Saliency-based Feature Selection Strategy in Stereoscopic Panoramic Video Generation

Haoyu Wang[1], Daniel J. Sandin[2] and Dan Schonfeld[3]
[1,3]Electrical and Computer Engineering Department , University of Illinois at Chicago, [2]Computer Science Department, University of Illinois at Chicago

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

Purpose

In this paper, we present one saliency-based feature selection and tracking strategy in the feature-based stereoscopic panoramic video generation system. Many existing stereoscopic video composition approaches aims at producing high-quality panoramas from multiple input cameras[1], [2], [3], [4], [5]; however, most of them directly operate image alignment on those originally detected features without any refinement or optimization. The standard global feature extraction threshold always fails to guarantee stitching correctness of all human interested regions. Thus, based on the originally commonly identified feature set, we incorporate the saliency map into the distribution of control points to remove the redundancy in texture-rich regions and ensure the adequacy of selected features in visual sensitive regions. The experiments show that our method can improve the stitching quality of visual important region without impairment to the human less-interested regions in the generated stereoscopic panoramic video.

Outline

In this paper, the proposed general disparity control strategy is established based on the construction of a commonly identified feature set. Then, we combine the disparity map, gradient map and saliency map into one energy map that indicates the visual importance of each pixel in the image. Given the total number of control points we intends to sample, we select those best-matched commonly-identified features in each grid. In the feature tracking process, we also do the local feature update based on the change of energy in each grid.

Commonly Identified Feature Generation

The stitching of binocular views based on those features from the same edge, corner, or object will maintain the stereo consistency between the left and right views. For simplicity, we only consider the general task of stitching two pairs of input rectified stereoscopic images I_{L1}, I_{R1}, I_{L2} and I_{R2}. We define four randomly chosen feature descriptors (e.g., SIFT or SURF) from the four images as d_1, d_2, d_3 and d_4 respectively. Each descriptor contains one vector $d_i.v$ to record the gradient in multiple directions, and two scalars $d_i.x$ and $d_i.y$ $ for the center point position. The score to evaluate the correspondence between them is defined as follows:

$$\epsilon(d_1,d_2,d_3,d_4) = \sum_{i=1}^{3}\sum_{j=i+1}^{4}\left\|d_i.v - d_j.v\right\|^2 + \alpha_1(\|d_1.y - d_3.y\|^2 + \|d_2.y - d_4.y\|^2)$$
$$+\alpha_2\left\|\frac{f*b}{d_1.x - d_2.x} - \frac{f*b\ *cos\theta}{d_3.x - d_4.x}\right\|_2$$

The first term above refers to the Euclidean distance between any two feature descriptors. The second and third terms are the vertical position difference between two matched center points. The last term is the difference in depth from triangulation between the left and right views. The symbol f is the focal length and b is the baseline.
Thus, the construction of the commonly identified feature set could be decomposed into multiple optimization problems for each extracted feature descriptor:

Fig. 1: Constructed commonly-identified feature set: The control points only connected by red line in vertical or horizontal direction will be rejected. Only the control points connected by yellow line in both of vertical or horizontal direction will be selected for Commonly-identified feature set.

Saliency-based Feature Selection

To generate visual sensitivity map with more sharp boundary, one energy fusion function [6] is used to combine the gradient map and GBVS-based saliency map [7] as:

$$e(i,j) = \alpha_1 * Dep(i,j) + \alpha_2 * Grad(i,j) + \alpha_3 * Sal(i,j)$$

For each grid $G_{p.q}$, its corresponding normalized weight is defined as:

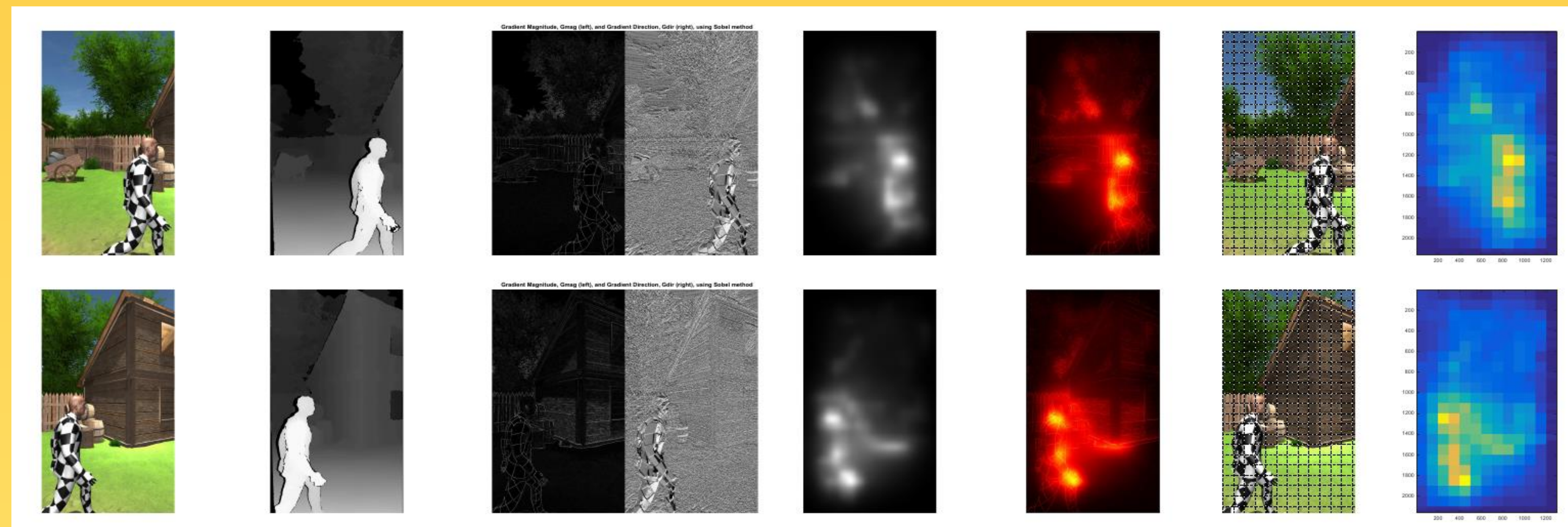
$$w'_{p,q} = \frac{w_{p,q}}{\sum w_{p,q}} \quad \text{where } w_{p,q} = \sum_{i,j\in G_{p,q}} e(i,j)$$

Given the total number of control points (dented as T), we select the best matched CIF(commonly-identified feature) in each grid as control points:

$$B_{p,q} = T \times w'_{p,q}$$

$$R(d_1,d_2,d_3,d_4) = \frac{\beta_1}{\epsilon(d_1,d_2,d_3,d_4)} + \beta_2\sum_{i=1}^{4}[\sqrt{(d_{i,x} - d'_{i,x})^2 + (d_{i,y} - d'_{i,y})^2}]$$

$B_{p,q}$ is the number of cp we sample in each grid, R is the matching score for candidates



(a) (b) (c) (d) (e) (f) (g)

Fig. 2 Images from left to the Right are: (a) Original images; (b) Depth map; (c) Gradient map; (d) Saliency map; (e) Energy map; (f) Grid map; (g) Grid map with assigned visual weight.

