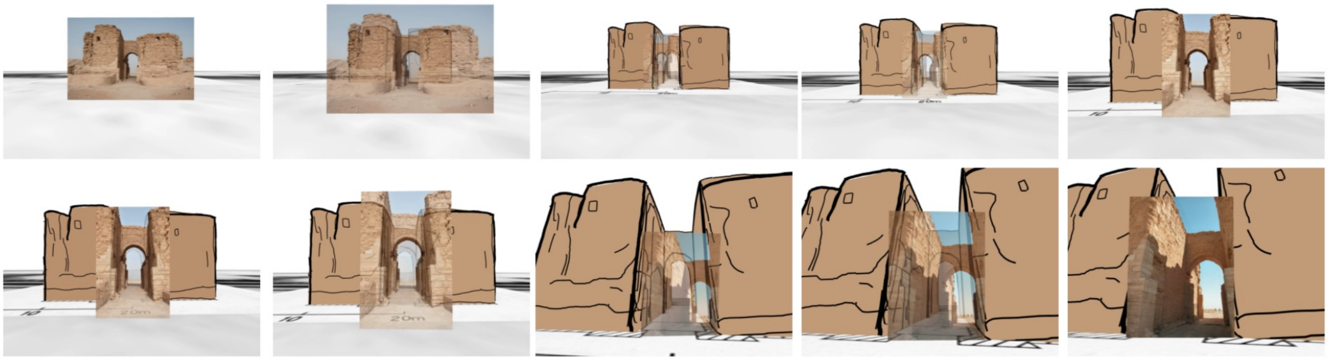

Figure 6: *Left to right, top to bottom: a sequence of sparsely sampled frames from the video tour with alpha-fading effect enabled (responsible for "ghosting" effect in intermediate frames). Full video is included in the supplementary material.*

whole set of 15,030 images to 220 by eliminating those whose metadata labels are unrelated to "Palmyrene Gate". Among the 220 images, we only kept field photos of the gate itself and removed images of inscriptions, graffiti, and nearby structures. In addition to field photos, we also looked for plans of the entire Dura-Europos site and plans of the Palmyrene Gate. The former is helpful for situating the gate in its proper context, and the latter is helpful for placing 3D scaffolds more precisely. We used Pleiades and YDEA to geo-locate the Palmyrene Gate within Dura-Europos with exact longitude and latitude numbers. In addition to images available from Artstor, we incorporated data from the Manar al-Athar image archive, which contains more recent photos of Dura-Europos.

After narrowing down the image set with the heuristics described above, we selected 15 images in total. Nine of them offered good views of the facade and two towers of the gate, and were fed through the contour extraction/line drawing vectorization pipeline to generate contours. These nine images/contours were then imported into 3D to assist reconstruction. The basic geometry was established with three rectangular parallelepipeds, which serve as 3D scaffolds that receive projected strokes (Figure 5). After the main structural outlines were formed via projection, some editing was performed to erase extraneous strokes, emphasize certain shapes, and add polygon fillers as occluding surfaces. Finally, we imported the remaining six images and placed them accordingly. Camera bookmarks were created at the same time as the images were placed in order to generate a final animation (Figure 6). The placement of images took two minutes per image on average, while editing of the projected strokes took around 15 minutes but could take shorter or longer depending on the user's preference.

5. Conclusion

In this paper we have proposed a new set of tools to facilitate virtual reconstruction of cultural heritage–fetching and organizing source material from linked open data, deep contour extraction, and virtual video tours. Despite poor data consistency and the consequent ineffectiveness of photogrammetric techniques, preconditioning SfM based on existing knowledge of structural details can be an exciting direction for future work. We will also expand our software by introducing more relevant tools, such as point clouds and surface reconstruction. Stroke projection and camera animation through bookmarks can potentially become plugins to existing systems such as Maya and Blender. We hope that our work will inspire further efforts in creating systems that empower artists, archaeologists, and computer scientists interested in documenting cultural heritage.

References

[BS19] Bessmeltsev M., Solomon J.: Vectorization of Line Drawings via Polyvector Fields. *ACM Trans. Graph. 38*, 1 (Jan. 2019). 2

[CMH*10] Chen X., Morvan Y., He Y., Dorsey J., Rushmeier H.: An integrated image and sketching environment for archaeological sites. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops* (2010), pp. 35–42. 1

[DXS*07] Dorsey J., Xu S., Smedresman G., Rushmeier H., McMillan L.: The Mental Canvas: A Tool for Conceptual Architectural Design and Analysis. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)* (2007), pp. 201–210. 2

[LLMe*19] Li M., Lin Z., Mˇ ech R., Yumer E., Ramanan D.: Photo-Sketching: Inferring Contour Drawings from Images. In *WACV* (2019). 2

[OCR*19] Olivier P., Chabrier R., Rohmer D., De Thoisy E., Cani M.-P.: Nested Explorative Maps: A new 3D canvas for conceptual design in architecture. *Computers & Graphics 82* (2019), 203–213. 2

[PKM*11] Paczkowski P., Kim M. H., Morvan Y., Dorsey J., Rushmeier H. E., O'Sullivan C.: Insitu: Sketching Architectural Designs in Context. *ACM Trans. Graph. 30*, 6 (2011), 182. 2

[RLT*17] Rudakova V., Lin N., Trayan N., Sezgin T. M., Dorsey J., Rushmeier H. E.: CHER-ish: A Sketch- and Image-based System for 3D Representation and Documentation of Cultural Heritage Sites. In *GCH* (2017), pp. 195–199. 1

[SSS06] Snavely N., Seitz S. M., Szeliski R.: Photo Tourism: Exploring Photo Collections in 3D. In *ACM siggraph 2006 papers*. 2006, pp. 835–846. 2

[SSS*08] Sinha S. N., Steedly D., Szeliski R., Agrawala M., Pollefeys M.: Interactive 3D Architectural Modeling from Unordered Photo Collections. *ACM Trans. Graph. 27*, 5 (2008), 159. 2

[WSA*18] Wang Z., Shi W., Akoglu K., Kotoula E., Yang Y., Rushmeier H.: CHER-Ob: A Tool for Shared Analysis and Video Dissemination. *J. Comput. Cult. Herit. 11*, 4 (Nov. 2018). 2