Methods for Compositing Historical Urban Landscapes

Frank M. Palermo
Carnegie Mellon University

Figure 1: Given modern (left) and historical (center) views of an urban landscape, techniques were devised to assist the creation of composite views (right) combining structures from both time periods.

Abstract

This paper discusses the process of compositing historical and modern photographs of an urban landscape into a novel view depicting structures from both time periods. Issues unique to this application of image compositing are addressed using techniques inspired by recent work in data-driven methods, and the use of these methods is demonstrated on an example scene. Finally, extensions of these techniques to other forms of historical photographs are discussed and potential areas of further investigation are identified.

1. Introduction

“Then and Now” photographs have long been a popular means of visualizing changes in the urban landscape. Books of such images have been published for most major cities, and a “Then and Now” group on Flickr (http://flickr.com/groups/thenandnow/) contains over 1,000 items at the time of this writing. Traditionally, the “then” and “now” photographs in such cases have been presented separately in a side-by-side configuration. It may be desirable, however, to instead composite structures which are no longer extant into the modern view of the scene (or alternately to composite modern structures into the historical view). Such composite views may be of interest to historians, educators and others.

To state the problem somewhat more precisely, let us assume that the user has provided an historical photograph for use in the composite. The methods discussed in this paper address the issues of finding a modern photograph of the scene taken from a sufficiently similar vantage point, performing image warping to partially correct for remaining differences in perspective, and partially automating the extraction of the matte required to create the composite image.

Before proceeding, it would be remiss not to mention the “4D Cities” work on spatio-temporal reconstruction from historical imagery by Dellaert et. al. [1]. The author notes that while the “4D Cities” project strives to construct a time-varying 3D model of a city utilizing structure from motion algorithms, the methods discussed in this paper remain within the realm of 2D image processing and compositing. Additionally, while the “4D Cities” project primarily models the actual evolution-over-time of the city in question, the methods discussed here are intended to combine structures from different eras in ways which may never have existed in reality.

2. Methods

Given a user-provided historical photograph upon which to base the composite, the first issue which must be addressed is that of finding a suitable modern photograph of the scene. Replicating the precise location from which the historical photograph was taken is at best difficult and may in fact be impossible (e.g. in the case of photographs taken from the roof of a now-demolished building). It is demonstrated below, however, that precisely matching the location and other characteristics (e.g. field-of-view, depth-of-field, etc.) of the historical photograph is not strictly necessary to achieve an aesthetically pleasing composite of a single structure. This problem, therefore, is framed as a simplified derivative of the “IM2GPS” work of Hays and Efros [2]. Whereas Hays and Efros have already demonstrated the ability to estimate the location of an image by searching for semantically similar scenes in a GPS-tagged database, here we may assume that the general location of the historical scene is already known. Thus the proposed solution to this issue is to search only a subset of a GPS-tagged image database (e.g. only those images tagged “Pittsburgh”) for semantically similar scenes and return a set of probable matches. The single image chosen for use, however, must reflect the artistic desires of the user producing the final composite; the final selection, therefore, would not be automated in the proposed technique. Due to the close relationship of this problem to previous work (and, indeed, to the relative ease of locating suitable photographs manually), it was not investigated in depth by this project. Instead, let us now turn to a discussion of specific issues surrounding the compositing process itself.

Having obtained source images taken from similar (but not identical) perspectives, as in the left and center portions of Figure 1 above, it is necessary to apply some form of image warping to one or both images so that they correspond as closely as possible prior to compositing. Even though the basic assumptions of a projective transform are violated in this situation (the camera has been allowed to translate even though the scene is not planar), experimental results of this work show that projective transforms nonetheless provide subjectively pleasing results for this application. Indeed, even a projective transform estimated from a small number of control points (less than 10 in the examples shown here) performs acceptably as the application is not as demanding as e.g. registration for panorama stitching. This is fortunate considering that as the time delta between the historical and modern photograph grows larger, the decreasing number of structures common to both images will increase the difficulty of finding suitable control points (whether manually or by automated means). The proposed method therefore assumes only a rough approximation of the underlying transform between the images, specifically a projective transform estimated using < 10 manually-selected control points. It should be noted that the requirement to manually specify control points when dealing with historical imagery is not uncommon: the web site of the “4D Cities” project mentioned previously notes that a human must specify the top corners of buildings in each historical image before their structure from motion algorithm is run on the data set. Therefore, this discussion will now proceed to the issue of deriving the matte used in the compositing process.

