A NATURAL SHAPE-PRESERVING STEREOSCOPIC IMAGE STITCHING

Haoqian Wang[1,2]*, Yaling Zhou*[1]*, Xingzheng Wang*[1,2]*, Lu Fang*[1,2]

[1]Key Laboratory of Broadband Network & Multimedia, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

[2]Shenzhen Institute of Future Media Technology, Shenzhen 518071, China

ABSTRACT

This paper presents a method for stereoscopic image stitching, which can make stereoscopic images look as natural as possible. Our method combines a constrained projective warp and a shape-preserving warp to reduce the projective distortion and the vertical disparity of the stitched image. In addition to provide a good alignment accuracy and maintain the consistency of input stereoscopic images, we add a specific restriction into the projective warp, which establishes the connection between target left and right images. To optimize the whole warp, a energy term is designed. It can constrain the shape of straight line and vertical disparity. Experimental results on a variety of stereoscopic images can ensure the efficiency of the proposed method.

Index Terms— Stereoscopic image, Image stitching, Image warping, Image distortion, Vertical disparity

1. INTRODUCTION

Image stitching is a well studied topic and most widely used in computer vision and image processing. However, most of the current image stitching methods are designed for monocular image stitching [1–5]. Naively extending monocular image mosaic algorithms to the stereoscopic image stitching may damage the consistency of source stereoscopic images, since two images in a stereoscopic pair contain an extra disparity dimension, known as depth information. Moreover, a small discrepancies in stitched images can bring an uncomfortable viewing experience to viewers, even make them feel dizzy.

Recently, dedicated methods [6, 7] have been proposed to stitch stereo images, which combined part of monocular image stitching algorithms. A user study in [8] shared a similar framework with a monocular mosaic method in using a quad mesh to determine the warp. In addition, he added specific feature correspondence constraints to reduce the perspective distortion. But it is limited by large camera translations. In [9], Yan *et al.* combined the half-projective and content-preserving warp. Although it can reduce the prespective distortion, it breaks the stereoscopic consistency. Zhang *et al.* [10] proposed a representative three-step algorithm to keep the consistency of stitched left and right images by using the target disparity map. Nevertheless, the application of parallax-tolerant method [11] may introduce projective distortion in the left panorama image. As shown in Fig.1(b), resultant images have seriously projective distortion in buildings and trees (detailed in red rectangles). There also exists curved line in the road and vertical disparity in the letter area (detailed in yellow rectangles and enlarged in the top right corner).

Research in [12] shows some issues that greatly affect a natural 3D viewing experience of final stitched stereoscopic images, and suggests to appropriately minimize effects of these factors. A directly way to reach the purpose is computing the binocular parallax to get the depth information at each pixel from input images, then reconstructing the real 3D scene, and projecting source images to a new coordinate system by new camera configuration. However, this method requires lots of calculations and depends much on the accuracy of the 3D reconstruction. Other methods have been proposed for reducing these issues such as keystoning [13], distortion [14] and disparity [15], which carefully control the horizontal disparity to make the image look more pleasant but ignore the vertical disparity.

This paper aims to develop an algorithm for users to create natural stereoscopic images with less distortion and vertical disparity. To this end, we propose a novel method to stitch stereoscopic images, which enormously reduce the projective distortion and can obtain a natural high-quality images. Our method firstly warps the images using a new point constraint combined the target left and right images. Secondly, a shape-preserving warp are accordingly adopted to stitch the images. Finally, we use new energy terms to reduce the vertical disparity, and linearly blend the warped input images to a final stereoscopic image.

The rest of this paper is organized as follows: Section 2 formulates and presents the proposed stereoscopic image stitching method. Experimental results are shown in Section 3. Conclusions are briefly remarked in Section 4.

This work is partially supported by the NSFC fund (61571259, 61531014, 61471213), Shenzhen Fundamental Research fund (JCYJ20160331185006518, JCYJ20170307153051701), Shenzhen Public Technology Platform fund (GGFW2017040714161462)

(a) (b)

Fig. 1. (a) Input left images. (b) The result samples from Zhang's [10].

2. THE PROPOSED METHOD

Our method proposes a new homography matrix to find a warping function that maps pixels from the target images to the reference images. It generates the correspondences from input left images, warped-left and -right image. The new homography matrix can preserve the disparity consistency of the source images with new point constraints. The image rectification based on shape-preserving is used in non-overlapped region to reduce the projective distortion. We note that straight lines have high visual significance, the straight line term and other constraints are utilized to optimize the energy terms.