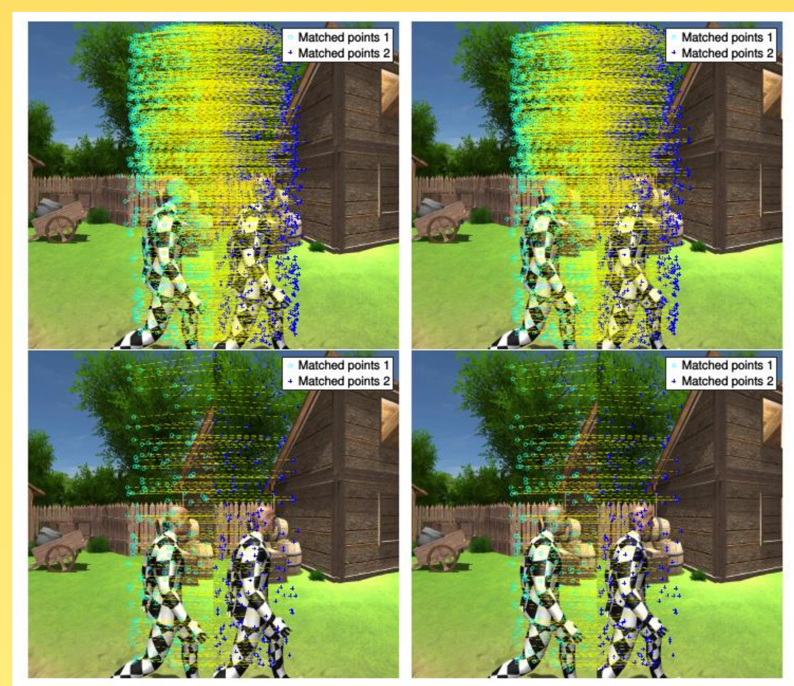


Fig. 3 Matched control point pair before and after feature selection.

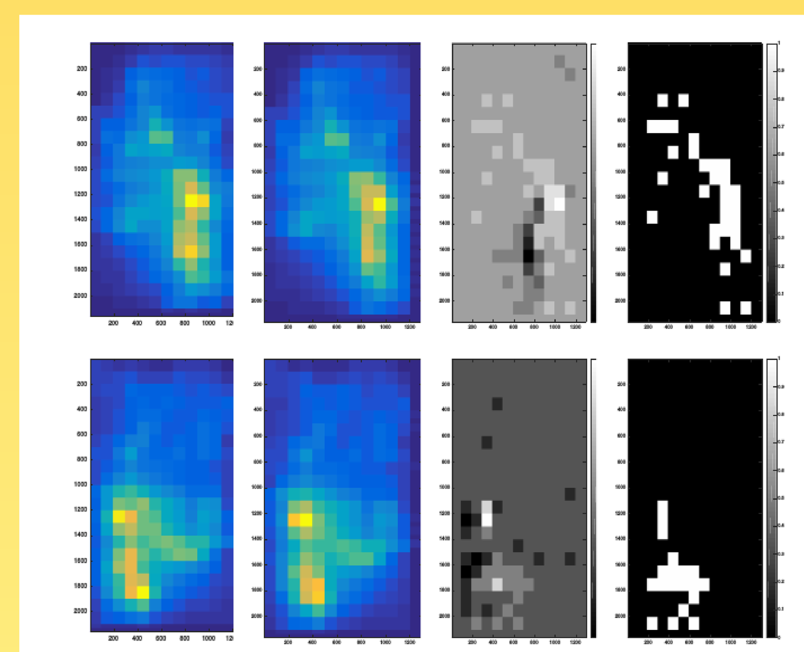


Fig. 4 Images from Left to the Right are: (a) Grid map between consecutive frames (b) Grid-level energy change map (c) Indicator map.

Local Feature Update Strategy

Step1: Grid classification to determine which grid need feature update
Step2: Run KLT tracking in adjacent camera views simultaneously
Step3: Filter the tracking result with proposed depth-related constraints
Step4: Fill those missing control point slot with retrieved position

