EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2017)
T. Schreck and T. Weyrich (Editors)

CHER-ish: A sketch- and image-based system for 3D representation and documentation of cultural heritage sites

Victoria Rudakova[1], Nathan Lin[1], Natallia Trayan[1], Tevfik Metin Sezgin[2], Julie Dorsey[1] and Holly Rushmeier [1]

[1]Yale University, USA [2]Koç University, Turkey

Abstract

We present a work-in-progress report on a sketch- and image-based software called "CHER-ish" designed to help make sense of the cultural heritage data associated with sites within 3D space. The software is based on the previous work done in the domain of 3D sketching for conceptual architectural design, i.e., the system which allows user to visualize urban structures by a set of strokes located in virtual planes in 3D space. In order to interpret and infer the structure of a given cultural heritage site, we use a mix of data such as site photographs and floor plans, and then we allow user to manually locate the available photographs and their corresponding camera positions within 3D space. With the photographs' camera positions placed in 3D, the user defines a scene's 3D structure by the means of stokes and other simple 2D geometric entities. We introduce the main system components: virtual planes (canvases), 2D entities (strokes, line segments, photos, polygons) and provide a description of the methods that allow the user to interact with them within the system to create a scene representation. Finally, we demonstrate the usage of the system on two different data sets: a collection of photographs and drawings from Dura-Europos, and drawings and plans from Horace Walpole's Strawberry Hill villa.

CCS Concepts

•***Human-centered computing*** → *Scientific visualization; Information visualization;* •***Applied computing*** → *Archaeology; Computer-aided design;*

1. Introduction

A challenge in cultural heritage is making sense of the data associated with a site. Diverse types of visual data may be available for a site –such as images, illustrations, maps and drawings. Images may include a range including historic photographs, modern tourist snapshots and satellite imagery. Illustrations range from artists impressions to measured drawings. Maps range from surveys of past excavations to modern Google or Bing maps. The goal of the system described here, CHER-ish (**C**ultural **HER**itage - **i**nformation for **s**ite **h**istory), is to assist a user by placing all of the available visual data and its interpretation into a single 3D virtual space.

Previous work on this problem was described by Chen et al. [CMH*10]. The system was based on a sketching design system that used strokes conveyed by "canvases" in 3D space [DXS*07, PKM*11]. Chen et al. augmented the sketching system with images positioned in the same 3D space. Modern, spatially dense images can be positioned with respect to one another using computer vision techniques as described by Snavely et al. [SSS06]. For historic photographs and drawings computer vision techniques fail for a variety of reasons – inadequate number of images, poor image quality, etc. Chen et al. proposed a technique for estimating the relative positioning of images where automated techniques failed. The technique requires the user to make estimates of the ground plane and horizon in each image, and to find common edges to position image canvases in 3D. The technique works for photographs with visible rectangular structures.

One aspect of sense-making for 3D sites is creating 3D reconstructions of structures that no longer exist. There are many sources of data, but they are spatially sparse and temporally inconsistent, making automatic methods ineffective. The reconstructions must be a human interpretation, and documentation is required to trace how the 3D was defined based on the various pieces of 2D visual evidence. Researchers have developed techniques to connect evidence with reconstructions. For example, Bale et al. [BAG*11] created a collaborative annotation system to attach original source material to a 3D reconstruction of the no longer existent 1938 British Empire Exhibition. These annotations are critical for subsequent users to understand the accuracy of the model.

The system described by Chen et al. demonstrated how a 3D reconstruction could be developed with a direct connection to 2D visual evidence. Early work by Debevec et al. [DTM96] showed how photographs could be used in an interactive application to create 3D models. Inspired by this, the Chen et al. system allows images and drawings to be positioned, and then strokes drawn on the images are used to create the 3D representation. The result is a reconstruction that has the 2D evidence embedded in the 3D space. Further, the 3D representation is clearly an interpretation as a result of the sketch-like appearance resulting from the strokes. Incorporating evidence as part of the process of generating a reconstruction avoids accidental omission of information in adding annotations afterwards. It also allows future users of the model to more easily assess the accuracy of the interpretation.

The system presented by Chen et al. used older technology, and included features more suited for creative design than data organization. The technique for positioning images proved to be cumbersome, and not general enough for arbitrary images and drawings that are not strict perspective renderings. In this work-in-progress, we have developed a new software system from scratch. We have used two very different test cases to exercise the system, and developed new interaction techniques to facilitate organization and reconstruction.

2. System overview

CHER-ish is implemented in C++ using Qt and OpenSceneGraph (`https://vicrucann.github.io/cherish/`). The basic concept follows Chen et al. and here we give just a brief description. A more detailed view of the system is available in "Beginner-Guide.pdf" included as supplementary material.