2.1. Constrained Projective Warp

To present our method in detail, we first denoted $I_1 = (I_l^1, I_r^2)$ and $I_2 = (I_l^2, I_r^2)$ as two input stereopairs. Each stereoscopic image has a left and right image, and I_l^1, I_r^1 are the reference images. As with similar approach of finding point correspondences, we follow [1] to use RANSAC to remove outliers from matched SIFT feature points. The homography matrix H_1 is estimated by the point correspondences from (I_l^1, I_l^2) (I_r^1, I_r^2) and (I_l^2, I_r^2). Sets of point-wise matches are defined as:

$$\begin{aligned}(I_l^1, I_l^2) =& \{(p_l^1(i), p_l^2(i)) | p_l^1(i) = [x_l^1(i)\ y_l^1(i)]^T, \\ & p_l^2(i) = [x_l^2(i)\ y_l^2(i)]^T, i = 1 \cdots n_1\}\end{aligned} \tag{1}$$

$$\begin{aligned}(I_r^1, I_r^2) =& \{(p_r^1(j), p_r^2(j)) | p_r^1(j) = [x_r^1(j)\ y_r^1(j)]^T, \\ & p_r^2(j) = [x_r^2(j)\ y_r^2(j)]^T, j = 1 \cdots n_2\}\end{aligned} \tag{2}$$

$$\begin{aligned}(I_l^2, I_r^2) =& \{(p_l^2(k), p_r^2(k)) | p_l^2(k) = [x_l^2(k)\ y_l^2(k)]^T, \\ & p_r^2(k) = [x_r^2(k)\ y_r^2(k)]^T, k = 1 \cdots n_3\}\end{aligned} \tag{3}$$

where n_1, n_2 and n_3 denote the number of matched feature points. $(p_l^1(i), p_l^2(i))$, $(p_r^1(j), p_r^2(j))$ and $(p_l^2(k), p_r^2(k))$ are sets of all matched feature points in (I_l^1, I_l^2), (I_r^1, I_r^2) and (I_l^2, I_r^2), respectively. Let the first left imageI_l^1 be the reference image. So formulate the transformation $p_l^1 = Hp_l^2$:

$$\begin{bmatrix} x_l^1 \\ y_l^1 \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} x_l^2 \\ y_l^2 \\ 1 \end{bmatrix} \tag{4}$$

where the transformation H is donoted in a 3×3 vector, which can be estimated using DLT in APAP warp [4] from the weighted problem. The same in right images transformation. So for each grid mesh, a local homography is

$$h_k = arg \min_{h_k} \sum_{i=1}^{n_1} \omega_k^i \left\| a_i h \right\|^2,\ s.t. \left\| h \right\| = 1 \tag{5}$$

where a_i is the two linearly independent rows to contain monomials. The weight ω_k^i describes the influence of each pair of point correspondence on the $i^t h$ grid, i.e., $w_k^i = max(exp(-\left\| p_l^2(k) - p_l^2(i) \right\|^2 / \sigma^2), \lambda)$. The σ is the scale parameter and $\lambda \in [0, 1]$ is used to prevent numerical issues.

To keep the consistency of the stereoscopic image pairs, we introduce the energy terms E_p to limit the local homography. The energy term can be calculated as:

$$\begin{aligned} E_p = & \frac{1}{n_1} \sum_{m=1}^{n_1} \gamma_l(m) \left\| \frac{1}{\varphi_m} h p_l^2(m) - p_l^1(m) \right\| \\ & + \frac{1}{n_2} \sum_{s=1}^{n_2} \gamma_r(s) \left\| \frac{1}{\varphi_s} h p_r^2(s) - p_r^1(s) \right\| \\ & + \frac{1}{n_3} \sum_{q=1}^{n_3} \left\| \frac{1}{\varphi_q} h p_l^2(q) - p_r^2(q) \right\| \end{aligned} \tag{6}$$

where $\gamma_l(m)$ and $\gamma_r(m)$ are binary values. If matched feature points are in the left images, $\gamma_l(m) = 1$, otherwise, it equals to 0. $\gamma_r(m)$ is denoted likewise for the right image. φ_m, φ_s and φ_q are the weight parameters, expressed as Euclidean distance, i.e., $\varphi_m = \sum_{i=1}^{n_1} \left\| p_l^2(m) - p_l^2(i) \right\|^2$, similarly, φ_s and φ_q can be obtained.

