Vanishing Point Guided Natural Image Stitching

Kai Chen, Jian Yao, Jingmin Tu, Yahui Liu, Yinxuan Li, and Li Li
School Remote Sensing and Information Engineering, Wuhan University

***Abstract*—Recently, works on improving the naturalness of stitching images gain more and more extensive attention. Previous methods suffer the failures of severe projective distortion and unnatural rotation, especially when the number of involved images is large or images cover a very wide field of view. In this paper, we propose a novel natural image stitching method, which takes into account the guidance of vanishing points to tackle the mentioned failures. Inspired by a vital observation that mutually orthogonal vanishing points in Manhattan world can provide really useful orientation clues, we design a scheme to effectively estimate prior of image similarity. Given such estimated prior as global similarity constraints, we feed it into a popular mesh deformation framework to achieve impressive natural stitching performances. Compared with other existing methods, including APAP, SPHP, AANAP, and GSP, our method achieves state-of-the-art performance in both quantitative and qualitative experiments on natural image stitching.**

***Index Terms*—Natural Image Stitching, Vanishing Point, Global Similarity Prior.**

I. Introduction

Image stitching is a classical computer vision task that combines multiple images into a panorama with a wider field of view. Methods started to flourish since 2007, when Brown and Lowe [1] proposed to use SIFT features [2] to fit a global homography model for stitching. Since then, various methods were developed to further improve the stitching performance, including the spatially varying methods [3], [4] with a higher degree of freedom for good alignment accuracy, and the combined-constraints based methods [5], [6], [7] for improving stitching robustness. The technique of mesh deformation [8] is also adopted to stitching since it has high alignment quality and is highly scalable to some specific stitching purposes, such as stereoscopic stitching [9], [10], [11]. Unfortunately, these methods usually can not deal with the non-overlapping region well owing to the lack of effective constraints, and as a result, they suffer severe projective distortions. Once number of images gets too large, the distortion would accumulate and propagate among images, leading to unnatural rotation, scaling and stretch.

Natural image stitching methods [12], [13], [14], [15] are developed to reduce distortions. Up till now, a widely acceptable fact for natural stitching is to make use of the shape-preserving property [16] provided by the similarity transformation. Moreover, the concept of image global similarity prior (*a.k.a.*, a scale factor s and a 2D rotation angle θ) [13] is proposed to be estimated to further improve the visual quality. Inappropriate rotations between adjacent images often induce obvious unnaturalness in results[1] (as shown in Figure 1). Thus, the key issue for natural stitching becomes how to properly estimate the image similarity prior. Most existing methods [12], [15], [13] determine them with matched feature points or line segments. Note that these schemes merely utilize the pairwise correspondences between adjacent images. The absence of a global constraint, which would offer robust guidance for the similarity prior estimation, makes these methods unstable when the overlap between images is small, or the number of the involved images is large. As a result, these methods produce unnatural artifacts.

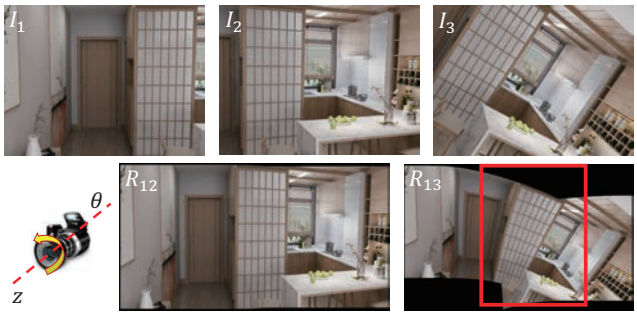


Figure 1. Top: Two pairs of input images (I_1, I_2) and (I_1, I_3). They share the same left image I_1, while I_2 and I_3 differ merely by a 2D rotation θ. Bottom: Two panoramas produced by the SPHP [12]. R_{12} and R_{13} have very different appearances caused by θ.

Considering the mentioned issues, in this paper, we propose to take the vanishing point (VP) as an effective global constraint, and develop a novel similarity prior estimation method for natural image stitching. We focus on the problem of estimating θ, and exploit the VP guidance by taking its two advantages: (1) utilizing the orientation clues from VPs to estimate the initial 2D rotations for input images; (2) making use of the global consistency of VPs in Manhattan world, by which a novel scheme is proposed to estimate the prior robustly. After that, the determined similarity prior is feed into a mesh deformation framework as global similarity constraints to stitch multiple images into a panorama with a natural look.

In summary, the main contributions of this paper are:

- We design a robust scheme to determine the image similarity prior from the VP clues of the scene, based on which a novel natural stitching method named VPG is developed to improve the naturalness of output panoramas significantly.
- We provide a degradation mechanism to make the proposed VPG can be well-used in general scenes. When the scene holds the Manhattan assumption, VPG manages to produce a more natural panorama than other methods. Otherwise, it automatically falls back to a standard stitch-

[1]For the presented two-image case, the zero-rotation switch in SPHP was turned on to illustrate the problem. In practical application, the switch has limited effect in improving panorama naturalness for general multiple-image cases. Relevant results will be presented in the following of the paper.

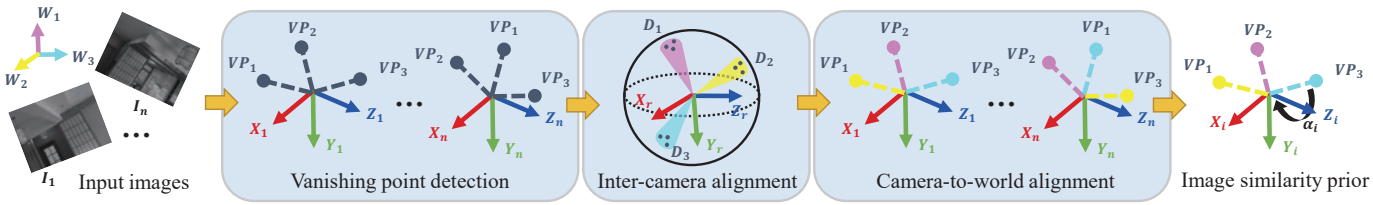


Figure 2. A flowchart for initial rotation estimation $\{\alpha_i\}_{i=1}^N$. VPs are first detected through line segments extraction and clustering. Then, VPs of different images are aligned on a unified sphere surface, on which three dominant VP directions D are estimated. Finally, VPs are associated with the ideal vanishing directions by D to compute $\{\alpha_i\}_{i=1}^N$.

ing scheme. The output still has a relatively natural look that is not affected by wrong VP guidance.
- We conducted more analyses upon the proposed VPG algorithm. The results further reveal that VPG has two additional good properties: First, it is not influenced by different reference selections; Second, it is compatible with other high-alignment-accuracy stitching frameworks to achieve a coordination between good naturalness and high alignment accuracy.

