Panorama to Cube: A Content-Aware Representation Method

Zeyu Wang, Xiaohan Jin, Fei Xue, Xin He, Renju Li, Hongbin Zha*
Key Laboratory of Machine Perception, School of EECS, Peking University

Figure 1: *A novel representation method for panoramas using Content-Aware Cube Unwrapping.*

Abstract

As panoramas provide a brand new viewpoint for the public, relevant cameras and software such as RICOH Theta, Microsoft Photosynth are embracing more and more users. However, the display methods for panoramas remain monotonous. In this paper, we propose a novel representation method called Content-Aware Cube Unwrapping using the effective and interactive techniques of orientational rectification, image modification and energy estimation. Thus, a number of fascinating applications will come into reality. For instance, six surfaces of a Rubiks cube can be automatically rendered from a vertically oriented panorama, without cutting any person or significant object apart. Moreover, seam carving and inserting are applied to each surface to enhance the key content and to make the scenery more consistent.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms; I.4.0 [Image Processing and Computer Vision]: General—Image Displays;

Keywords: panorama, vanishing point, image unwrapping, energy estimation, seam carving

1 Introduction

Nowadays, panoramas are gaining more and more attention and preference because of their complete delivery of visual information. In the field of panoramic image processing, much attention has been paid to image stitching [Ying et al. 2009]. However, while the majority are making every possible endeavor to produce better panoramas, few people are interested in representing and amplifying their central information friendly when panoramas have been already generated by panoramic cameras such as RICOH Theta. As nearly all images and visual resources are digital, is there any possibility to convert a panorama to an actual 3D model that has an intimate relationship to humans, without any obvious aesthetic flaw or perceptible information loss?

We focus on Content-Aware Cube Unwrapping by precisely estimated fold lines. Our method involves orientational rectification through the classical method of vanishing point estimation based on random or interactive unwrapping, energy estimation and seam carving for emphasizing salient human or objects, final unwrapping and reprojection. Theoretically by our method, a spherical panorama can be projected onto an arbitrary 3D shape without separating key content to different surfaces.

2 Related Work

Line detection and vanishing point estimation are widely used in orientational rectification for non-panoramic images. We popularize them to panoramic image processing by cube unwrapping. Progressive probabilistic Hough transform (PPHT) [Matas et al. 2000] is well adapted for detecting long lines. Some research focus on calculating cluster of parallel lines [Kim et al. 2011; Li et al. 2012]. In [Lezama et al. 2014], point alignments are based on line segment detection [von Gioi et al. 2010] and are defined in both image primal and dual domains, but parameters are more complicated.

Clear content representation is also required. Originally, image energy such as intensity gradient is designed for edge detection, but its improved version can be regarded as a criterion for image content measurement and salient detection [Avidan and Shamir 2007; Chen et al. 2010; Cheng et al. 2015] to determine the most appropriate unwrapping solution [Carroll et al. 2009].

Based on the above, we propose a novel method called Content-Aware Cube Unwrapping. Thereafter, with our effective and interactive techniques, designers and industrial practitioners will be empowered to develop various amusive applications.

*Email: {1200012927, 1200012629, 1200018415, hex}@pku.edu.cn, {lirenju, zha}@cis.pku.edu.cn

SA'15 Technical Briefs, November 02 – 06, 2015, Kobe, Japan.
 ISBN 978-1-4503-3930-8/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2820903.2820911

3 Orientational Rectification

3.1 Representations of Panorama

Representations of panorama include spherical projection, cubic projection, planar fisheye images and so on. Spherical projection is usually provided as a 360° field-of-view virtual scene roaming. Fisheye images are standard and intact. Cubic ones consist of images from six perspectives, which satisfy the Manhattan world assumption and can help orientational rectification based on perspective transformation properties. For every pixel (p, q) on the cube surface, its corresponding point (x, y, z) in the spherical panorama and pixel (u, v) in the fisheye image can be deducted from these equations:

$$(x, y, z) = \boldsymbol{\mathcal{L}}_z + \left(p - \frac{w}{2}\right) \boldsymbol{\mathcal{L}}_x + \left(q - \frac{w}{2}\right) \boldsymbol{\mathcal{L}}_y \quad (1)$$

$$u = \frac{\pi - \arctan \dfrac{x}{z}}{2\pi} W + \frac{1}{2} \quad (2)$$

$$v = \frac{\dfrac{\pi}{2} - \arccos \dfrac{z}{R}}{\pi} H + \frac{1}{2}, \quad (3)$$

where $\boldsymbol{\mathcal{L}}_z = (x_c, y_c, z_c)$ is the center of a cube surface in the global coordinate system, w is the side length of the surface, $\boldsymbol{\mathcal{L}}_x$ and $\boldsymbol{\mathcal{L}}_y$ are two coordinate axes besides $\boldsymbol{\mathcal{L}}_z$ in the target cubic panorama, W is width of the fisheye panorama, and H is height.

3.2 Vanishing Point Estimation

To calibrate the cubic panorama, line clustering and orthogonal vanishing point estimation can be applied to each surface. We use PPHT [Matas et al. 2000] to detect reliable straight lines, which works by exploiting the difference in the fraction of votes with minimized supporting points. Weight W is calculated according to distance between the chosen point V_i with all other intersection points in its adjacent space $\Omega_i (r = 20, \epsilon > r)$ [Kim et al. 2011].

$$V = \underset{i}{\operatorname{argmax}} \left[W = \left(\sum_{j \in \Omega_i} \frac{\epsilon - D_{ij}}{\epsilon} \right) \right]_{i=0,1,2,\ldots} \quad (4)$$

$$D_{ij} = \underset{j \in \Omega_i}{dist} \{V_i, V_j\} \quad (5)$$

The vanishing point V is the clustering center of these intersection points with maximum weight. The Manhattan world assumption points out that there exists only three orthogonal vanishing points, which is also a constraint. The angle between the former and new z-axis is the rotating angle ω. According to the rotation matrix, we can remap the pixels to a new calibrated fisheye panorama.

4 Panorama Unwrapping

After orientational rectification, we have a prefect panorama with a perceptually-friendly orientation. Instead of the traditional representation of panoramic sphere and viewing from the fixed center point, we commit ourselves to finding new mapping methods from panoramas to other three-dimensional centrosymmetric shapes, such as cubes, cuboids, cylinders, etc. However, despite potential information loss, it also involves a problem of edge excision. How to render images of all the surfaces without cutting apart important people and salient scenery? In order to solve this, we first introduce several engery functions for salient detection, which satisfies our following aesthetic criteria.

- People and important objects are not divided to different sides.

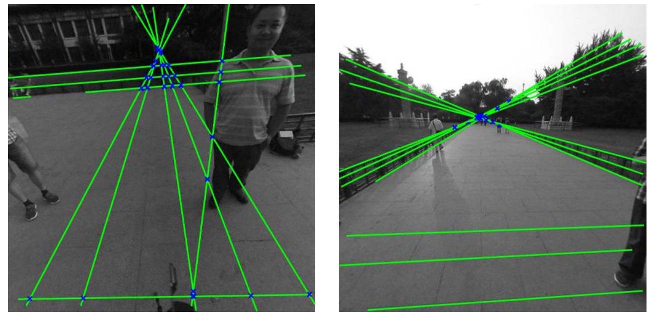


Figure 2: *Line detection based on progressive probabilistic Hough transform and vanishing point estimation by weighing the clustering center of the intersection points.*

- The scenery stays consistent in content on adjacent surfaces.
- Meaningless information is deemphasized in a proper way.