For each local homography in equation (5), we calculate its energy term using equation (6). So loop the process until the iterative time equals to the initial set value, initialized as 2000. The new homography transformation H_1 can be obtained with the minimum energy value E_p. Then I_l^2 and I_r^2 can be transformed into the coordinate system of reference images I_l^1 and I_r^1. We define transformed target images as $I_l^{2\prime}$ and $I_r^{2\prime}$.

2.2. Shape-Preserving Warp

After applying the new homography matrix H_1 to warp the target image, the overlapping regions of two left images are better aligned compared to APAP [4] and global transformation, as shown in Figure 2 (red and yellow boxes). Nevertheless, the result suffers from obvious distortions in the non-overlapping regions, like stretched shapes and non-uniform scaling. Various previous work uses content-preserving method to fine-tune images [7, 16], but this still can not meet the requirement. For the non-overlapping regions, A global similarity transformation is used to preserve the shape.

In view of Chum *et al.* [17], changing the coordination benefits to expose the distortion characteristics in projective transformation. So the original coordination of the target image (x^2, y^2) is rotated to to the new coordinate system (u, v), denoted as:

$$\begin{bmatrix} x_l^2 \\ y_l^2 \end{bmatrix} = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{7}$$

where $\theta = arctan(h_8/h_7)$. Then put the Eq. (6) into Eq. (3) to construct the mapping between (x_l^1, y_l^1) and (u, v):

$$\begin{bmatrix} x_l^1 \\ y_l^1 \end{bmatrix} = \frac{1}{1-cu} \begin{bmatrix} c_1 h_1 + c_2 h_2 + h_3 \\ c_1 h_4 + c_2 h_5 + h_6 \end{bmatrix} \tag{8}$$

where $c = \sqrt{h_7^2 + h_8^2}$, $c_1 = ucos\theta - vsin\theta$ and $c_2 = usin\theta + vcos\theta$. Then divide the overlapping and non-overlapping regions by the line $u = u_*$, learned from [18], which combines the constrained projective warp in overlapping areas with the shape-preserving transform H_2 in non-overlapping areas. To remain the continuous of the two warps in the region boundary, let $H_2(u_*, v) = H_1(u_*, v)$ and the shape-preserving warp H_2 is solved as following:

$$\begin{aligned} H_2(u,v) = \frac{1}{1-cu_*}(&\begin{bmatrix} (h_5cos\theta - h_4sin\theta)(u-u_*) \\ (h_1sin\theta - h_2cos\theta)(u-u_*) \end{bmatrix} \\ &+ \begin{bmatrix} h_1c_1 + h_2c_2 + h_3 \\ h_4c_1 + h_5c_2 + h_6 \end{bmatrix}) \end{aligned} \tag{9}$$

2.3. Energy Optimization

In this section, target images are divided into a grid mesh, so we can transfer the mesh warping problem into an energy terms minimization problem containing alignment, similarity, straight line and disparity constrains. To keep the natural shape of source images, straight line and vertical disparity constrains are especially added. The final warp is fine-tuned by minimizing total energy terms. Next, total energy terms are depicted in detail.

Straight line term: To keep straight after warping, the straight line term is introduced. We first use Hough transform to extract an infinite line (ρ, α), where $\rho \geq 0$ and $\alpha \in [0, 2\pi]$. Hence, a pair of matched corresponding lines are $(s_l(\alpha_l, \rho_l), s_r(\alpha_r, \rho_r))$, where $0 < |\alpha_l - \alpha_r| < 20°$. The straight line term is written as:

$$E_{sl} = \int_s \left\| \sum_s [sin\widetilde{\alpha}_s, cos\widetilde{\alpha}_s] \cdot \begin{bmatrix} c\ cos\overline{\alpha}_s \\ -sin\overline{\alpha}_s \end{bmatrix} \right\|^2 ds \tag{10}$$

where $\overline{\alpha}_s$ is the orientation of the warped line segment. $\widetilde{\alpha}_s$ is the target orientation of the segment s_l or s_r.

Disparity term: Vertical and horizontal terms are used to constrain the whole disparity. The disparity of both warped feature points in image pairs should be as close as possible to the original disparity.

$$E_d = \sum_{j=1}^{n_3} \left\| p_{l,y}^2(j) - p_{r,y}^2(j) \right\|^2 + \sum_i \left\| D_i - d_i \right\|^2 \tag{11}$$

where $D_i = \overline{p}_{l,x}^2(i) - \overline{p}_{r,x}^2(i)$, represented the difference between abscissa of warped left and right images. d_i is the target horizontal value computed from the I_l and I_r.