2.1. Canvas

The primary means of representing a site is through sketching and other 2D drawing procedures on a virtual plane. We define a *canvas* to represent a plane in 3D space. Since each canvas is only a means to perform 2D drawing and, in theory, is infinite in its extent, we show it as a finite bounding rectangle which embraces all the contained elements (e.g., strokes, polygons, etc.). This representation makes it easy to see the canvas orientation in 3D. The canvas can be positioned and rotated in 3D space using traditional graphics conventions. An initial view of the 3D space with a single canvas is shown in Fig. 1.

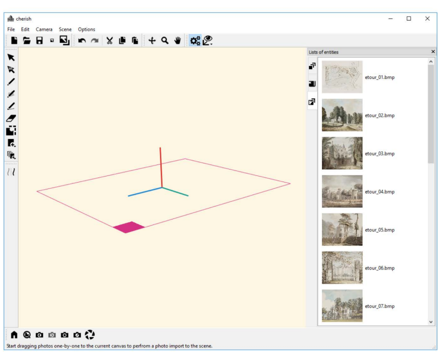

Figure 1: *An initial view with a single canvas in the 3D space.*

To denote what canvas the user is about to draw on, each canvas has a state: *current*, *previous* or the *rest*. The state is shown by the color of canvas bounding rectangle, e.g., magenta - for the current, purple - for the previous and gray for the rest of canvases. Each canvas has a *pickable area* -i.e., when clicking on it with the mouse, the canvas gets selected and becomes *current*, while the previously selected canvas changes its state to *previous*, and the previously previous canvas takes a status of the *rest*.

Elements such as 2D strokes are added to a canvas using a stylus on a tablet or screen. By using a mouse to rotate, pan and zoom the camera position, the user is able to draw on the canvas from any camera position, i.e., the camera does not have to be strictly perpendicular to the canvas as the user sketches. The pen position is automatically projected onto the appropriate point on the canvas surface.

2.2. 2D Entities

Any element that can be drawn within a given canvas is called a 2D *entity*. The types of entities are: stroke, line segment, photo (textured rectangle) and polygon. The type of entity to use depends on the type of the structure, e.g., a stroke would be best to define weathered and partial structures , a line segment would be best when reconstructing rigid structures with clearly defined features, the photo entity serves as documentation and the polygon entity can help to create an occluding surface. Occluding surfaces are useful for avoiding the look of a wired structure that becomes confusing in complex scenes. Photo entities are assigned a transparency level, to allow the user to see through the photo when positioning it relative to strokes that have been positioned to represent a 3D structure.

Once some entities are introduced on a canvas, it is possible to perform entity selection and manipulation. Once a set of one or more entities is selected, a modification tool allows the set to be rotated, scaled and positioned. The selected set can also be copied or cut from the canvas and pasted within the same or different canvas. As in [DXS*07] strokes can be "pushed' from one canvas to another taking into account perspective distortion. Most of the modifications for the entities and canvases are incorporated into undo-redo framework.

2.3. Bookmark

A *bookmark* is defined as a saved camera position with fixed camera parameters (e.g., focal length) with an attached state of the scene graph (e.g., visibility of each of the canvas and transparency levels of each of the photographs, etc.). Like any type of 3D graphics system, the software supports the change of the current camera position by means of zoom, rotation and pan. The bookmarks are particularly useful for using photographs to introduce changes by means of drawing into the 3D model. For example, when a user is working with the scene using two photographs, it is possible to change views from one semi-transparent photo to another in order to verify the scene structure from the both views.

There are two ways to create a bookmark: first is to take a snapshot of the current camera position, and the second is to manually place a camera within the current canvas. The second way proves to be especially useful for datasets when not only the ground plans are provided, but also the approximate camera position and its corresponding field of view are known or easily inferred by the user. After the camera is posed using the manual method, the user is offered the option to create a canvas perpendicular to the camera "look at" direction, and then to drop in a photo which corresponds to that particular view. By doing so, the whole procedure now allows the user to introduce more details into the 3D scene by using the dropped photo as a documentation reference.

3. System usage

3.1. Users

Rather than develop the system and then do a user study, throughout the development of the system users with different backgrounds were asked to use the system and provide feedback. Users included cultural heritage professionals and computer scientists. From cultural heritage professionals the main feedback was that far more guidance was needed for the purpose and usage of the system, and this led to more extensive documentation (including the BeginnersGuide.pdf). From the computer scientists came the practice of locating and specifying cameras on the ground plane (rather than trying to mark up correspondences between images and strokes already in the scene).