4.1 Energy Function

Our approach to content-aware unwrapping is to find fold lines in a judicious manner. Intuitively, we have to avoid projecting a continuous object on separated surfaces. This leads to the following simple and popular energy function based on partial derivatives of intensity, which can be simply implemented by Sobel operator.

$$e_{Intensity}(\boldsymbol{I}) = \left| \frac{\partial}{\partial x} \boldsymbol{I} \right| + \left| \frac{\partial}{\partial y} \boldsymbol{I} \right| \quad (6)$$

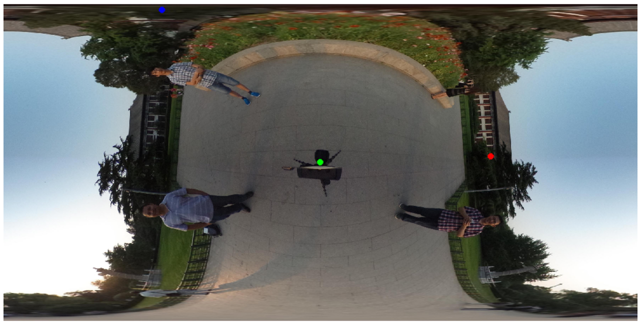
Taking principles of information theory into consideration, entropy measures the amount of information in a certain window Ω, whose size is 9×9 in our calculation. p denotes the probability of intensity and we certainly do not want folds to cross information-rich area.

$$e_{Entropy}(\boldsymbol{I}) = - \sum_{i \in \Omega(\boldsymbol{I})} p_i \log p_i \quad (7)$$

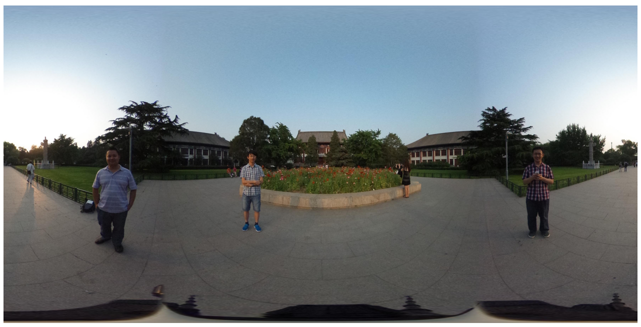
Except intensity, color consistency is also a vital criterion of energy measurement. If the panorama is stitched together with several separated images, whose exposure varies, focusing only on intensity may detect an unreal edge. By converting RGB domain to HSV domain, we can calculate the hue value and its partial derivatives of each pixel, thus have a basic understanding of color variation.

$$e_{Color}(\boldsymbol{I}) = \left| \frac{\partial}{\partial x} Hue(\boldsymbol{I}) \right| + \left| \frac{\partial}{\partial y} Hue(\boldsymbol{I}) \right| \quad (8)$$

Despite the simplicity of these definitions, their linear combination can obtain fairly ideal results, where the weight is set through multiple experiments. Other optimizing methods are still available, such as texture analysis based on Gabor transform and frequent feature classification, histogram of oriented gradients in intensity domain, detection of face and region of interest, etc. It is also a better way to use adaptive scale when revising and improving the energy function. For example, the energy function of a forest image may be extremely high because of foliage and space in between. To solve this problem, we may do brief segmentation and labeling automatically or interactively, and then choose a proper level of image pyramid via smoothing or sharpening filters. Optionally, we can adjust the window size when calculating gradients and entropy of intensity and hue as well.



(a)



(b)

Figure 3: *(a) Projection of estimated three axes in the original image, where red point stands for x-axis, blue point y-axis and green point z-axis; (b) The rectified scene.*

4.2 Panorama Resizing

Before unwrapping a panorama to any 3D shape, it is a wise decision to do some resizing pretreatment for appropriate content. For example, photographers usually use a tripod to capture vertical images as mentioned, but its residue at bottom in the panorama is regarded as meaningless information, which we can simply cut it off. However, how to recover the original width-height ratio without stretch and distortion? We introduce a method called seam carving and inserting [Avidan and Shamir 2007] using the idea of dynamic programming.