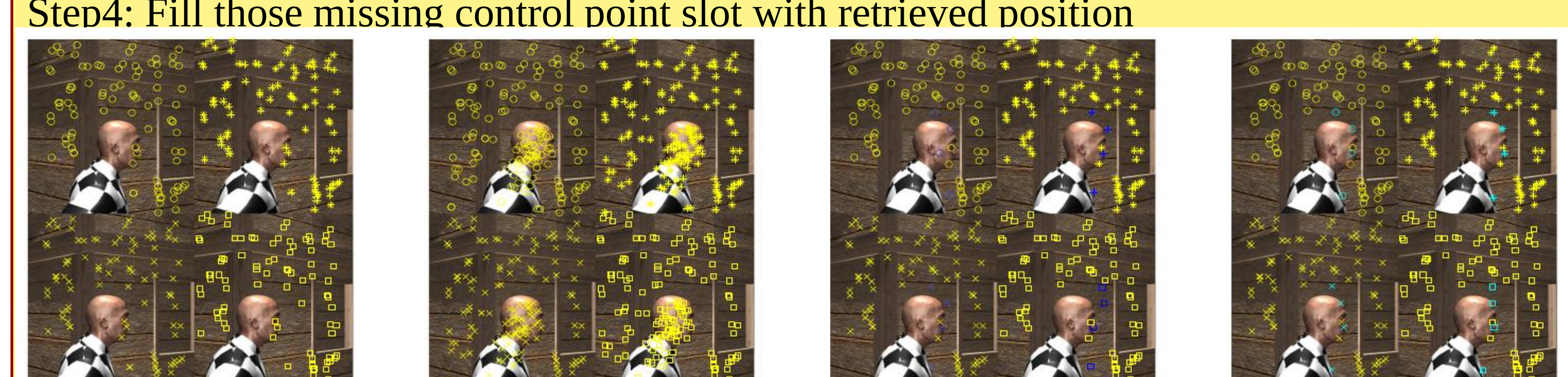


Fig. 5 Update result comparison between different strategies: (a) original detected features; (b) purely detected; (c) purely tracking; (d) proposed strategy.

