FrameBreak: Dramatic Image Extrapolation by Guided Shift-Maps

Yinda Zhang
National University of Singapore

Jianxiong Xiao
Massachusetts Institute of Technology

James Hays
Brown University

Ping Tan
National University of Singapore

Figure 1. Our method can extrapolate an image of limited field of view (left) to a full panoramic image (bottom right) with the guidance of a panorama image of the same scene category (top right). The input image is roughly aligned with the guide image as shown with the dashed red bounding box.

1. Introduction

When presented with a narrow field of view image humans can effortlessly imagine the scene beyond the particular photographic frame. In the computational domain, however, no existing technique can significantly extrapolate a photo because this requires implicit or explicit knowledge of scene layout. Inspired by large-scale database of panoramic photographs [3], we ask the question: is it possible to dramatically extend the field of view of a photograph with the guidance of a representative wide-angle photo with similar scene layout?

Specifically, we seek to extrapolate the FOV of an input image using a panoramic image of the same scene category. An example is shown in Figure 1. The input to our system is an image (Figure 1, left) roughly registered with a guide image (Figure 1, top). The registration is indicated by the red dashed line. Our algorithm extrapolates the original input image to a panorama as shown in the output image on the bottom right. The extrapolated result keeps the scene specific structure of the guide image, *e.g.* the two vertical building facades along the street, some cars parked on the side, clouds and sky on the top, etc. At the same time, its visual elements should all come from the original input image so that it appears to be a panorama image captured at the same viewpoint. Essentially, we need to learn the shared scene structure from the guide panorama and apply it to the input image to create a novel panorama.

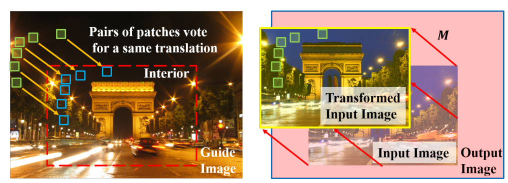

Figure 2. Left: in the guide image, the green patches vote for a common shift vector, because they all can find a good match (blue ones) with this shift vector; Right: The red rectangle is the output image canvas. The yellow rectangle represents the input image shifted by a vector voted by the green patches in the guide image. The data cost within these green patches is 0. The data cost is set to C for the other pixels within the yellow rectangle, and set to infinity for pixels outside of the yellow rectangle.

2. Method

We approach this FOV extrapolation as a constrained texture synthesis problem and address it under the framework of shift-map image editing [2]. We assume that panorama images can be synthesized by combining multiple shifted versions of a small image region with limited FOV. Under this model, a panorama is fully determined by that region and a shift-map which defines a translation vector at each pixel. We learn such a shift map from a guide panorama and then use it to constrain the extrapolation of a limited FOV input image. The shift map is obtained by a graph optimization which minimize the following energy,

$$E(M) = \sum_{q} E_d(M(q)) + \sum_{(p,q)\in N} E_s(M(p), M(q)). \quad (1)$$

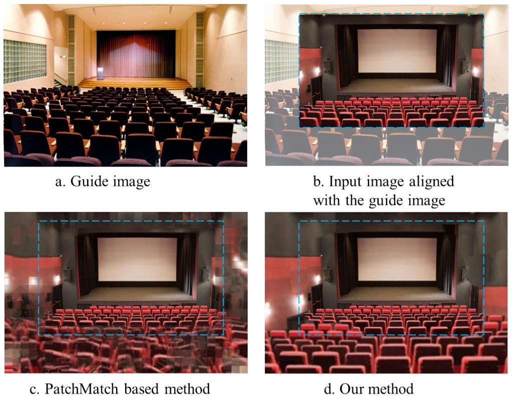

Figure 3. Comparison between our method and a PatchMatch based method.

Here, q is an index for pixels in output image canvas, N is the set of all neighboring pixels. $M(q)$ is a transformation defined on q. $E_d(\cdot)$ is the data term to measure the consistency of the patch centered at q and the position of q after transformation $M(q)$ in the guide image, which is defined as shown in Figure 2. $E_s(\cdot, \cdot)$ is the visual consistency between neighboring pixels in output image. Minimizing E_d would inherit more structural information from the guide image, and minimizing E_s could reduce visual artifacts in the output image.

Because a panoramic scene typically contains surfaces, boundaries, and objects at multiple orientations and scales, it is difficult to sufficiently characterize the self-similarity using only patch translations. Therefore we generalize the shift-map method to optimize a general similarity transformation, including scale, rotation, and mirroring, at each pixel. However, direct optimization of this "similarity-map" is computationally prohibitive. We propose a hierarchical method to solve this optimization in two steps. In the first step, we fix the rotation, scaling and reflection, and optimize for the best translation at each pixel. Next, we combine these intermediate results together with a graph optimization similar to photomontage [1].

3. Experiments

3.1. Comparison with the PatchMatch based method

Figure 3 shows a comparison of our method with a PatchMatch based method in which each patch directly copies source patch from the input image according to the guide image self-similarity. Our method apparently performs better because our optimization makes a proper tradeoff between conforming to the guide image and visual appearance of the output image. The optimization can also handle considerable registration error, and prevent unfavorable self-similarity transferred from the guide image.

Figure 4. Panorama synthesis result. The left column is the input image. On the right are the guide image and the synthesized image.

3.2. Panorama Synthesis

When the guide image is a panoramic image, our method can synthesize the input image to a panorama. However, the transformation space has to be much larger in order to cover the whole panorama image domain, which requires huge memory and computation.

To solve this problem, we first divide the panoramic guide image into several sub-images and synthesize the output for each of these sub-image one by one. Then, we combine all these intermediate results to a full panorama by photomontage, which involves another graph cut optimization. More panorama synthesis results are shown in Figure 4. The success of this divide and conquer approach also demonstrates the robustness of our method, because it requires that all the sub-images be synthesized correctly and consistently with each other.

References

[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graph. (Proc. of SIGGRAPH)*, 23(3):294–302, 2004. 2

[2] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *Proc. ICCV*, pages 151–158, 2009. 1

[3] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Proc. CVPR*, 2010. 1