Two users took part in testing the system once it became mature, and built demonstration models: a user with a technical (computer science) background, and a user with artistic background. Both of the users were asked to test the system by creating different scenes of different scale complexity - starting from a small scene which is supported by one photograph, e.g., a room, and finishing by compilation of dozens of photographs as support for one model. Both of the users were able to comprehend the concepts used in CHER-ish and produce a 3D reconstruction of fair complexity.

3.2. Site Types

Two vastly different heritage sites were used for testing. One is the Dura Europos site used by Chen et al. The visual data set includes plans, vintage and modern photographs and diagrams. The site includes natural features and irregular (i.e. not modern or rectangular) structures. The other site is Horace Walpole's eighteenth century Strawberry Hill villa. The visual data set includes water color paintings and a site plan with approximate positions of the views shown in the paintings. The collection of paintings includes both interior and exterior views.

3.3. Typical workflow

A typical workflow evolved as users interacted with the system.The users found that the 3D reconstruction process relies heavily on the availability of floor or ground plans. Importing the floor or ground plan is the first natural step in scene reconstruction. The plan helps to identify the approximate location of structures and create a first draft of the site. While this may seem to be a difficult requirement, in practice there is generally some Google or Bing map to work with and associated satellite image data where some structure can be located. Users find this more logical than relating images pairwise in the way computer vision algorithms relate images to compute 3D locations.

The next step involves going through other documentation and photographs of the site, one-by-one, and trying to embed them within the 3D space as if they were taken from their corresponding camera positions. While for some datasets like Strawberry Hill, we had the information on the approximate camera positions for each of the drawings, it is not the case for most of the other datasets like Dura-Europos. Therefore, the main difficulty was to use the manual camera positioning tool and find the camera position relative to the floor plan. Note that the initial guess for camera position may be incorrect, but the position can be adjusted (moved, rotated or tilted) as needed.

Once the approximate camera position is found for the first photograph or drawing in relation to the 3D scene (where the floor plan is already embedded), the user starts to define the model's structural elements based on the details from the photograph. Since the located camera position is saved as a bookmark, the user can always go back to it, and introduce changes in the scene from the same camera position perspective. The addition of scene details is often done by tracing over the photograph, and carrying the traced data over to the scene in the form of strokes, line segments and polygons which are projected on the corresponding canvases within the 3D scene. The example result of this step is shown in Fig. 2 when the user introduced one drawing of Strawberry Hill site and recovered the height of a particular wall. Figure 3 shows a final result when several more pictures were added, and with their help the whole building structure was recovered.

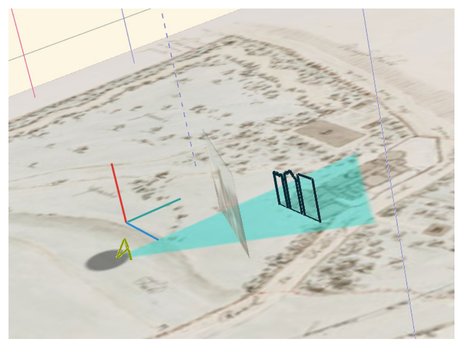

Figure 2: *Defining initial structure using bookmark functionality and tracing over the painting.*

Figure 3: *Building defined using manual camera positioning.*

3.4. Results and Challenges

Figures 4 and 5 show screenshots of reconstructions of Dura-Europos architectural structures, and Fig. 6 shows the Strawberry

Hill upper floor interior reconstruction. All the reconstructions are based on the photographic and/or painting data. Gifs are included as supplemental material that show more views of these examples.

Since CHER-ish does not rely on computer vision algorithms for inferring the perspective transformation between the reconstructed 3D model and an image, one of the main challenges is the manual positioning of the photos and finding their corresponding camera parameters. While CHER-ish is based on finding the camera positions and doing the whole reconstruction *approximately*, the current camera positioning workflow still involves substantial readjustment.

Another notable issue which occurred mainly when working with Strawberry Hill data is the distortion of perspective parameters which has an artistic (not optical) nature. Most of the Strawberry Hill data are paintings and drawings produced by different artists two centuries ago. After trying to perform the reconstruction using indoor and outdoor views, we found out that perspective in the sources is often not geometrically correct, e.g., by trying to locate a vanishing point in one-point perspective drawing and realizing the lines perpendicular to the picture plane do not converge strictly at the same point. One of the ways to deal with such illustrative images was to use the image as a visual reference but not to use it for direct tracing and projection onto the canvases of 3D scene. Therefore, the whole interior dataset of Strawberry Hill was drawn using the image-as-reference method.

