Mem-Pano: Constructing Individualized Google Streetview Panoramas

Ada Zhou†
Stanford University
Stanford, CA
94305, USA
zhouc@stanford.edu

Michelle Gan
Stanford University
Stanford, CA
94305, USA mailto:
mgan@stanford.edu

Jennifer Luo
Stanford University
Stanford, CA
94305, USA
luojen@stanford.edu

Nicholas LaRosa
Stanford University
Stanford, CA
94305, USA
nclarosa@stanford.edu

ABSTRACT

Current digital representations of locations in the world exist as photographs or panoramas, such as in Google Street View. To transport users to locations around the world, virtual tours utilize the 360-degree panoramic format, however, panoramic representations cannot fully capture how an individual experiences or remembers the location. We present Mem-Pano, a panorama-editing system that allows novice photo editors to make a series of panoramic image edits to reconstruct a location to better reflect their own personal memories and experience of a location. Through a two-part initial study of users creating their individualized panoramas and viewers experienced individualized panoramas we find that (1) Mem-Pano allowed users to create panoramas that better represented their perspectives and memories of location and (2) individualized panoramas provided distinct stories of the represented location to viewers but were confusing without context.

KEYWORDS

Image editing, object relation, object insertion, object resizing, object replacement, scene reconstruction, panoramas

INTRODUCTION

In 'The Image of The City', Kevin Lynch defines the city as a form of temporal art. Much like sculptures, cities are spatial structures, but unlike sculptures, "cities undergo change at every moment." [1] As spaces evolve over time, the gap between digital representations of a specific location and individual experiences of a specific location widens. Existing digital representations of given locations currently take the form of images, virtual reality reconstructions, or Google Street View 360-degree panoramas. [2] [3] Images allow users to convey specific details relevant to a given location, but do not allow users to synthesize these details into a complete portrait of a location in which users can model the relationships between these specific details. Virtual reality reconstructions allow creators to mold and texture places flexibly beyond real-world representations of a location, however, this is resource intensive, inaccessible to non-technical individuals and thus not widespread. Panoramic representations of locations may prove more accessible for users, however, panoramas only reflect a specific location at a given moment in time, and no current tools exist to simultaneously modify and view panoramas. Photoshop poses the closest tool to easily modifying panoramas, yet Photoshop comes with a steep learning curve, making it inaccessible for novice users.

We present Mem-Pano, a panorama-editing system that allows novice photo editors to make a series of panoramic image edits to reconstruct a location to better reflect their own personal memories and experience of a location. Users provide a location and images as details for a panorama edit. Our system allows users to make low-level panorama-edit operations by replacing, resizing, and removing objects in their panorama. Users can view this panorama in a 360-degree panorama viewer integrated into our system. For individuals who have lived in locations that have experienced drastic change, including natural disasters, gentrification, and displacement, Mem-Pano works with novice users, utilizing low-level panorama alterations, to bridge the gap between an inaccessible past and accessible present, resulting in a digital representation of a location that closely reflects individual experiences of a given location. Mem-Pano makes edits accessible and integrates detailed user-provided images into a 360-degree panorama viewer to allow novice users to seamlessly edit and experience locations significant to them. We aim to allow novice users to create personal artifacts and individually guided tours to further understanding of the history at a given location.

We perform a preliminary evaluation of our system, asking 6 novice users to utilize Mem-Pano to make panoramic edits to a location of significance, and a follow-up study asking 12 novice users to evaluate the panoramic edits.

RELATED WORK

Prior work falls under three umbrellas: (1) refinement of replicating physical locations, (2) creating an accurate representation of an individual's spatial models and (3) image editing that empowers novice users. Prior research, as in The Photo Tourism Project, utilized collections of user photos to create 3-D models of scenes [4]. This work creates the foundation for stitching together photos into panoramas. Prior research has toyed with different mediums for representing models, such as maps in the Automatic Generation of Tourist Maps, which automatically generates a simplified 2-D map attuned to the mental model of tourists navigating cities [5], and Memory through Design, which used a 2-D mapping tool for their users to record their mental models on home place making and created 3-

D physical models based on the user's 2-D work [6]. Prior research has also sought to empower novice users, as with Eevee, a system built to empower novice users in the task of editing images [7]. We build upon Eevee, a simple image-editing system, to allow users to transform images into the panorama space for the immersive experience. Our edits are inspired by the model for manipulating certain elements in an image for different perspectives presented by ZoomShop: Depth-Aware Editing of Photographic Composition [8].