Considering we are short of some rows in a picture, we can either insert some horizontal seams or delete some vertical seams, which are defined below:

$$\begin{aligned} \boldsymbol{s^x} &= \{s_i^x\}_{i=1}^n = \{(x(i), i)\}_{i=1}^n, \\ & s.t. \forall i, |x(i) - x(i-1)| \leq 1 \end{aligned} \tag{9}$$

$$\begin{aligned} \boldsymbol{s^y} &= \{s_j^y\}_{j=1}^m = \{(j, y(j))\}_{j=1}^m, \\ & s.t. \forall j, |y(j) - y(j-1)| \leq 1 \end{aligned} \tag{10}$$

The first step is to traverse the image from the second row to the last row and compute the cumulative minimum energy M_x for all possible connected seams for each entry (i, j). M_y can be calculated in a similar way.

$$\begin{aligned} & M_x(i, j) = e(i, j) + \\ & \min\{M_x(i-1, j-1), M_x(i-1, j), M_x(i-1, j+1)\} \end{aligned} \tag{11}$$

$$\begin{aligned} & M_y(i, j) = e(i, j) + \\ & \min\{M_y(i-1, j-1), M_y(i, j-1), M_y(i+1, j-1)\} \end{aligned} \tag{12}$$

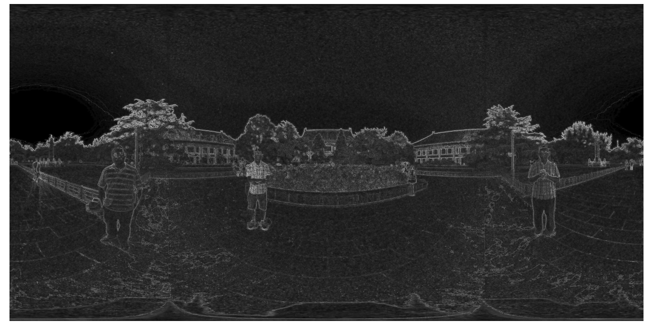


Figure 4: *The illustration of our energy function, which is the weighed combination of energy of intensity, entropy and color.*

(a) (b) (c)

Figure 5: *Some special cases are better handled using optimized techniques. (a) The different exposure can be eliminated in hue domain; (b) Gaussian blur weakens the importance of foliage; (c) Face and ROI detection is also an improvement method.*

While dynamic programming, we can set a marking array to record the path of seams. If there are t rows of shortage, we should just insert some horizontal seams and/or carve some vertical seams with lowest total energy to get a new panorama with recovered size.

4.3 Geometric Constraints

To unwrap fisheye panorama to six-surface cube as an example, the corresponding boundaries can be decided by geometric constraints. For a panorama shot from RICOH Theta, the size is 1792×3584 and we take it as baseline. As is shown in Figure 6, there are four equidistant vertical lines in a panorama, which denotes four edges on side surfaces, and eight arcs for edges on the top and bottom surfaces. These arcs are projected from the center of sphere to edge lines, so each coordinate (r, θ, ϕ) of arc points satisfies:

$$\tan\theta = \frac{1}{\cos\phi} \tag{13}$$

We define a point set $\boldsymbol{F}(y_0)$, all of whose elements compose twelve edges of a cube after projection. y_0 is the vertical translation and our algorithm is to find a y_0 to minimize the total energy, i.e. the best unwrapping solution.

$$Opt = \min_{y_0} \sum_{\boldsymbol{I} \in \boldsymbol{F}(y_0)} e(\boldsymbol{I}) \tag{14}$$

5 Experiments and Conclusion

5.1 Surface Editing

There is still something to do even with the rendered surface images. Based on energy functions we have calculated, image editing

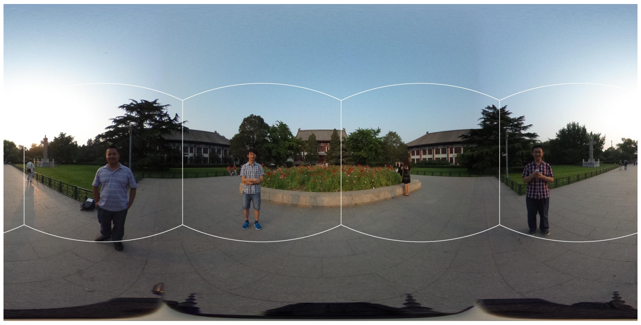


Figure 6: *Lines and arcs which are projected to cube edges, as a result of energy minimization and geometric constraints.*

techniques including seam carving and inserting with scaling [Avidan and Shamir 2007] are powerful tools for content amplification, object removal and so on.