Abundant quantitative and qualitative experiments both demonstrate that the proposed vanishing point guided method (VPG) outperforms the state-of-the-art methods, including APAP [4], SPHP [12], AANAP [15], and GSP [13].

II. Related Work

We briefly review the most related works in three aspects: mesh-deformation based stitching methods, natural image stitching methods, and some existing successful practices in other fields that are associated with the VP guidance.

A. Mesh-deformation based Image Stitching

Mesh deformation technique [8] is adopted to image stitching since its flexibility. It first divides input images into a series of uniform grid meshes, and then estimates deformed mesh vertexes by minimizing an objective function. Various constraints are utilized to build the objective function in order to improve robustness and stitching quality. Early feature point based methods [17], [11], [18] detected and matched key-points in the overlapping region, then they achieved alignment by warping matched points to close positions. Li *et al.* [5] tailored a dual-feature model that considers both key-point and line segment correspondences to perform a robust stitching for low-textured scenes. Lin *et al.* [6] were inspired by optical flow estimation and performed alignment by minimizing the overall intensity difference among regularly sampled points. In addition, Xiang *et al.* [19] locally regulated image content by penalizing the straightness of line segments, and Lin *et al.* [20] preserved image structures through maintaining the contour shapes. Although these mentioned methods achieve good pairwise alignment accuracy, they are not suited for natural stitching since a lack of valid constraints for the non-overlapping region and the global image content.

B. Natural Image Stitching

Most existing methods [21], [12], [13], [15], [10], [22]achieved natural stitching by integrating the similarity transformation into a spatially varying homography model. The key issue is how to estimate the global similarity prior for each input image. SPHP [12] determined the similarity transformation through analyzing the pairwise image homography. Then, it enhanced the naturalness via smoothly changing the stitching model from projective to similarity transformation. AANAP [15] first computed a bunch of 2D rotations by feature matching, then it empirically selected the similarity transformation with the smallest rotation angle as the optimal one. GSP [13] solved the global similarity prior by feature matching as well as the 3D rotation relationship between adjacent images. VL [22] built similarity constraints for orthophotos by taking use of the orientations of line segment clusters. Considering that all these methods lack an effective global guidance, they suffer failures in challenging scenes.

C. Vanishing Point Guidance

Vanishing point (VP) [23] is widely adopted in computer vision tasks since its predominance in two aspects. First, VPs contain strong orientation and geometric clues of a scene which could be useful guidance. Lee *et al.* [24] interpreted the scene structure from VPs extracted from a single image. Lee and Yoon [25] recovered the camera orientation with a joint estimation of VPs. Huang *et al.* [26] exploited VPs in image completion by detecting the planar surfaces and regularity with VPs. Furthermore, VPs were applied in layout estimation [27] to pre-align the image to be level with the floor, and also were utilized to offer geometric context for road detection and recognition [28]. Second, VPs are globally consistent in Manhattan world, and therefore provide an effective global constraint for many optimization-based problems which help to yield stable performance for robot navigation [29]. Camposeco and Pollefeys [30] adopted VPs to improve the accuracy of visual-inertial odometry, and Li *et al.* [31] leveraged them to build a robust monocular SLAM system. Inspired by these successful applications of VPs, we apply them to natural stitching, and propose a robust similarity prior estimation method making use of the VP guidance, as shown in Figure 2.

III. Our Methodology

Given N images $\{I_i\}_{i=1}^N$, we stitch them into a natural panorama using the mesh deformation framework as described

previously. Let V be the set of vertexes in the uniform grid mesh that placed on input images, the stitching process is formulated into a mesh deformation problem through finding the optimal warped vertex set $\hat{V}$. Usually, it turns into an optimization problem by minimizing an objective function with the following classical form [21], [5], [13]:

$$E(V) = E_a(V) + E_l(V) + E_g(V), \tag{1}$$

where E_a, E_l and E_g denote the alignment term, the local shape-preserving term and the global similarity term respectively.

As mentioned, previous methods improve alignment accuracy and conserve image content by regulating E_a and E_l, while recent natural stitching methods focus on E_g to produce more natural looking panoramas. E_g is built based on the image similarity priors (a scale factor s and a 2D rotation θ), but existing methods fail to offer a robust scheme to estimate them. They usually treat each input image separately [22], or determine the prior merely with pairwise image correspondence information [13], [12], [15], which are not robust enough in practical applications. Therefore, we focus on developing a robust similarity prior estimation method with the guidance of VPs, in which we mainly focus on the estimation for 2D rotations $\{\theta_i\}_{i=1}^N$.

A. Rotation Estimation with VP Guidance

Let $\mathbf{G}$ be the stitch graph of N input images. $\mathbf{J}$ denotes its edge set, in which each edge (i, j) corresponds to a pair of adjacent images (I_i, I_j). In general, the $\mathbf{G}$ could be manually specified or automatically verified by the probabilistic model [1]. Let's define $\mathbf{P} = \{\mathbf{P}_{ij} | (i, j) \in \mathbf{J}\}$ as the set of matched feature points. We apply LSD [32] to detect line segments on I_i, and then loosely follow the scheme in [24] to find three orthogonal VPs $[\mathbf{v}_1^i, \mathbf{v}_2^i, \mathbf{v}_3^i]$ without the intrinsic parameters. As illustrated in Figure 2, we subsequently obtain the initial rotation estimation $\{\alpha_i\}_{i=1}^N$ through two steps: (1) *Inter-camera alignment*. VPs from different images are first aligned in the same reference coordinate system, in which we can estimate dominant directions using these roughly aligned VPs. (2) *Camera-world alignment*. VPs are associated with the ideal vanishing directions in Manhattan world to produce initial rotation estimation $\{\alpha_i\}_{i=1}^N$. After that, we design a robust estimation scheme to get optimal rotation result $\{\theta_i\}_{i=1}^N$.