MEM-PANO

Mem-Pano, leveraging Google Street View maps located at user-specified coordinates and building on prior work constructing low-level image edits from high-level goals, transforms an existing Google Street View panorama.

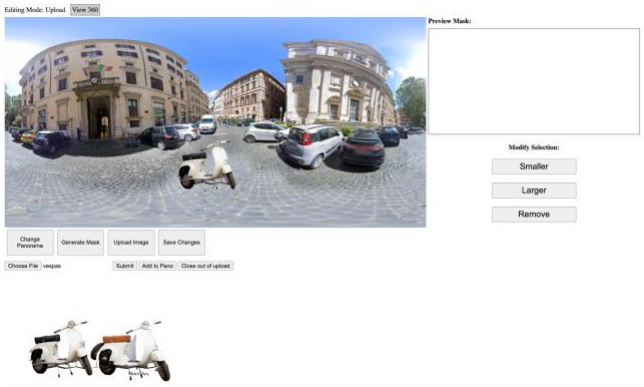


Figure 1: Edited panorama with a Vespa inserted.

Mem-Pano asks users to provide (1) coordinates of a specific geographic location and (2) personal images, or hand-drawn sketches, of a specific geographic location. Based on user-provided coordinates, Mem-Pano loads the existing Google Street View panorama at the location. Using a direct manipulation interface, the user then specifies high-level edits to perform on the image on specific landmarks or objects in the image. They are able to adjust the size of an object, adjust the existence of a particular object, and replace a given object with any user-provided images. Mem-Pano will perform the edits, render a 360-degree panorama, and re-upload the edited panorama to a 360-degree panorama viewer. The user can then use the 360-degree panorama viewer to explore the panorama at the specified location.

The goals of Mem-Pano are to:

- G1: Enable users to make a series of high-level panoramic image edits that reflect their mental unique perspectives of a location.
- G2: Enable viewers to view edited panoramas to gather a more distinctive narrative about a location.

WORKFLOW

Users provide the coordinates of a location of interest. Mem-Pano loads the Google Street View panorama at the specified coordinates.

Figure 2: Unedited Panorama, NYC Chinatown

Users are then able to make changes to the resulting panorama to better match their unique experiences of the location through a variety of high level editing options. These include 1) altering the size of a particular object 2) removing objects and 3) uploading their own images, all of which are done using image masking and editing.

Using Mem-Pano, users trace around the object of interest to generate a mask. Then, they can click "larger," "smaller," or "remove" buttons to make the necessary adjustments to this mask. Users can also upload their own image, outline an object of interest in their image, and add the object to the panorama. This object can be similarly adjusted using the labeled buttons. This allows users to add objects, like crowds, shops, animals, and remove other objects, like newer buildings or vehicles.

Figure 3: Edited Panorama (inserted crowd of parents), NYC Chinatown

After the user is satisfied with their edits, they click the "Save Changes" button. Our application then uploads the resulting image to Imgur, where any edited images are stored, and re-uploads it to the web application. The user can then view the pano in 360 degrees.

EVALUATION

Methodology

1.1 Study 1

We first performed a formative evaluation of our system with 6 participants aged 20-57 with limited experience in sketching and image editing. We aimed to evaluate the aforementioned system goal G1 and to better understand how users interact with our system.

For each participant, we asked them to provide a location of interest (either coordinates or an address) and find or sketch pictures of landmarks, people, and other details that pertain to their experience at the location. They scanned and uploaded these images. We then had the participant describe the location and its significance to them, taking note of features like buildings/notable landmarks, density, etc. Participants could supplement these oral descriptions using resources on the Internet, sketches, personal photos, etc. Then, the participant utilized our system to reconstruct a panoramic image that is more reflective of their unique experience of the location through the provision of high-level edits (eg. “remove this tree”, “make this building bigger”).

After completion, we evaluated how well our system addressed our project goals by asking participants to compare our system with other tools and to rate the following statements on a Likert scale (1-7):

- The system helped me transform the image to match my perspective of a location.
- The reconstructed image matched my expectations of the edits I suggested.
- The system was easy to use.
- I prefer the edited panorama over the normal panorama.

For Study 1, we hypothesized the following.

