# Human-centric Panoramic Imaging Stitching

Tomohiro Ozawa
University of Electro-Communications
Chofu, Tokyo, Japan
ozawat@vogue.is.uec.ac.jp

Kris M. Kitani
Carnegie Mellon University
Pittsburgh, PA USA
kkitani@cs.cmu.edu

Hideki Koike
University of Electro-Communications
Chofu, Tokyo, Japan
koike@is.uec.ac.jp

## ABSTRACT

We introduce a novel image mosaicing algorithm to generate 360° landscape images while also taking into account the presence of people at the boundaries between stitched images. Current image mosaicing techniques tend to fail when there is extreme parallax caused by nearby objects or moving objects at the boundary between images. This parallax causes ghosting or unnatural discontinuities in the image. To address this problem, we present an image mosaicing algorithm that is robust to parallax and misalignment, and is also able to preserve the important human-centric content, specifically faces. In particular, we find an optimal path between the boundary of two images that preserves color continuity and peoples' faces in the scene. Preliminary results show promising results of preserving close-up faces with parallax while also being able to generate a perceptually plausible 360° panoramic image.

## Categories and Subject Descriptors

I.4.8 [**Scene Analysis**]: Color,Surface fitting, and Object recognition

## General Terms

Computer Vision, Image Stitching

## Keywords

Content-based image stitching, Face detection, Panorama image

## 1. INTRODUCTION

Photographs are perhaps the most prevalent medium used by people to **augment human memory**. For example, photographs taken in the context of sightseeing are used to keep a record of important people and places. For any given snapshot, we usually choose between a portrait style human-centric photo or a landscape style photo. In fact, due to the wide use of these styles, face detection and landscape mode are a standard feature of most digital camera. To offer a wider field-of-view (FOV) to capture landscape style photos, digital cameras and camera equipped mobile devices often have built-in applications that allow the user to generate panorama images from a collection of images.

**Figure 1: ATT CrowdPhoto [1]. High-resolution stadium image mosaic. Images blended together irrespective of image content.**

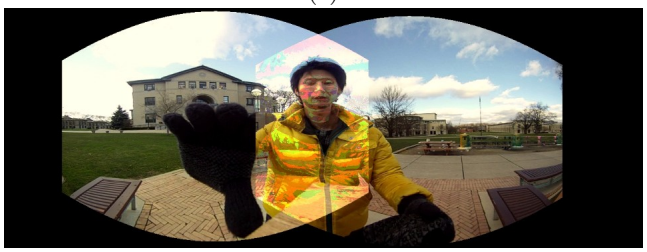
There is a dilemma however, when the user wants to capture close-up faces and a wide field-of-view landscape photo simultaneously. To capture a panoramic image of the landscape, the size of people in the scene become small. Inversely, capturing a close-up view of a person will limit the amount of the landscape that can be captured by the camera. Image mosaicing software, such as Autostitch [5] and Microsoft's Image Composite Editor (MICE) [13] can be use to create a wide FOV image from a collection of smaller images. However, current image stitching algorithms tend to fail when there is a parallax effect between objects at the boundaries between images. This is because most techniques assume that objects are (1) sufficiently far from the camera and (2) objects in the world are static. Therefore, since current image mosaicing software are not designed to deal with parallax effects, nearby objects and moving objects that that lie at the boundary between images cause ghosting or unnatural discontinuities (see Figure 1).

In addition to the use of panoramic imaging for personal digital camera, large panoramic views for large scale spectator sports is also growing in popularity [1]. However, the same problem remains between capturing both the wide FOV and capturing spectator faces. For example in Figure 1, we can see a large high-resolution panoramic image [1] of an American football stadium. Up-close, however, one can



AH '12, March 08-09, 2012, Megève, France.

(a)



(b)

**Figure 2: (a) feature matching (b) sum image**

see that the images have been blended together irrespective of the image content and spectator faces have been blended unnaturally. In this case, the unsatisfactory image blending is caused by the motion of the spectators and the temporal displacement between image acquisition.

To address this gap between portrait style photos and panoramic imaging, we present an image mosaicing algorithm that is (1) robust to parallax caused by nearby or moving objects and (2) is also able to preserve the important human-centric content, namely faces, in panoramic photographs. In particular, we make use of Dijkstra's algorithm to find an optimal path between the boundary of two images that preserves color continuity and face continuity. Qualitative results show how our proposed method is able to preserve human-centric content (human faces) while generating a perceptually plausible 360° panoramic image.