1) Inter-Camera Alignment.: We project VPs of different images onto a unified sphere surface in order to achieve alignment. As a preparation, based on the matched point set $\mathbf{P}$, we first estimate the 3D rotation $\mathbf{R}_i$ for each image I_i by bundle adjustment method [1]. After that, without the loose of generality, we regard I_r as a reference and set $\mathbf{W} = \mathbf{R}_r$. VPs then are projected as:

$$[\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i] = \mathbf{W}\mathbf{R}_i^{-1}[\frac{\mathbf{v}_1^i}{\|\mathbf{v}_1^i\|}, \frac{\mathbf{v}_2^i}{\|\mathbf{v}_2^i\|}, \frac{\mathbf{v}_3^i}{\|\mathbf{v}_3^i\|}]. \tag{2}$$

Assuming that the scene satisfies the Manhattan world and considering the possible parallax and estimation noises, these aligned VPs should roughly congregate around three dominant directions $\mathbf{D}_{3\times3}$. We propose to estimate $\mathbf{D}_{3\times3}$ by:

$$\mathbf{D} = \arg\min_{\mathbf{D}} \sum_{i=1}^{N} \|\mathbf{D}_{3\times3} - [\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i]\|_F^2, \tag{3}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm. In order to improve the stableness, we decompose $\mathbf{D}$ into $\mathbf{D}_0\mathbf{D}_t$, where $\mathbf{D}_0$ is the initial dominant direction hypothesis, and $\mathbf{D}_t$ is a 3×3 relative rotation matrix. We collect two roughly orthogonal VPs as the first two directions $\mathbf{d}_1$ and $\mathbf{d}_2$ in $\mathbf{D}_0$. They produce the third direction by $\mathbf{d}_3 = \mathbf{d}_1 \times \mathbf{d}_2$. After that, $\mathbf{d}_2$ is revised by $\mathbf{d}_2 = \mathbf{d}_1 \times \mathbf{d}_3$ to ensure the orthogonality so that we get a complete dominant direction hypothesis $\mathbf{D}_0 = [\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3]$. Given $\mathbf{D}_0$, we can estimate $\mathbf{D}_t$ by:

$$\mathbf{D}_t = \arg\min_{\mathbf{D}_t} \sum_{i=1}^{N} \|\mathbf{D}_0\mathbf{D}_t - [\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i]\|_F^2. \tag{4}$$

It is obvious that solving Equation 4 is much more stable than directly optimizing Equation 3. It can be effectively optimized by Gauss-Newton iteration method. After traversing all possible dominant direction hypotheses, we obtain the final result with the minimal residual error as the optimal dominant direction $\hat{\mathbf{D}} = [\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \hat{\mathbf{d}}_3]$.

2) Camera-to-World Alignment.: Let the three global VPs associated with three dominant directions in Manhattan world be $[\mathbf{v}_1^{\mathrm{w}}, \mathbf{v}_2^{\mathrm{w}}, \mathbf{v}_3^{\mathrm{w}}] = [[1, 0, 0]^{\mathsf{T}}, [0, 1, 0]^{\mathsf{T}}, [0, 0, 1]^{\mathsf{T}}]$. We assume that people rarely twist the camera severely relative to the horizon when capturing a picture, which can be a relative loose assumption than the one in [1] and [13]. Hence, we link $\hat{\mathbf{D}}$ with $[\mathbf{v}_1^{\mathrm{w}}, \mathbf{v}_2^{\mathrm{w}}, \mathbf{v}_3^{\mathrm{w}}]$ by:

$\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3]$, which is determined by

$$\mathbf{m}_i = \arg\min_{\mathbf{v}_j^{\mathrm{w}}} \|\hat{\mathbf{d}}_i - \mathbf{v}_j^{\mathrm{w}}\|^2, i, j = 1, 2, 3. \tag{5}$$

Meanwhile, we rearrange $[\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i]$ for I_i in order to make them correspond to $\hat{\mathbf{D}}$, that is, $\hat{\mathbf{v}}_k^i \leftrightarrow \hat{\mathbf{d}}_k, k = 1, 2, 3$. Then, the transformation from the i-th camera to the global Manhattan world is formulated as:

$$\mathbf{R}_i^{\mathrm{w}} = \mathbf{M}[\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i]^{-1}, \tag{6}$$

where $\mathbf{R}_i^{\mathrm{w}}$ is a 3×3 rotation matrix. We decompose it to get α_i, which is a 2D rotation angle with respect to z axis.

3) Robust Estimation.: Instead of directly taking α_i as θ_i, we propose to estimate $\{\theta_i\}_{i=1}^N$ through minimizing an objective function with regarding α_i as the data term. In addition, we consider the relative rotation between adjacent images as the smoothness term. Specifically, for image pair (I_i, I_j), the relative 2D rotation $\beta_{i,j}$ is obtained by decomposing $\mathbf{R}_j\mathbf{R}_i^{-1}$. We represent α_i by a unit 2D vector $(\phi_i, \omega_i)^{\mathsf{T}}$, and the rotation $\{\theta_i = (u_i, v_i)^{\mathsf{T}}\}_{i=1}^N$ are obtained by minimizing:

$$\sum_{i=1}^{N} \| \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} \phi_i \\ \omega_i \end{bmatrix} \|^2 + \lambda \sum_{(i,j)\in\mathbf{J}} \|\mathbf{R}(\beta_{i,j}) \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} u_j \\ v_j \end{bmatrix} \|^2, \tag{7}$$

where $\mathbf{R}(\beta_{i,j})$ denotes the $2D$ rotation matrix specified by $\beta_{i,j}$, and $\lambda = 10.0$ is a balance weight. In Equation 7, we

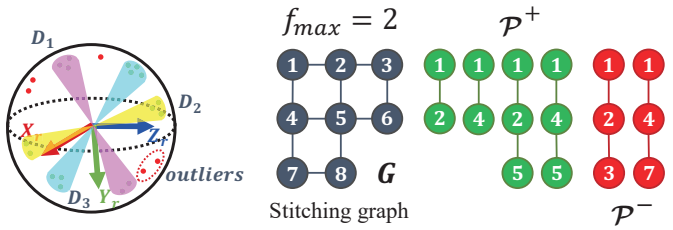


Figure 3. A illustration of two strategies adopted for robustly estimating $\{\theta_i\}_{i=1}^N$. Left: Given the VPs alignment result and dominant directions, images whose VPs has large residuals are marked as outliers. Right: Initial rotations are propagated along the stitching graph, and are weighted using a path voting scheme.

consider each α_i equally, which is easily affected by possible noises existing in $\{\alpha_i\}_{i=1}^N$. Therefore, as shown in Figure 3, two strategies are further designed to deal with this problem.