Re-creating a non-flat surface presented another difficulty when dealing with outdoor scenes which typically contained buildings. Since the basic surface of CHER-ish is a canvas, it became a challenge to represent non-flat surfaces, e.g., cylindrical, spherical, or uneven walls of building. Of course, there are different ways one can approximate those surfaces using a set of canvases discretely placed within the 3D space, however it can become tedious when working with large-scale and non-trivial surface scenes. For the most of such surfaces within the tested datasets, we found the approximation can produce visually satisfying results - since the whole concept of CHER-ish is based on approximation and allows artistic interpretation to obtain a complete result.

For the larger scale scenes, it can become difficult to interact with the scene elements using the current user interface, i.e., when the complexity of the scene grows, the number of canvases becomes larger accordingly. This leads to difficulties in navigating through the large number of canvases and selecting a canvas. One way to overcome this difficulty is to give a textual description as a name to each canvas. However, working with very large scenes (50+ canvases) is still difficult, since there is currently no way to re-organize the order of the canvases or sort them.

4. Conclusion

Our goal is to develop a system that helps to make 3D sense of cultural heritage data using photographs and other visuals, e.g., paintings, drawings and floor plans. For this purpose, we use previously developed 3D sketching techniques and expanded them in order to embed the available data by manually positioning it within the system.

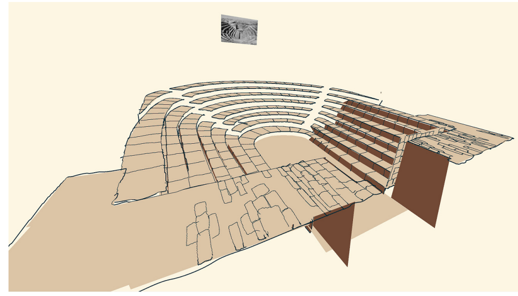

Figure 4: *Dura-Europos amphitheater reconstruction.*

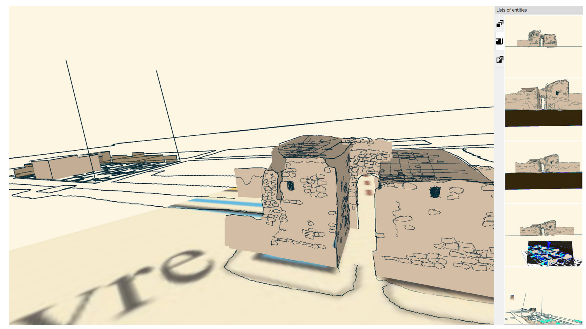

Figure 5: *Dura-Europos main gate reconstruction.*

For future work, the process of camera positioning can be improved by developing methods for approximating and refining camera position and projection parameters based on user-provided information. For example, it would be possible to extract a homography matrix from the four point correspondence between an area of a photograph and an area of a floor plan. The method however needs to be simpler than the method described in Chen et al., and should take advantage of the context provided by having a ground plan and initial simple 3D models. Another future improvement is to allow organizing canvases into groups (e.g., folders), and to sort their order within the canvas widget. This would be especially useful in a function to merge multiple sub-scenes into one big scene, i.e., a scene import functionality.

This material is based upon work supported by the US National Science Foundation under Grant No. 1302267.

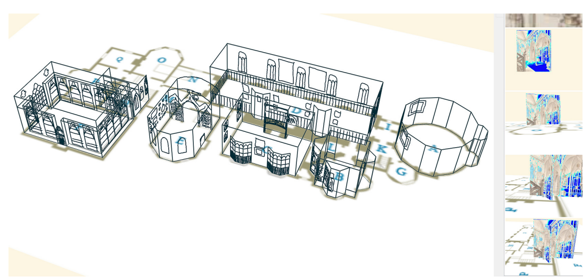

Figure 6: *Strawberry Hill interior reconstruction.*

References

[BAG*11] BALE K., ABBOTT D., GOWIGATI R., PRITCHARD D., CHAPMAN P.: Linking evidence with heritage visualization using a large scale collaborative interface. In *Proceedings of the 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST* (2011), Eurographics Association, pp. 121–128. 1

[CMH*10] CHEN X., MORVAN Y., HE Y., DORSEY J., RUSHMEIER H.: An integrated image and sketching environment for archaeological sites. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* (2010), IEEE, pp. 35–42. 1

[DTM96] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 11–20. 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 *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on* (2007), IEEE, pp. 201–210. 1, 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), 182–1. 1

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)* (2006), vol. 25, ACM, pp. 835–846. 1