Wide FOV imaging is an important aspect of many HCI applications such as remote robots [15], remote meeting interfaces (Microsoft RoundTable) and augmented reality applications [11]. While omnidirectional cameras can be used to avoid issues of image mosaicing and parallax, high camera cost or low resolution can be a limiting factor in many applications. Therefore, exploring novel algorithms for augmenting human perception from collections of image (or video) is of great relevance for HCI research.

Many approaches have been proposed for generating wide FOV images. A classical yet highly efficient approach is to stitch images and blend the seams between images with linear blending (Google street-view). When more computation time is available, misalignment at images seams can be gain compensated and merged with multi-band blending [6]. However, these approaches assume little or no parallax and have difficulties dealing with objects such as cars or people who are standing near the camera. Parallax and moving objects result in blurry regions in the panorama image.

When there is significant parallax or motion between images, it may be more desirable to select a single best pixel value to generate a crisp focused image. Davis addressed this issue in [8] and used the difference between pixels belonging to different images in overlapped regions to find optimal cutting boundary between images using Dijkstra's algorithm. In computer graphics it has also been shown that image cutting (seam carving) algorithms are well suited for dealing with images with complex textures [12, 2].

(a) (b) (c) (d)

**Figure 3: (a) left image (b) right image (c) alpha-blended image (d) optimal image cut**

In our work, we combine ideas from previous work on image stitching, seam carving and face detection to explore a novel human-centric wide FOV imaging framework. Our work is different from previous work in that our method is able to generating a perceptually plausible 360° panoramic image and is specifically designed to preserve the presence of human faces in overlapping image boundaries.

## 2. HUMAN-CENTRIC IMAGE STITCHING

When generating a human-centric panorama image it is important to take into account the presence of people in the scene as well as preserving the continuity of the background landscape. In this section, we describe our proposed approach and explain how our approach is able to deal with parallax caused by nearby and moving objects; with a explicit focus on preserving human faces. It is important to note that our method aims to generate a perceptually plausible mosaic (not necessarily a physically correct mosaic) by minimizing the adverse effects of parallax during the image mosaicing process.

### 2.1 Generating the panorama

Before finding the correspondences between images it is important that the images are calibrated using the internal parameters of the camera. While it is possible to estimate internal parameters as part of a global optimization process, we use the internal parameters estimated using a calibration rig [18]. Since the cameras used in our experiments where arranged in a cylindrical formation, we project the images into a cylindrical coordinate system [16], then proceed to compute the displacement between images.

To estimate the displacement between images, we compute the planar homography transformation between images using sparse feature point correspondences of SURF [3] features and compute an optimal homography using a random sample consensus algorithm (RANSAC). Figure 2(a) shows the correspondences used for computing the homography matrix. Figure 2(b) shows a visualization of the overlapping region by taking a simple sum of pixel values. We can see that the parallax effect causes bad alignment over the face and body of the person.

At this point, many image stitching techniques use image blending techniques to hide any moderate parallax effects at image boundaries. Blending strategies will produce awkward ghosting effects as in Figure 3(c).

Instead, we aim to understand the content of image bound-

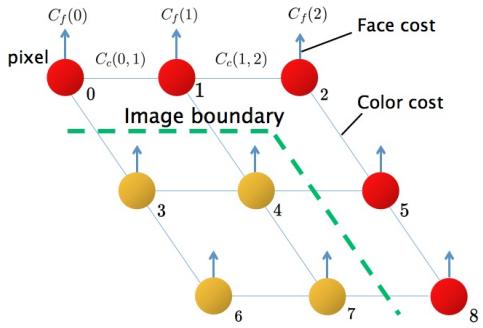



**Figure 4: Cost function over the image graph**

aries to ensure that important objects, like people, are preserved during the image stitching process. To this end, we borrow ideas from texture synthesis [12] and computer graphics [2], and perform a global search for the best image seam carving path that preserves peoples faces. The problem is therefore reduced to a shortest path search algorithm. In this work we use Dijkstra's algorithm for simplicity but other dynamic programming based approaches or graph cut algorithms can also be utilized. Furthermore, accelerated algorithms using other heuristics may also be viable to speed up the computation time (e.g., Fibonnacci heaps[9], A*[10]). For applications to video, spatio-temporal graph-cuts [12] or perturbation theory [4] maybe used to enforce temporal continuity. It may also be possible to using algorithms such as the Floyd-Warshall algorithm [14] to find more optimal start and end points. Currently, start and end points are hard-coded.