Outlier rejection. Previously, we have roughly aligned VPs of different images and have obtained their three dominant directions $\hat{\mathbf{D}}$. Starting from the global consistency of VPs in Manhattan world, we compute the residual difference e_i for I_i according to its $[\hat{\mathbf{v}}_1^i, \hat{\mathbf{v}}_2^i, \hat{\mathbf{v}}_3^i]$ and $\hat{\mathbf{D}}$. The image whose e_i is larger than a threshold τ is marked as an outlier. We only compute α_i for inliers, and adopt them as the data term in Eq. 7 to estimate the image similarity prior.

Path voting. The scheme of outlier rejection actually is a hard constraint. It has to face a dilemma that a too small τ may wrongly filter out many images. Otherwise, it may introduce some fallacious α_i. A path voting scheme acts as a soft constraint to cope with this dilemma.

As shown in Figure 3, given the stitching graph $\mathbf{G}$ with N images, we collect all valid paths $\mathcal{P}_i = \{p_i^k\}_{k=1}^{c_i}$ within a maximal length f_{max} for each inlier image I_i, where c_i denotes the total number of valid paths for I_i. p_i^k starts from I_i and ends at its neighboring inlier images. A path is valid only when it does not pass through any outlier image. We then divide $\mathcal{P}_i$ into two parts: the supporting set $\mathcal{P}_i^+$ and the opposing set $\mathcal{P}_i^-$. Such a division is achieved based on the relative rotation angle $\beta_{i,j}$ between adjacent images and by judging each p_i^k whether it supports the estimation result α_i for image I_i or not. For $p_i^k \in \mathcal{P}_i^+$, we directly take its path length $L(p_i^k)$ as the supporting length $S(p_i^k)$. Otherwise, its opposing length $O(p_i^k)$ is calculated as $f_{max} + 1 - L(p_i^k)$. Then, we weight the corresponding data term with ψ_i, which is defined as:

$$\psi_i = \sigma\left(\frac{\sum_{p\in\mathcal{P}_i^+} S(p)}{\sum_{p\in\mathcal{P}_i^+} S(p) + \sum_{p\in\mathcal{P}_i^-} O(p)}\right), \tag{8}$$

where $\sigma(\cdot)$ is a sigmoid-form non-linear mapping function. After the above two schemes, the data term in Eq. 7 is developed into:

$$\sum_{I_i\in\Psi} \psi_i \| \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} \phi_i \\ \omega_i \end{bmatrix} \|^2, \tag{9}$$

where Ψ denotes inlier images. $\{\theta_i\}_{i=1}^N$ are obtained by minimizing Equation 7.

B. Stitching by Mesh Deformation

After collecting $\{\theta_i\}_{i=1}^N$, we need to determine the scale factor $\{s_i\}_{i=1}^N$ to build a complete global similarity constraint. Chen and Chuang [13] propose to estimate s_i by the ratio of focal length, yet this scheme relies too much on camera intrinsic matrix. In contrast, we resort to the matched point set $\mathbf{P} = \{\mathbf{P}_{ij} | (i,j) \in \mathbf{J}\}$ to determine s_i. For a pair of adjacent images (I_i, I_j), we first estimate their relative scale η^{ij} by:

$$\eta^{ij} = \frac{c(h_i)}{c(h_j)}, \tag{10}$$

where h_i and h_j are two convex hulls that are determined by $\mathbf{P}_{ij}$, and $c(\cdot)$ returns the perimeter of a convex hull. After that, we estimate the absolute scale s_i for I_i by solving:

$$\arg\min_s \sum_{(i,j)\in\mathbf{J}} \|\eta^{ij} s_j - s_i\|^2, s.t. \sum_{i=1}^N s_i = N. \tag{11}$$

It is a quadratic constrained minimization problem, and can be efficiently solved by any linear system. With $\{\theta_i\}_{i=1}^N$ and $\{s_i\}_{i=1}^N$, we take the deformation objective function (Eq. 1) from [13] as our baseline, but boost its E_g with our improved image similarity prior. The final stitched panorama eventually is generated by texture mapping technique.

IV. Experiments and Results

In this section, we compared our proposed VPG with four state-of-the-art methods: APAP [4], SPHP [12], AANAP [15], and GSP [13]. Besides the widely used qualitative comparison manner, two metrics were designed based on the collected synthetic image sets to quantitatively asses the panorama naturalness produced by different methods.

A. VPG Dataset

36 sets of images were collected to form the VPG dataset. As shown in Figure 4, it includes 12 sets of synthetic images and 24 sets of real images. All synthetic images were generated through 3Ds Max rendering[2] hence the associated camera parameters are known. All real images were captured by ourselves with a mobile phone. The VPG dataset contains both indoor scenes and outdoor street-view scenes. Specifically, 12 synthetic sets are composed of 6 indoor cases and 6 street-view cases. Similarly, 24 real sets consist of 12 indoor scenes and 12 street-view scenes. More details can be found in Figure 4. It is noteworthy that all images were carefully collected to ensure that they satisfy the Manhattan assumption. The number of images involving in stitching in each set ranges from 5 to 72.