H1: Current digital ways of experiencing locations are not sufficient at representing individual perspectives and memories of locations.
H2: User-edited panoramas can better represent the user's individual perspective and memories of a location.

1.2 Study 2

Next, we conducted an analysis of the panoramas constructed by users in Study 1 with 12 participants, aged 19-26. All of the viewers had limited photo editing experience. We aim to evaluate the aforementioned system goal G2 and to better understand how users perceive panoramas that have been edited using Mem-Pano.

Each participant was shown one set of panoramic images, each consisting of a pair of edited and unedited panoramas created by participants in part one of the study. Given that we had 6 sets of panoramas and 12 participants in Study 2, each set of panoramas was viewed twice.

For the panorama viewings, we alternated the panorama that was viewed first by the user. For each set of panoramas, one user viewed the original panorama first, and another user viewed the edited panorama first. For each panorama they were shown, users were asked to answer two open-ended questions for each of the panoramas:

- What is your impression of this location based on the panoramas?
- Can you describe the setting and story depicted in this panorama?

They were also asked to respond to two Likert Scale questions (1-7) for each panorama:

- I have a good understanding of the location presented in the panorama..
- I understood a distinctive or unique perspective on this specific location from this panorama.

Finally, participants read the textual description provided by the creator of the panoramic images in Study 1. They then responded to the following questions:

- On a scale from 1 to 7, with 1 being strong disagree, and 7 being strongly disagree, please rate the following statements:
- The panorama matches the description for the original panorama.
- The panorama matches the description for the edited panorama.
- Which panorama contributed more to your understanding of the location? Why?
- Which panorama do you prefer? Why?

For Study 2, we hypothesized the following.

H3: Viewers prefer looking at user-edited panoramas because they can gather more distinctive narratives about a location.

Results

2.1 Study 1

H1: Current digital ways of experiencing locations are not sufficient at representing individual perspectives and memories of locations.

We confirm this hypothesis primarily based on users’ qualitative responses and how they chose to edit their panoramas. Users stated that the panorama matched what is physically at the locations they chose but described their experience and memories of the location differently. Coding the themes from user responses on what changes would create a representation that matched their memory or perspective, we found they fell under one of the following categories:

- Users wanted to add objects prominent to their associations of the place and remove objects that conflicted with their image of the place, even if not geographically or physically accurate.
- Users wanted to modify the panorama to match their emotions around the place (e.g. “more sunny”).

The concrete edits users leaned towards adding details to the panorama compared to removing. 6/6 participants added details to the panorama. Though we expected users to add in elements of their personal photos, they opted to add stock images to the panorama as many did not have personal photos with the elements they wanted to add. Only 2/6 participants opted to remove objects from pano. This is consistent with previous research showing people systematically overlook subtractive changes [9].

H2: User-edited panoramas can better represent the user's individual perspective and memories of a location.

Participants responded to the question "the system helped me transform the panorama to match my perspective of a location" with an average Likert scale rating of 5.67/7. In interviews, participants noted that they were pleased with the ability to reconstruct their memories of the location into one place. Participant 6 noted, "It centered memory as opposed to the objective truth [so] I was able to do a patchwork of my memory from things that maybe I didn't see together at one time but were all part of my associations of that place."
Participants also responded to the question "the reconstructed panorama matched my expectations of the edits I suggested" with an average Likert scale rating of 6.4/7. Accordingly, Participant 5 stated, "The panorama helps construct the picture I want. I want to restore some old memories I had there, and the panorama was useful for that." Participant 3 reacted with contentment to the edited panorama, "This is exactly what I imagined."

Statement	Mean Likert Scale Rating
The system helped me transform the image to match my perspective of a location.	5.67
The reconstructed image matched my expectations of the edits I suggested.	6.4
The system was easy to use.	5.1
I like the edited panorama over the normal panorama.	5.5

Figure 4: Table with average Likert Scale Ratings from Study 1.

2.2 Study 2

H3: Viewers prefer looking at user-edited panoramas because they can gather more distinctive narratives about a location.

We found that viewers had mixed responses to user-edited panoramas. First, the viewers who preferred the edited panorama saw elements in the panorama that matched elements in the user's description of the place. For example, Participant 5 preferred the edited panorama because "it has people and feels realistic, I can imagine being there and seeing those people [from the description]." Others valued the journey of uncovering the story about the location in the edited panorama. Participant 12 explained that "although the unedited panorama gives more context, I like the [process of] trying to figure out the background story of the edited panorama."