Manually extracting a matte of sufficient quality to produce an aesthetically pleasing composite of a structure into a new scene could prove tedious even for a skilled user of image editing software. This is especially true for structures such as that depicted in the center portion of Figure 1, a former Pittsburgh river crossing commonly referred to as “Point Bridge II”. In this case, a proper matte would require isolating the individual structural members of the bridge such that the river and other scenery behind the bridge were left visible in the modern photograph (rather than erroneously transferring them from the historical photograph). To both reduce difficulty for the user and improve the quality of the resulting composite, the following method for automatically calculating an initial draft of the required matte is proposed:

First, a number of modern photographs similar to the one selected for the composite are registered to the chosen photograph. Following from the earlier observation that a projective transform provides suitable results for this application, the registration is accomplished using a robust homography extraction via RANSAC similar to that discussed in [3]; features are chosen utilizing the implementation due to Vedaldi and Fulkerson [4] of Lowe's SIFT feature descriptor [5]. It should be noted that, as compared to the panorama-stitching application discussed in [3], this application requires a substantially higher number of RANSAC iterations and a significantly larger error tolerance before control points are rejected as outliers. A number of historical images are registered similarly; it should be noted that this is done after the historical image selected for the composite has already been manually registered to the chosen modern image as discussed previously. Upon completion of the automatic registration process, the algorithm due to Weiss [6] is utilized to estimate intrinsic reflectance images for both the historic and modern scene. The reflectance images are then adjusted to have the same mean intensity and an initial draft of the matte is created based on the remaining differences in pixel intensity between the two images as shown in Figure 2.

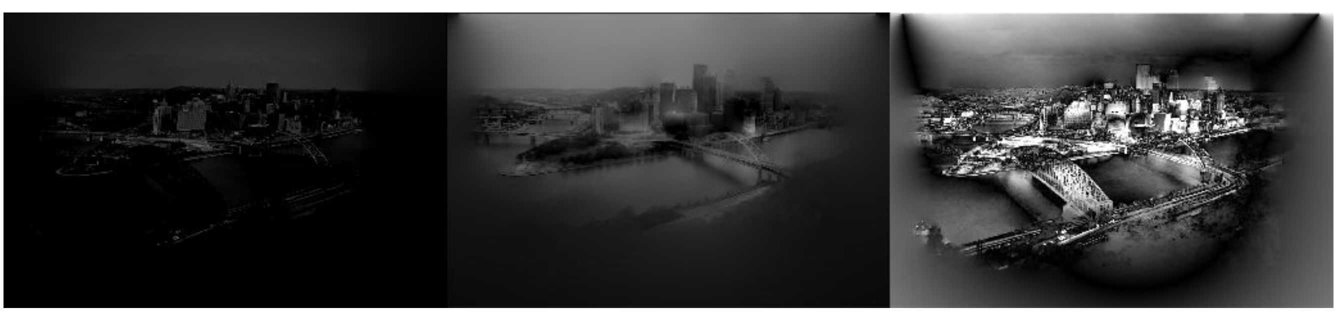

Figure 2: Using the algorithm due to Weiss [6], estimated reflectance images for the historical scene (using 5 images) and modern scene (using 15 images) are shown left and center respectively. The initial draft matte produced by the proposed method is shown at right.

As the structures of interest are entirely absent from either the historical or modern scene, they tend to cause large intensity differences between the derived reflectance images and thus be among the brightest regions in the image shown at right in Figure 2. Subsections of this image therefore provide useful mattes for compositing the desired structures with relatively little manual refinement. Two example mattes are shown in Figure 3; of particular interest is the matte at right which exhibits the desired isolation of most of the bridge's structural members (as well as a semi-transparent blending region for the bridge's shadow) without requiring that either be manually created by the user.