B. Quantitative Metrics

Before specific experiments, two quantitative metrics first are introduced for the assessment of panorama naturalness. Note that comprehensively evaluating the panorama quality is still an open research problem [33], [34], and it is not the main concern of this paper either. Therefore, we simply start from the observation that a panorama produced by stitching methods

[2]https://www.autodesk.com/

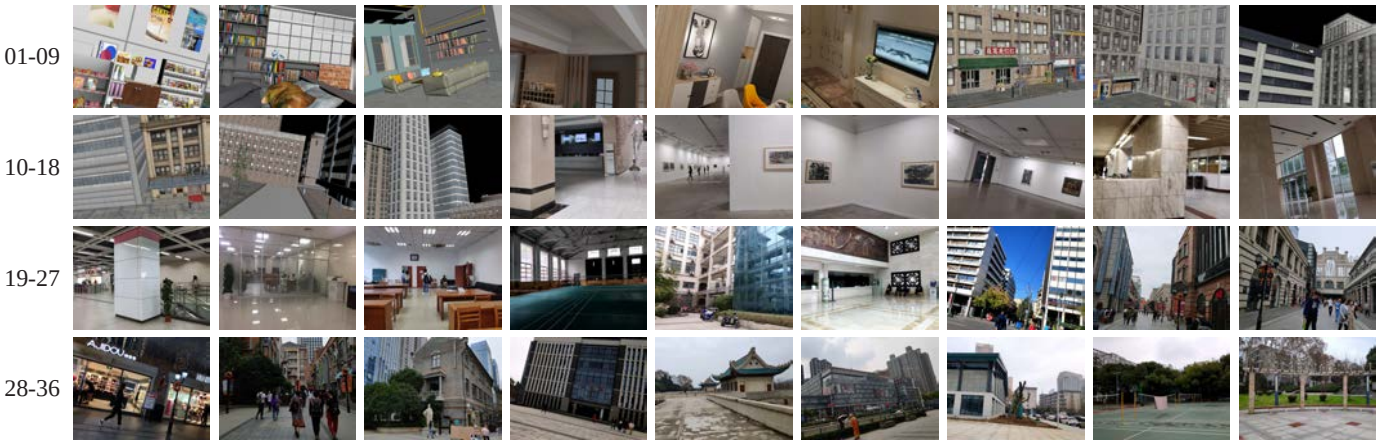


Figure 4. An overview of the VPG dataset. It consists of 12 synthetic sets (01-12) and 24 real sets (13-36). 01-06 and 13-24 are indoor cases. 07-12 and 25-36 are outdoor street-view cases.

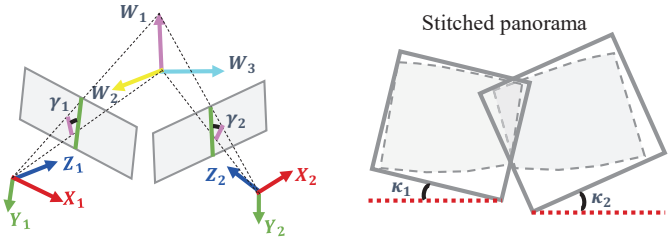


Figure 5. An illustration of the proposed GDIC metric. With the assist of external parameters, we approximate the direction of image content by a 2D angle γ with respect to the absolute vertical direction. Accordingly, we approximate the direction of stitched image content by κ, the orientation of the image bounding rectangle. The GDIC then measures the average difference between these two types of directions.

usually suffers two kinds of unnaturalness: local projective distortion, and global unnatural rotation. Accordingly, we put forward two indexes for quantitative evaluation: Local Distortion (LD) and Global Direction InConsistency (GDIC).

1) ***LD-index****:* LD is used to evaluate the local projective distortion. Let Q be a quad with four vertexes, the local homography H can be computed from its original and deformed coordinates. We first use a similar way as in [12] to measure the local area change at image position (x, y):

$$\mathcal{M}(H, (x, y)) = \det J_H(x, y), \tag{12}$$

where J_H is the Jacobian of H. We then calculate the mean μ_Q and the standard deviation σ_Q of $\mathcal{M}$ from all pixels within Q, and take the coefficient of variation (c.v.) σ_Q/μ_Q as the measurement of projective distortion for Q. Subsequently, for I_i, we compute the c.v. for all quads that located on the non-overlapping region, and take their average value as the measurement $\mathcal{D}_i$ for I_i. Finally, we define the LD index as:

$$\text{LD} = \max(\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_N). \tag{13}$$

2) ***GDIC-index****:* GDIC aims to measure the global unnatural rotation. Assuming that the external parameters are known for I_i, as shown in Figure 5, we compute a rotation γ_i for I_i on the image plane. It is a 2D angle between the camera y-axis and the absolute vertical direction. γ_i approximates the direction of image content in the world coordinate system. Accordingly, on the output panorama, we estimate the bounding rectangle with the minimal area for deformed vertexes of I_i, and take the rectangle orientation κ_i as the direction of I_i after stitching. We think that the relative direction of image content should be preserved as much as possible if images are stitched with a natural look, and the GDIC then is defined as:

$$\text{GDIC} = \frac{\sum_{i=1, i\neq r}^{N} |(\kappa_i - \kappa_r) - (\gamma_i - \gamma_r)|}{N - 1}, \tag{14}$$

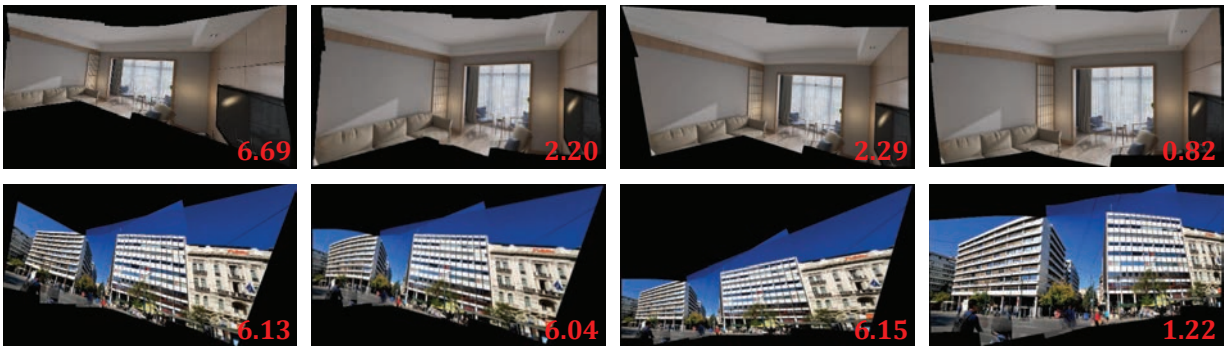
where r denotes the selected reference for relative direction computation. I_r is fixed when computing GDIC for different methods. Since the proposed GDIC requires the external parameters, it is available only for synthetic image data in our experiments.