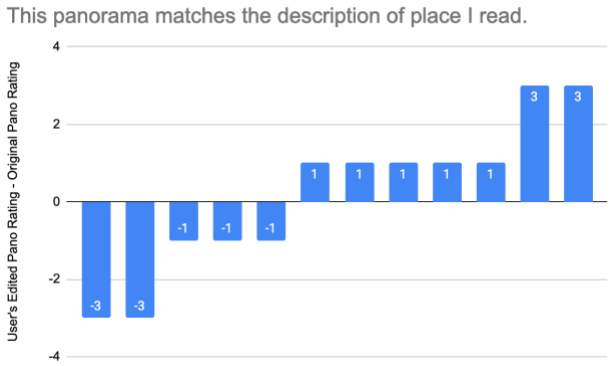
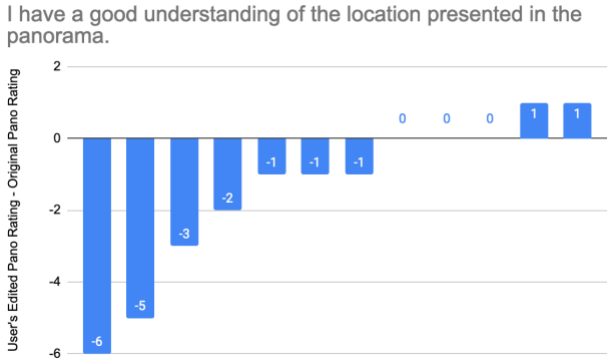


Figure 5: Original panorama ranking subtracted from the edited panorama ranking (higher scores are correlated with preferring the edited panorama).

Other viewers who preferred the original panorama were confused by edits that either weren't explicitly featured in a user's description of the location or didn't match their own projected expectation of the location. One such example of the latter is Participant 6's edited panorama of a neighborhood in Orlando, Florida. In Study 1, Participant 6 added an image of their mother and of a peacock, which they claimed "were everywhere" in their neighborhood. When Participant 11 viewed this panorama in Study 2, they expressed confusion over the peacocks present in the edited panorama. Participant 9 commented, "With this [panorama], I feel like I'm in a Gabriel García Marquéz novel because you don't everyday come across a peacock in the middle of the street, so it's a bit uncanny." Their assumption that peacocks wouldn't exist in this neighborhood influenced their perception of the panorama. They further noted that this called them to question the panorama editor's intentions, whether this panorama was "supposed to be real" or was "deliberately uncanny". This suggests that the experience of panorama viewing can be improved if the editor's intentions for the panorama are clearly communicated with the viewer.



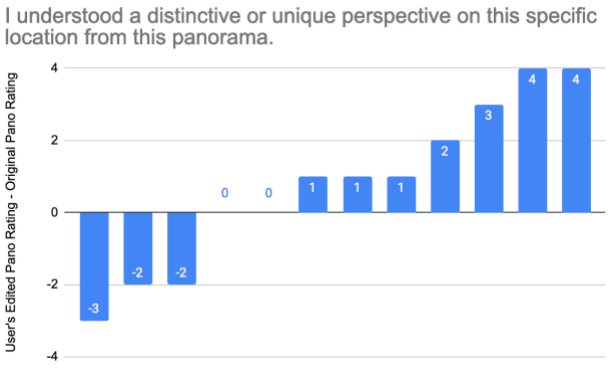


Figure 6 & 7: Original panorama ranking subtracted from the edited panorama ranking (higher scores are correlated with preferring the edited panorama over the original panorama).

A further analysis of the quantitative ratings provided by viewers for the original and edited panoramas demonstrates that many viewers did not feel like the edited panorama provided a strong understanding of the location compared to the original panorama. In interviews, they revealed that the poor graphic quality of these edited panoramas hindered its utility and understandability, especially in regards to removing objects. Participant 4 stated that "the [edited panorama] was kind of too chaotic for me and made me not want to look at it anymore." Still, 7/12 viewers agreed that the edited panorama provided a distinctive or unique perspective on the location compared to the original panorama.