The cost function we minimize to discover the most optimal carving path is a sum of two terms (Figure 4),

$$C_t(i,j) = \alpha_c C_c(i,j) + \alpha_f C_f(j), \tag{1}$$

where $C_c(i,j)$ is the cost of color continuity between two pixels and $C_f(j)$ is the cost based on the detection of a face in the image. The index $i$ is the current pixel and $j$ is a neighboring pixel. The weight parameters $\alpha_c$ and $\alpha_f$ are the weights of the color cost function and face cost function, respectively.

The pairwise color continuity cost between two pixels $i$ and $j$, $C_c(i,j)$, is computed as the color difference between the left image $\mathbf{A}$ and right image $\mathbf{B}$,

$$C_c(i,j) = \frac{1}{N_c}\left\{||\mathbf{A}(i) - \mathbf{B}(i)||_{L2} + ||\boldsymbol{A}(j) - \boldsymbol{B}(j)||_{L2}\right\}, \tag{2}$$

where $\boldsymbol{A} = [r_A, g_A, b_A]^T$ and $\boldsymbol{B} = [r_B, g_B, b_B]^T$ are 3 dimensional color vectors. The normalization constant ensures the value lies between 0 and 1.

The unary face region cost function $C_f(j)$ is defined as,

$$C_f(j) = \frac{1}{N_f}\left\{\frac{r^2}{4} - ||\boldsymbol{\mu}_c - \boldsymbol{x}(j)||\right\}, \tag{3}$$

where $\boldsymbol{\mu}_c$ is center of the face and $r$ is the radius of the face detected by the Viola and Jones face detector [17]. The vector $\boldsymbol{x}(j)$ is the pixel location of $j$ and $N_f$ is the normalization factor.

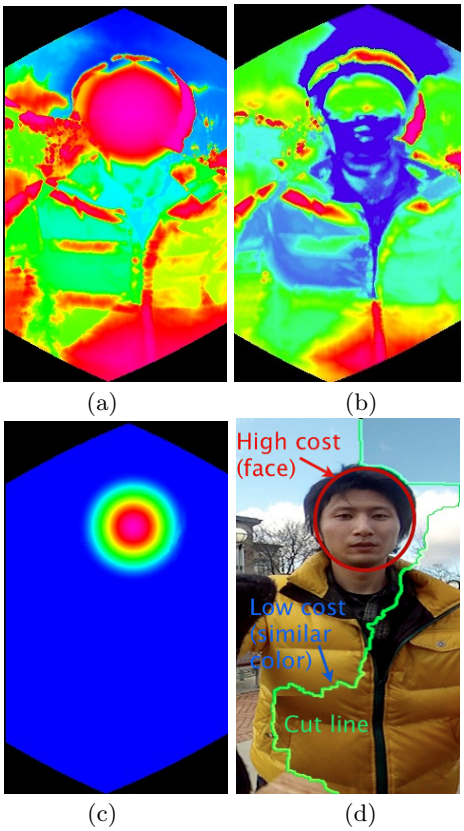



**Figure 5: Visualization. (a) total cost (b) color cost (c) face cost (d) image boundary**

We calculate the optimal carving path by Dijkstra's algorithm, where the objective is to finding the optimal path $\boldsymbol{\lambda} = \{i_n, j_n\}_{n=1}^{N}$ (set of edge transitions from $i_n$ to $j_n$) that minimizes the total path cost,

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{n} C_t(i_n, j_n). \tag{4}$$

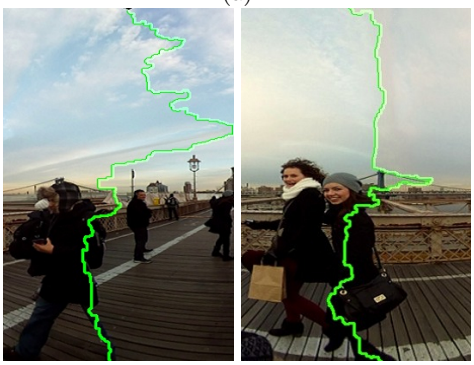
Figure 5(a) is a visualization of the total cost of terminating at that pixel. High cost areas are pink-red and low cost areas are blue-green. The cost of color continuity and face region are shown in FIgure 5(b) and Figure 5(c), respectively. Figure 5(d) shows the carving path computed by Dijkstra's algorithm. We can see that that the path avoid faces and prefers regions that have similar color (sky area).