C. Comparison with APAP [4] and SPHP [12]

We first compared VPG with two early state-of-the-art algorithms: APAP [4] and SPHP [12]. They were tested using the source code provided by the authors. Since APAP and SPHP have limitations on the field of view, we had to reduce the number of involved images from tens to 3-9 during our experiments. Table I reports the quantitative results on synthetic sets 04-09. VPG outperforms other two competing methods in both LD and GDIC metrics. Figure 6 further provides qualitative comparisons on one synthetic set (04) and one real set (29). As we can see, panoramas produced by APAP and SPHP exhibit severe projective distortions and unnatural rotations. Adjusting the zero-rotation switch for SPHP has very limited effects in improving naturalness. In contrast, the proposed VPG produces panoramas with apparently higher visual quality. The qualitative comparison also is consistent with the quantitative evaluation.

D. Comparison with AANAP [15] and GSP [13]

We compared the proposed VPG with two recent state-of-the-art natural stitching methods: AANAP [15] and GSP [13].



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

Figure 6. Qualitative comparisons with APAP [4] and SPHP [12]. (a) results from APAP. (b) results from SPHP with the zero-rotation switch on. (c) results from SPHP* with the zero-rotation switch off. (d) results from our VPG. The corresponding LD values are given in red text.

Table I. Quantitative comparisons with APAP [4] and SPHP [12] on synthetic image set 04-09. SPHP* denotes the zero-rotation switch is turned off during stitching.

Metrics	Methods	04	05	06	07	08	09
LD ($\times 10^{-2}$) ↓	APAP	6.69	8.36	6.70	4.01	6.21	5.55
	SPHP	2.20	3.93	7.32	2.66	3.07	2.32
	SPHP*	2.29	2.08	3.17	2.81	2.29	2.16
	VPG	**0.82**	**1.35**	**1.11**	**1.08**	**0.19**	**1.15**
GDIC (deg) ↓	APAP	18.94	22.76	11.44	11.48	4.76	7.02
	SPHP	11.21	20.16	5.32	4.92	8.06	13.27
	SPHP*	8.75	3.94	6.58	8.52	6.30	12.70
	VPG	**1.21**	**0.81**	**0.95**	**0.95**	**1.16**	**2.13**

GSP was tested using the source code provided by authors and both the 2D solution and the 3D solution were tested. AANAP was tested using our own re-implementation. Table II offers quantitative comparisons on synthetic image sets $01-12$. The proposed VPG has a comparable performance with AANAP and GSP in LD metric since they all share a similar mesh deformation framework to reduce projective distortion. Moreover, in most cases, VPG steadily produces the smallest GDIC values, which means the best global natural look among four methods of comparison. Figure 7 and Figure 8 further present typical qualitative comparisons on synthetic and real images respectively, from which the superiority of proposed VPG can be observed intuitively. AANAP empirically estimates the image rotation with the smallest angle, GSP-2D assumes the zero rotation for images, and GSP-3D determines the rotation with pairwise 3D rotation relationships. Their results exhibit obvious unnaturalness since the lack of effective global constraints. In contrast, the proposed VPG manages to improve the panorama naturalness significantly.

E. User Study

Since naturalness is a subjective feeling, we further conducted a user study to investigate whether the proposed VPG is preferred by users. In practice, we invited 20 participants, including 10 researches/students with computer vision/graphics backgrounds and remaining 10 volunteers outside this community. We randomly selected 20 groups of stitching results in different scenes (*e.g.*, indoor and street-view) for the user study. There were 4 unannotated panoramas in each group that were produced by 4 methods: AANAP, GSP-2D, GSP-3D, and our VPG. Panoramas were shown on a screen in sequence, and the user was allowed to switch images back and forth for a convenient comparison. Then, each participant ranked four results in each group, and assigned each panorama with the corresponding score (from rank 1 to rank 4, scores varies from 5 to 2). Figure 9 shows the user study results. The VPG is substantially preferred. In addition, it indicates to some extent that the adopted two metrics, LD and GDIC, are consistent with the user's subjective evaluation.

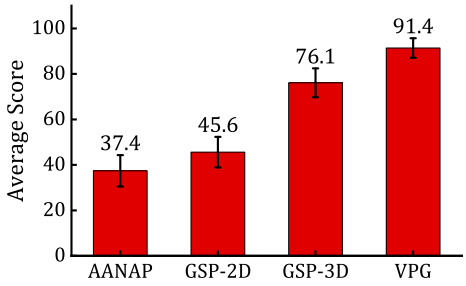


Figure 9. User study result. The average scores of 4 methods from 20 participants.

F. Validation of the Robust Estimation Scheme

VPG determines the image similarity prior using a two-step robust estimation scheme. In this section, we hope to quantitatively valid the effectiveness of this step. Using the VPG synthetic image set 01-12 and given the VP extraction results from different images, we manually added random noises to the VP coordinates before adopting them to extract VPs guidance. Then, we compared the associated GDIC values produced by VPG without the robust estimation scheme and by VPG with the robust estimation scheme. Table III reports the comparison results when gradually increasing the noise ratio from 0% to 20%. As we can see, metric values from

Table II. Quantitative comparisons with AANAP [15] and GSP [13] on 12 sets of synthetic images. 01-06 are indoor scenes, and 07-12 are outdoor street-view scenes.