Alignment and smooth terms: These two terms are designed to better align warped and reference images and keep the shape. The alignment term is defined below.

$$\begin{aligned} E_g = &\sum_{i=1}^{n_1} \lambda_i \left\| \overline{p}_l^2(i) - p_l^1(i) \right\|^2 \\ &+ \sum_{i=1}^{n_2} \mu_i \left\| \overline{p}_r^2(i) - p_r^1(i) \right\|^2 \end{aligned} \tag{12}$$

λ_i is a binary value. If the warped feature points are in left images, we set it 1, otherwise it is 0. The μ_i is obtained as the same way for right images. As the Hessian of the warping function should be zero, we use the constraints e_s in [19] as $E_s = \omega e_s$ to maintain smooth.

The final energy term can be enumerated as:

$$E = \beta_1 E_g + \beta_2 E_s + E_{sl} + E_d \tag{13}$$

β_1 and β_2 are the weighted parameters initialized as 0.8 and 0.3.

Finally, the α blending method is applied to create stitched left-view and right-view images. The final red-cyan anaglyph with the proposed warp is shown in Fig.2(b).

(a) (b)

Fig. 2. (a) Above:the result by warping with global homography. Below: the result of APAP [4]. (b) Above: our result with constrained projective warp. Below: final result of the proposed warp. Some details are highlighted and enlarged

3. RESULTS

In this section, a set of experiments are conducted in MATLAB R2014b development environment on a PC with a 3.20GHz CPU and 4GB RAM. Figure 3 shows part of the test images on stereoscopic image dataset[1] provided by Zhang et.al [10] using stereo cameras. To evaluate the efficiency of our method, we compared our method with other approaches, including APAP [4], AutoStitch [1], Zhang [10] and Yan's [7].

Figure 4 compares the results between APAP, AutoStitch and our method. We can see that APAP in cannot guarantee a good alignment in the overlapping regions, such as ghosts are occur in the short sleeve and hat (yellow boxes). There also exists large vertical disparity in APAP (orange boxes). Both APAP and AutoStitch cannot keep the straight shape of the street lamp detailed in blue boxes.

[1]http://web.cecs.pdx.edu/ zhangfan/stereostitch/index.html

Figure 5 compares Zhang's method [10] and the proposed method. Zhang's method (Fig. 5(a)) could align the input images without ghosts but ignored distortion (red boxes). As shown in the last column, there are vertical disparity in the enlarged region. In contrast, our method (Fig. 5(b)) can deliver a natural viewing experience with less distortion and vertical disparity.

Fig. 3. Test Images. Top to bottom: Building, Monument, Shop, Room and Campus.

Figure 6 shows final anaglyph versions of the stereopairs by Yan's method [7] and our method. In Fig. 6(a), Yan's results could lead to a good result with little vertical disparity. However, there are some distortions in edge areas. The proposed method (Fig. 6(b)) lessens the projective distortion.

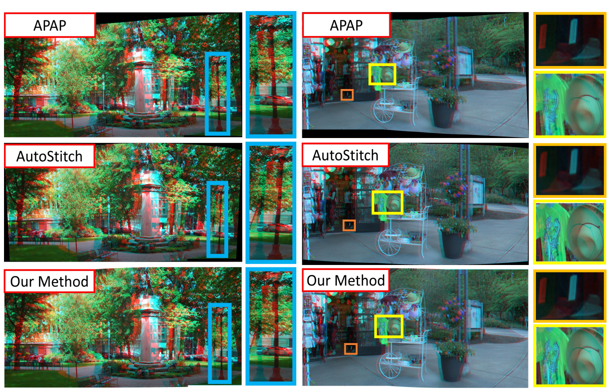


Fig. 4. Comparison results with APAP [4] and AutoStitch [1]. From top to bottom: APAP, AutoStitch, and Our method. For better comparison, some details are highlighted.