Although we have focused on faces in this work, our framework is not limited to only preserving faces. By adding additional terms to the cost function, such as body detection[7] or object detection we can preserve other important object categories.

## 3. QUALITATIVE ANALYSIS

We now show qualitative results of our proposed approach and show how our approach is robust to parallax caused by nearby faces in overlapping regions of the composite panorama image. Figure 6 (a) shows the composite image without any blending (a sum of image). It is clear that there is significant parallax in overlapping regions. Figure 6 (d) shows the final result of our proposed method by taking into account the head region to compute the optimal image boundary.

(a)



(b)

**Figure 7: (a) Success and (b) failure cases**

The performance of two commercial image stitching programs, namely, AutoStitch [5] and Microsoft Image Collection Editor [13] are also shown in Figure 6(b) and Figure 6(c), respectively. Notice that there is ghosting in Figure 6(b) near the head region due to the parallax caused by the proximity of the head to the camera. There is gross misalignment in the head region in Figure 6(c).

In Figure 7 we show both successful and unsuccessful image stitching results using our proposed method. Since our current work only considers regions that contain a face, the image boundary optimization is not always able to preserve the full body of people in overlapping areas of the composite image. Adding additional detectors or introducing heuristics about body position relative to face detections can be used to improve human-centric image stitching.

Our method relies on the fact that there is sufficient overlap between neighboring images. In our current implementation, each pair of images overlap by 50%. Our method is only able to preserve faces, when they are smaller than this overlapped region. When the face is extremely close to the camera the face region will extend beyond the overlapping region and our method will not be able to find an optimal image cut.

# 4. CONCLUSION

We have shown that our proposed method is able to generate a 360° FOV panorama image, while taking into account the presence of faces in the overlapping regions between sub-images. Qualitative results show that our method can outperform general purpose image stitching applications with respect to preserving human faces.

# ACKNOWLEDGMENTS

This research was supported by The Canon Foundation.

# 5. REFERENCES

[1] AT&T. Crowdphoto. `http://uverseonline.att.net/usc-crowdphoto`.

[2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26, 2007.

[3] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. *European Conference on Computer Vision*, pages 404–417, 2006.

[4] A. Briuno. Normal form in perturbation theory. In *8th International Conference on Nonlinear Oscillations, Proceedings*, volume 1, pages 177–182, 1979.

[5] M. Brown. Autostitch. `http://www.cs.bath.ac.uk/brown/autostitch/autostitch.html`.

[6] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[8] J. Davis. Mosaics of scenes with moving objects. In *Computer Vision and Pattern Recognition, Proceedings.*, pages 354–360, 1998.

[9] M. Fredman and R. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, 34(3):596–615, 1987.

[10] P. Hart, N. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics*, 4(2):100–107, 1968.

[11] V. Kuchelmeister, J. Shaw, M. McGinity, D. Del Favero, and A. Hardjono. Immersive mixed media augmented reality applications and technology. *Advances in Multimedia Information Processing*, pages 1112–1118, 2009.

[12] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics*, 22(3):277–286, 2003.

[13] Microsoft. Microsoft image composite editor. `http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice/`.

[14] C. Papadimitriou and M. Sideri. On the Floyd-Warshall algorithm for logic programs. *The Journal of Logic Programing*, 41(1):129–137, 1999.

[15] R. Sarvadevabhatla and V. Ng-Thow-Hing. Panoramic attention for humanoid robots. In *Humanoid Robots*, pages 215–222, 2009.

[16] M. Uyttendaele, A. Eden, and R. Skeliski. Eliminating ghosting and exposure artifacts in image mosaics. In *Computer Vision and Pattern Recognition, Proceedings.*, volume 2, pages 509–516, 2001.

[17] M. Viola, M. Jones, and P. Viola. Fast multi-view face detection. In *Computer Vision and Pattern Recognition, Proceedings.*, 2003.

[18] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

(a)

(b)

(c)

(d)

**Figure 6: Comparing stitching results. (a) naive composite image (b) AutoStitch (c) Microsoft Image Composite Editor (d) Proposed Method**

(a)

(b)

(c)

**Figure 8: Results on different scenes. (a) AutoStitch (b) Microsoft ICE (c) Proposed Approach**