Metrics	Methods	01	02	03	04	05	06	07	08	09	10	11	12
LD ($\times 10^{-2}$) ↓	AANAP	**1.21**	1.56	**1.20**	2.94	**1.47**	**1.10**	0.94	5.17	5.95	1.50	1.14	0.89
	GSP-2D	1.48	1.73	1.82	2.19	1.85	1.28	1.00	3.00	2.68	2.04	**1.01**	0.74
	GSP-3D	1.41	1.60	1.59	**1.81**	1.84	1.43	1.05	2.06	2.69	1.35	1.11	**0.67**
	VPG	1.44	**1.45**	1.57	2.06	1.68	1.53	**0.93**	**2.69**	**2.50**	**1.27**	1.19	0.73
GDIC (deg) ↓	AANAP	2.26	1.06	2.33	8.72	7.24	8.60	2.75	9.49	25.06	4.37	2.59	3.25
	GSP-2D	6.06	2.07	1.66	2.77	6.20	1.63	2.50	2.70	4.26	1.23	2.22	1.76
	GSP-3D	4.54	2.78	3.41	1.05	1.83	1.05	3.73	2.05	0.86	**1.11**	1.70	1.18
	VPG	**0.73**	**0.49**	**0.77**	**0.49**	**0.55**	**0.74**	**0.72**	**0.50**	**0.52**	1.41	**0.81**	**0.63**

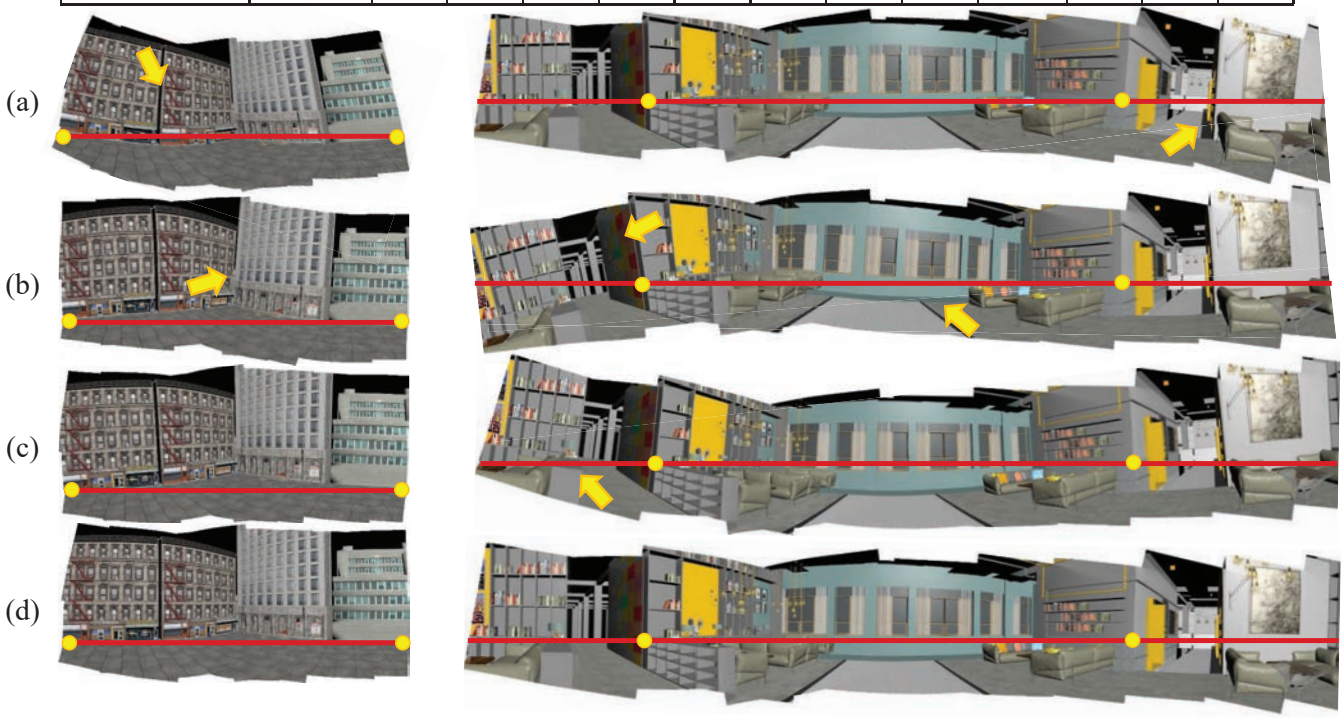


Figure 7. Qualitative comparisons with AANAP [15] and GSP [13] on synthetic set 03 and 08. (a) results from AANAP. (b) results from GSP-2D. (c) results from GSP-3D. (d) results from the proposed VPG. The yellow points indicating the anchor directions that have been aligned to the red horizontal lines for better visual comparison. The yellow arrows hightlight the unnatural artifacts. The same marks are adopted in the following figures.

VPG (w/o) increases significantly as the noises increase while values from VPG (w/) keep relatively stable in most cases. It demonstrates that the proposed robust estimation scheme effectively ensures the robustness of VPG and maintains the consistently high naturalness of the output panoramas.

V. More Analysis

We further analyzed the performance of VPG in four main aspects: (1) the adaptability for general scenes where the Manhattan assumption may not hold; (2) the stability for reference selection; (3) the scalability for higher alignment accuracy; (4) the time efficiency.

A. Adaptability for General Scenes

We explored the possibility that VPG adaptively falls back to a regular stitching scheme without using the VP clues when the scene disobeys the required Manhattan assumption. Since we have projected the VPs from different images on a unified sphere surface in previous Sections, we think these roughly aligned VPs can reflect the regularity of the scene.