LIMITATIONS

Our user studies suggest primary limitations of Mem-Pano are (1) the selection of the panorama location, (2) the graphic quality of edits, and (3) the types of edits a user could make to the panorama. We were limited by the panoramas available on Google Street View. There were several occasions when participants wanted to select certain locations but could not due to a lack of Street View data in the particular region. Users felt the graphic quality of edits made them less open to sharing their panoramas with others. For the study, users could only directly manipulate items and their sizes. We did not for example have access to more subtle alterations like color or opacity.

FUTURE WORK

In the future, we plan to investigate expanding editing options for Mem-Pano, including color manipulation, lighting manipulation, and automated broad image edits (e.g. removing all vehicles from a given image). We also plan to improve the graphic quality of panoramas by automating the process of ingesting Google Street View images and stitching together these image into 360-degree equirectangular images. Automating the process of curating equirectangular images will broaden our access to panoramas in regions that Google Street View fails to cover (e.g. expanding to incorporate panoramas from Baidu). We also plan to explore different methods for presenting panoramas alongside the context and stories they were developed within. We also plan to explore options for automating edit suggestions based on user oral and textual descriptions of locations leveraging natural language processing. Finally, we plan to investigate how Mem-Pano might automatically stitch together images and scenes sourced from Google user reviews to construct time-differentiated panoramas of a given location, such as to allow users to view how a given location has evolved over time.

We hope that expanding the capabilities of Mem-Pano can make it usable by urban studies scholars as well as open it up to more creative freedom from novice users. Current studies in collective memory and narratives of local history shape the ways people imagine a neighborhood's present situation and future development, but current avenues for uncovering these narratives are purely text-based [10]. Novice-created individualized panoramas would allow a visual format to understand the multiple narratives in locations. This could be particularly helpful in participatory community development approaches.

CONCLUSION

We have presented Mem-Pano, a panorama-editing system that allows novice photo editors to make a series of panoramic image edits to reconstruct a location to better reflect their own personal memories and experience of a location. Our preliminary evaluation shows that 1) Mem-Pano allowed users to create panoramas that better represented their perspectives and memories of location and (2) individualized panoramas provided distinct stories of the represented location to viewers but were confusing without context. This current system of Mem-Pano creates personal artifacts but our results show that user-edited panoramas have potential to uniquely help other users understand an individual's distinct perspective of a location.

REFERENCES

[1] Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: mapping the inequality of urban perception. PloS one, 8(7), e68400. https://doi.org/10.1371/journal.pone.0068400.

[2] Kwiatek, K. & Woolner, M. (2010). Transporting the Viewer Into a 360° heritage story: Panoramic interactive narrative presented on a wrap-around screen. 16th International Conference on Virtual Systems and Multimedia. https://doi.org/10.1109/VSMM.2010.5665980.

[3] Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the World at Street Level. Computer, 43(6). https://doi.org/10.1109/MC.2010.170.

[4] Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the World at Street Level. Computer, 43(6). https://doi.org/10.1109/MC.2010.170.

[5] Grabler, F., Agrawala, M., Sumner, R. W., & Pauly, M. (2008). Automatic generation of tourist maps. ACM Transactions on Graphics (TOG), 27(3), 1-11.

[6] Sabie, D., Sabie, S., & Ahmed, S. I. (2020, April). Memory through design: supporting cultural identity for immigrants through a paper-based home drafting tool. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[7] Michelle S. Lam, Grace B. Young, Catherine Y. Xu, Ranjay Krishna, and Michael S. Bernstein. 2019. Eevee: Transforming Images by Bridging High-level Goals and Low-level Edit Operations. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19). Association for Computing Machinery, New York, NY, USA, Paper LBW1513, 1–6. https://doi-org.stanford.idm.oclc.org/10.1145/3290607.3312929.

[8] Liu, S. J., Agrawala, M., DiVerdi, S., & Hertzmann, A. (2022). ZoomShop: Depth-Aware Editing of Photographic Composition.

[9] Adams, G.S., Converse, B.A., Hales, A.H. et al. People systematically overlook subtractive changes. Nature 592, 258–261 (2021). https://doi-org.stanford.idm.oclc.org/10.1038/s41586-021-03380-y.

[10] Aptekar, S. (2017). Looking Forward, Looking Back: Collective Memory and Neighborhood Identity in Two Urban Parks. Symbolic Interaction, 40(1), 101–121. https://www.jstor.org/stable/90000410.