In our experiments, we calculates the average absolute vertical disparity (AVD) to evaluate the quality of stereoscopic images. Here, the AVD is defined as:

$$AVD = \frac{1}{n}\sum_{i=1}^{n}|VD_i| \tag{14}$$

where m is the number of matched feature points, and VD_i denotes the absolute vertical disparity of the i^{th} matched feature points between stitched left and right image. In our experiments, we use SIFT to extract 5000 feature points. Table 1 shows the AVD of all test images in Fig. 3. From the table,

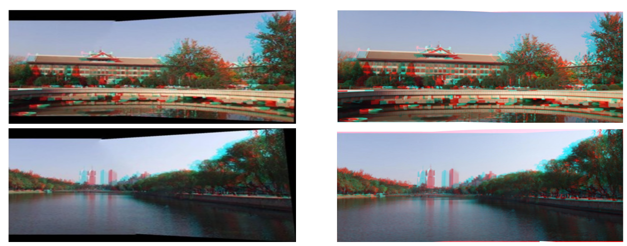
Table 1. Average Vertical Disparity(/pixel)

Dataset	AutoStitch	APAP	Zhang's	Proposed
Building	3.42	6.13	1.74	1.59
Monument	4.34	5.22	1.89	1.32
Shop	2.67	3.64	1.54	1.27
Room	2.06	4.80	1.07	0.91
Campus	7.96	8.69	1.98	1.68

APAP has the largest vertical disparity, while AutoStitch and Zhang's method have relatively small value. The proposed method can reduce the vertical disparity.

(a) (b)

Fig. 5. Comparison results with Zhang's [10]. (a) Results of Zhang's. (b) Our results.



(a) (b)

Fig. 6. Comparison results with Yan's [7]. (a) Anaglyph images of Yan's. (b) The result of our method.

4. CONCLUSION

This paper describes a natural shape-preserving stereoscopic image stitching method, which can create natural-looking anaglyph images and deliver a pleasent viewing experience to viewers. The proposed approach combines the constrained projective warp with the shape-preserving warp, which can handle vertical disparity and distortion to some extent. It weakly relies on the choice of the boundary of two warping regions. Experimental results have shown that the proposed method can produce a natural stereoscopic image.

5. REFERENCES

[1] Matthew Brown and David G Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, no. 1, pp. 59–73, 2007.

[2] Junhong Gao, Seon Joo Kim, and Michael S Brown, "Constructing image panoramas using dual-homography warping," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 49–56.

[3] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong, "Smoothly varying affine stitching," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 345–352.

[4] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter, "As-projective-as-possible image stitching with moving dlt," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2339–2346.

[5] Yu-Sheng Chen and Yung-Yu Chuang, "Natural image stitching with the global similarity prior," in *European Conference on Computer Vision*. Springer, 2016, pp. 186–201.

[6] Ho-Chao Huang and Yi-Ping Hung, "Panoramic stereo imaging system with automatic disparity warping and seaming," *Graphical Models and Image Processing*, vol. 60, no. 3, pp. 196–208, 1998.

[7] Weiqing Yan, Chunping Hou, Jianjun Lei, Yuming Fang, Zhouye Gu, and Nam Ling, "Stereoscopic image stitching based on a hybrid warping model," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[8] Song-Pei Du, Shi-Min Hu, and Ralph R Martin, "Changing perspective in stereoscopic images," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 8, pp. 1288–1297, 2013.

[9] Weiqing Yan and Chunping Hou, "Reducing perspective distortion for stereoscopic image stitching," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[10] Fan Zhang and Feng Liu, "Casual stereoscopic panorama stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2002–2010.

[11] Fan Zhang and Feng Liu, "Parallax-tolerant image stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3262–3269.

[12] Andrew J. Woods, Tom Docherty, and Rolf Koch, "Image distortions in stereoscopic video systems," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 1915, pp. 36–48, 2009.

[13] Lew B Stelmach, Wa James Tam, Filippo Speranza, Ron Renaud, and Taali Martin, "Improving the visual comfort of stereoscopic images," in *Proc. SPIE*, 2003, vol. 5006, pp. 269–282.

[14] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross, "Nonlinear disparity mapping for stereoscopic 3d," in *ACM Transactions on Graphics (TOG)*. ACM, 2010, vol. 29, p. 75.

[15] Yuzhen Niu, Wu-Chi Feng, and Feng Liu, "Enabling warping on stereoscopic images," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 183, 2012.

[16] Che-Han Chang, Chia-Kai Liang, and Yung-Yu Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 589–601, 2011.

[17] Ondřej Chum, Tomáš Pajdla, and Peter Sturm, "The geometric error for homographies," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 86–102, 2005.

[18] Che-Han Chang, Yoichi Sato, and Yung-Yu Chuang, "Shape-preserving half-projective warps for image stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3254–3261.

[19] Robert Carroll, Aseem Agarwala, and Maneesh Agrawala, "Image warps for artistic perspective manipulation," in *ACM Transactions on Graphics (TOG)*. ACM, 2010, vol. 29, p. 127.