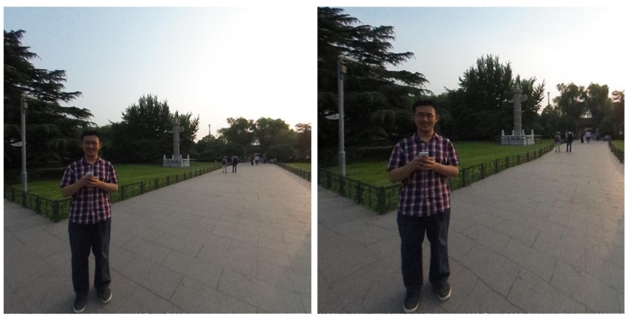


Figure 7: *Content amplification. From left to right: a combination of seam carving and scaling amplifies the content of original image.*

5.2 Applications

Since the invention of Rubik's Cube in 1974, this 3D combination puzzle have become a proverbially popular game and is regarded as the world's best-selling toy. What an inspiration! Once we project six rendered images on each surface, a panoramic magic cube is born. Because the unwrapping solution avoids separating important content on different sides, it is a challenging game tool as well as an exquisite souvenir. Our algorithm also has practical significance on content-aware mapping for other rectangular cuboids such as pen containers, packing boxes.

5.3 Future Work

In this paper, we propose a novel method to manufacture actual stereo shapes from amazing panoramas. The unwrapping solution is guaranteed to stay accordance with aesthetic criteria and effective information is preserved utmostly. We have tested a set of panoramas on our algorithm and 94.7% of them achieve rather ideal results, where people are never on separated surfaces. More results are provided in the supplementary material. Nevertheless, it makes matters worse when there are plenty of people or when people get so close to the camera, because of the strong geometric constraints of cube. In the future, we will explore the projection from panoramic sphere to cuboid, which will introduce two more degrees of freedom for the optimization problem, and more types of polyhedron and 3D printing will be probable to get involved as well. Further research can be also combined with augmented reality displays.

Figure 8: *A panoramic Ribik's cube.*

References

Avidan, S., and Shamir, A. 2007. Seam carving for content-aware image resizing. In *ACM Transactions on graphics (TOG)*, vol. 26, ACM, 10.

Carroll, R., Agrawal, M., and Agarwala, A. 2009. Optimizing content-preserving projections for wide-angle images. In *ACM SIGGRAPH 2009 Papers*, ACM, New York, NY, USA, SIGGRAPH '09, 43:1–43:9.

Chen, J., Miao, L., and Liu, X. 2010. Balanced energy for content-aware image resizing. In *Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on*, 24 – 29.

Cheng, M., Mitra, N., Huang, X., Torr, P., and Hu, S. 2015. Global contrast based salient region detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 37*, 3 (March), 569–582.

He, K., Chang, H., and Sun, J. 2013. Rectangling panoramic images via warping. *ACM Trans. Graph. 32*, 4, 79:1–79:10.

Kim, J. W., Kim, T. H., and Jo, K. H. 2011. Traffic road line detection based on the vanishing point and contour information. *Sice Annual Conference Proceedings of*, 495 – 498.

Lezama, J., Grompone von Gioi, R., Randall, G., and Morel, J.-M. 2014. Finding vanishing points via point alignments in image primal and dual domains. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, 509–515.

Li, B., Peng, K., Ying, X., and Zha, H. 2012. Vanishing point detection using cascaded 1d hough transform from single images. *Pattern Recognition Letters 33*, 1, 1 – 8.

Matas, J., Galambos, C., and Kittler, J. 2000. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding 78*, 1, 119–137.

von Gioi, R., Jakubowicz, J., Morel, J.-M., and Randall, G. 2010. Lsd: A fast line segment detector with a false detection control. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 32*, 4 (April), 722–732.

Ying, X., Peng, K., and Zha, H. 2009. Walkthrough in large environments using concatenated panoramas. In *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on*, IEEE, 286–291.