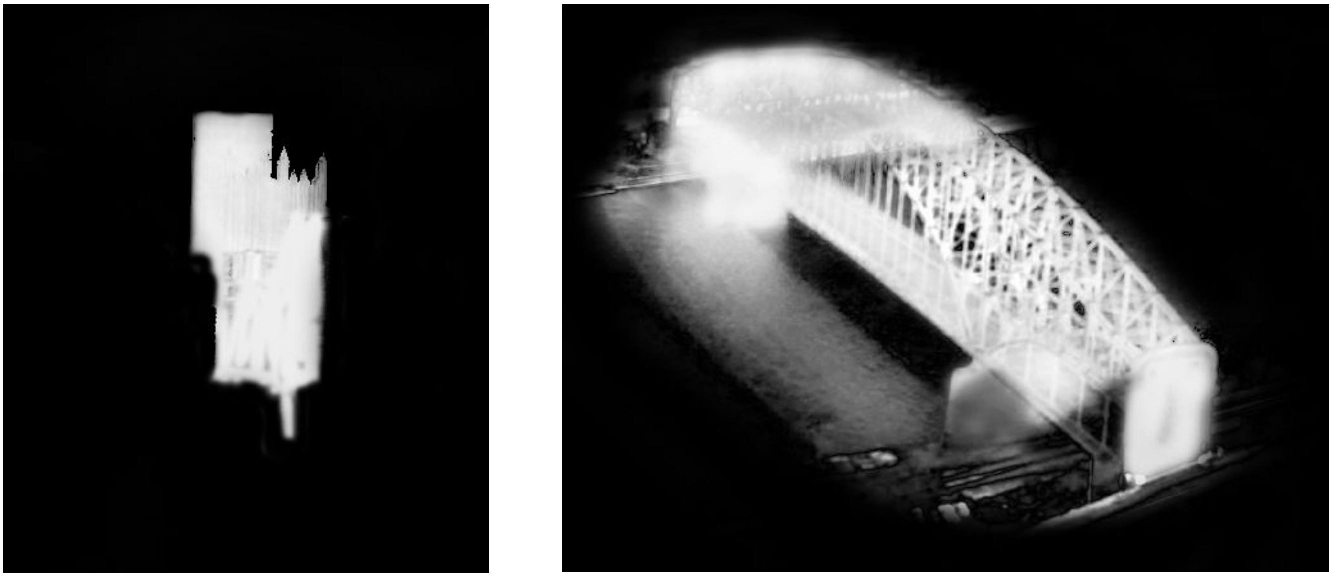

Figure 3: Final mattes used for compositing the U.S. Steel Tower and One PPG Place into an historical photograph (left), and Point Bridge II into a modern photograph (right).

3. Results

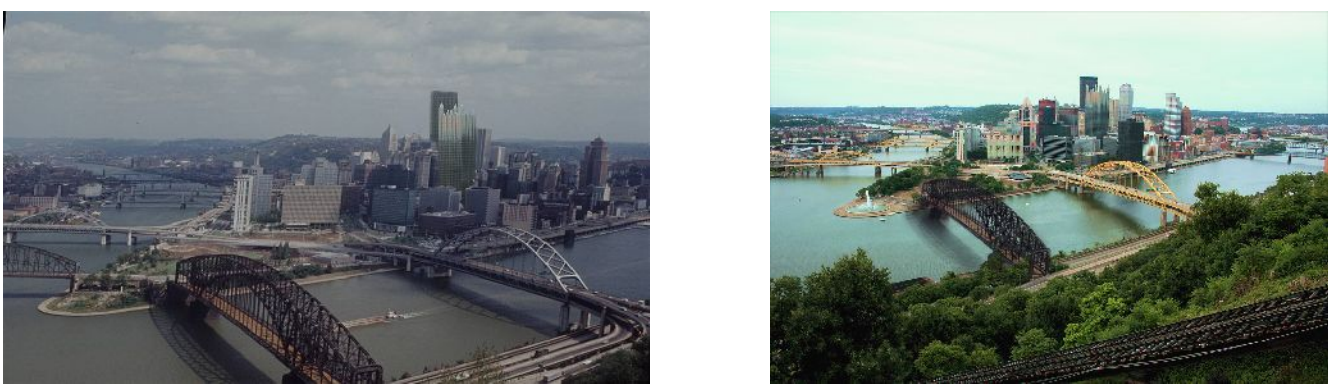

Figure 4: Sample results depicting the composition of modern structures into an historical photograph (left) and vice versa (right) using the described method. (The mattes are as depicted in Figure 3).

4. Discussion

This paper has presented methods for producing aesthetically pleasing composites of historical and modern photographs which provide an alternative to the traditional concept of “then and now” photography in the case of urban landscapes. However, the described model for producing an initial draft of the matte relies on the availability of at least a few similar historic images. Unlike modern images (for which Flickr provides ready access via published APIs), obtaining historical images may require research in a number of university, library, or historical society archives which provide no uniform means of access (and which may, of course, not be accessible online at all). It would therefore be desirable to reduce the requirements of historical imagery to only the single historical photograph selected by the user for the composite. This would require evaluating the performance of current algorithms for estimating intrinsic images from a single source image when confronted with typical historical imagery (and in particular the associated noise and resolution issues). Additionally, although it is currently not unusual for algorithms dealing with historical imagery to require manual intervention for feature selection, it would still be desirable to further automate the described method. It is thought that robust registration algorithms originally intended for images of different modalities (a topic of particular interest in the medical imaging community) may prove useful in this application. Finally, the proposed methods do not in any way address colorization of grayscale historical imagery, adjustments of color or contrast, etc. prior to compositing. While this area may also be pursued for future work, its utility is likely less than those previously stated as such adjustments tend to be subjective and a matter of the user's artistic desires for the result.

Acknowledgements

The author wishes to thank Alexei A. Efros for his advice and suggestions throughout the course of this work.

Images utilized in this research include "A view from atop Duquesne Heights" by Flickr user "Deepak & Sunitha" and various "Pittsburgh - Prominent Citizens" images from the LIFE Photo Archive.

References

[1] G. Schindler, F. Dellaert, and S. B. Kang. Inferring Temporal Order of Images From 3D Structure. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[2] J. Hays and A. A. Efros. IM2GPS: Estimating Geographic Information from a Single Image. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[3] M. Brown, R. Szeliski, and S. Winder. Multi-Image Matching using Multi-Scale Oriented Patches. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[4] A. Vedaldi and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms. http://www.vlfeat.org/, 2008.

[5] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.

[6] Y. Weiss. Deriving Intrinsic Images from Image Sequences. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2001.