Visual improvement in Stitched Panorama/Video

Fig. 6 Comparison of stereoscopic panorama stitching result

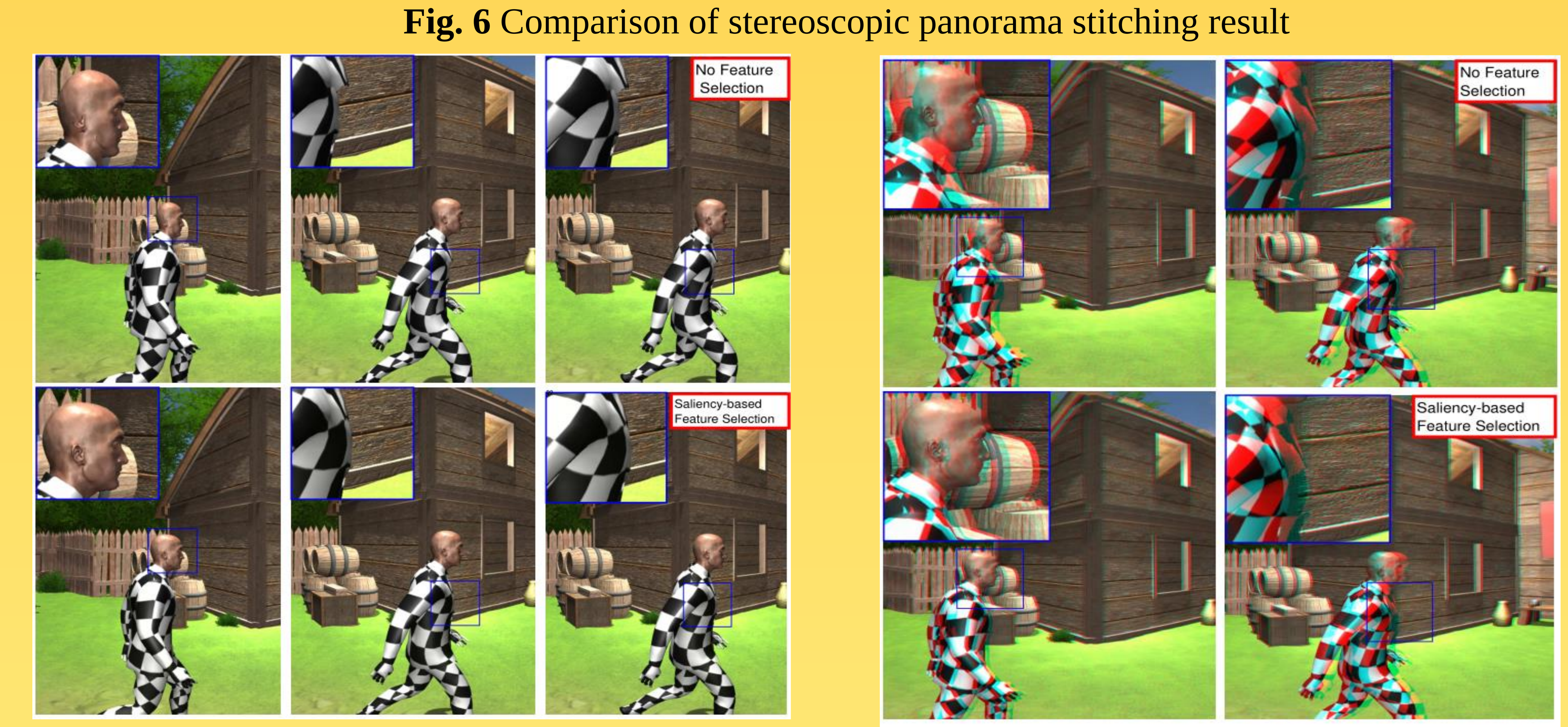


Fig. 7: Comparison of stereoscopic panoramic video stitching result

Quantitative Comparison

Table 1:
Numerical Comparison of panorama case.

	Autopano	PTGui	Hugin	Proposed
RMSE	19.61px	12.39px	9.04px	8.55px
Vertical Disp	0.39°	0.20°	0.31°	0.20°
Horizontal Dist	0.42°	0.47°	0.35°	0.34°

Table 2:
Numerical Comparison of video case

	RMSE	V Disp	H Dist
NFS+1.3m	9.73px	0.19°	0.94°
NFS+2.0m	2.46px	0.13°	0.63°
NFS+3.3m	1.81px	0.10°	0.35°
SFS+1.3m	8.45px	0.05°	0.12°
SFS+2.0m	2.06px	0.06°	0.11°
SFS+3.3m	1.03px	0.05°	0.11°

Conclusion

In this paper, we presented a feature selection and tracking strategy that optimizes the distribution of control points in panoramic video generation system according to the saliency change.

Acknowledgement

This publication is based on work supported in part by the National Science Foundation award CNS-1456638 for SENSEI. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agency.

References

[1] R. Szeliski, “Image alignment and stitching: A tutorial,” Foundationsand Trends R ? in Computer Graphics and Vision, vol. 2, no. 1, pp. 1–104, 2006.
[2] H.-C. Huang and Y.-P. Hung, “Panoramic stereo imaging system with automatic disparity warping and seaming,” Graphical Models and Image Processing, vol. 60, no. 3, pp. 196–208, 1998.
[3] S. Peleg, M. Ben-Ezra, and Y. Pritch, “Omnistereo: Panoramic stereoimaging,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 279–290, 2001.
[4] C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung, “Megastereo: Constructing high-resolution stereo panoramas,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1256–1263.
[5] F. Zhang and F. Liu, “Casual stereoscopic panorama stitching,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2002–2010.
[6] Tzu-Chieh Yen, Chia-Ming Tsai, and Chia-Wen Lin, “Maintaining temporal coherence in video retargeting using mosaic-guided scaling,” IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2339–2351, 2011.
[7] Jonathan Harel, Christof Koch, and Pietro Perona, “Graph-based visual saliency,” in Advances in neural information processing systems, 2007, pp. 545–552.