Given a set of aligned VPs, we define its associated VP divergence as follows:

$$\varepsilon = \frac{1}{\rho} \sum_{\hat{\mathbf{v}}_i \in inlier} \|\hat{\mathbf{v}}_i - \hat{\mathbf{d}}_i\|^2, \tag{15}$$

where $\hat{\mathbf{d}}_i$ is the dominant direction of $\hat{\mathbf{v}}_i$ that has been obtained previously, ρ denotes the inlier ratio that is estimated in the outlier rejection step. We consider a scene with a small ε as a Manhattan scene and perform the stitching process using the complete VPG scheme. Otherwise, we alternatively remove

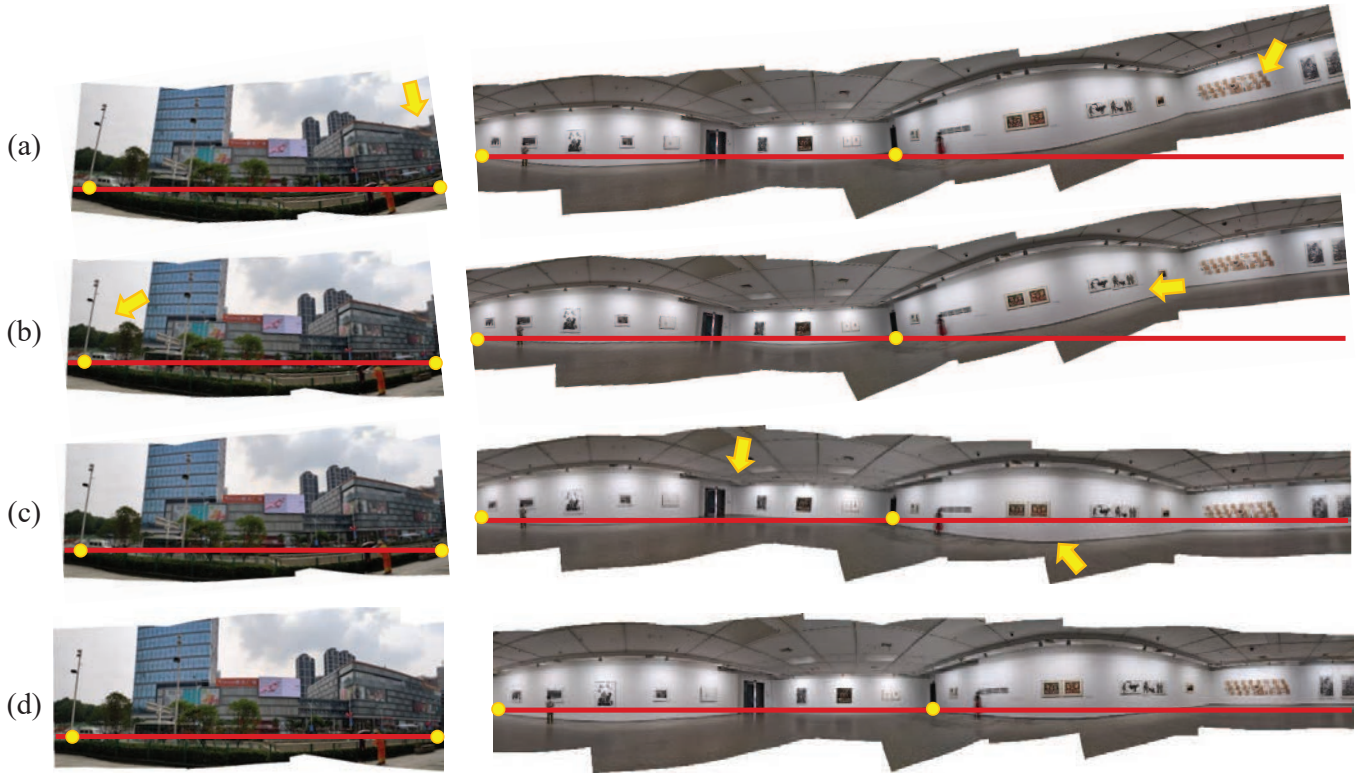


Figure 8. Qualitative comparisons with AANAP [15] and GSP [13] on real image set 16 and 33. (a) results from AANAP. (b) results from GSP-2D. (c) results from GSP-3D. (d) results from our VPG.

Table III. Quantitative validation of the proposed robust estimation scheme for the image similarity prior. The GDIC values of 12 synthetic image sets are reported.

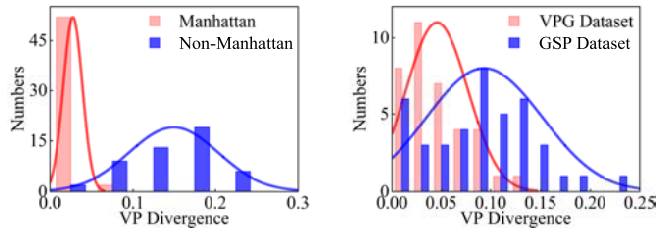
Noises	Methods	01	02	03	04	05	06	07	08	09	10	11	12
0% noises	VPG (w/o robust scheme)	**0.73**	**0.49**	**0.77**	0.82	1.01	**0.74**	**0.72**	0.65	0.96	2.41	**0.81**	**0.63**
	VPG (w/ robust scheme)	**0.73**	**0.49**	**0.77**	**0.49**	**0.55**	**0.74**	**0.72**	**0.50**	**0.52**	**1.41**	**0.81**	**0.63**
10% noises	VPG (w/o robust scheme)	2.97	0.44	1.28	1.74	0.90	0.80	0.65	1.65	0.81	2.30	1.01	1.79
	VPG (w/ robust scheme)	**0.53**	**0.44**	**0.64**	**0.84**	**0.59**	**0.65**	**0.65**	**0.54**	**0.38**	**1.37**	**0.94**	**1.53**
20% noises	VPG (w/o robust scheme)	4.74	2.57	4.47	2.29	2.47	2.67	1.67	2.21	3.19	8.55	1.65	1.13
	VPG (w/ robust scheme)	**0.56**	**0.42**	**0.63**	**0.47**	**0.57**	**0.63**	**0.74**	**0.71**	**0.56**	**3.75**	**0.94**	**0.85**

the VP guidance[3] in Eq. 7 to make the proposed stitching algorithm fall back to the regular scheme as in [13].

In order to determine the threshold for ε, we collected another 130 sets of image. Half of them were captured in Manhattan scenes, and another half of them were captured in natural scenes (non-Manhattan). Their VP divergence values were computed according to Eq. 15 and are presented in Figure 10. As we can see, ε from a Manhattan scene usually is small and tends to be limited in a narrow range. On the contrary, ε of a natural scene usually has a relatively large value. $\varepsilon_0 = 0.10$ seems to be a valid sentry threshold to distinguish the Manhattan scenes from non-Manhattan scenes.

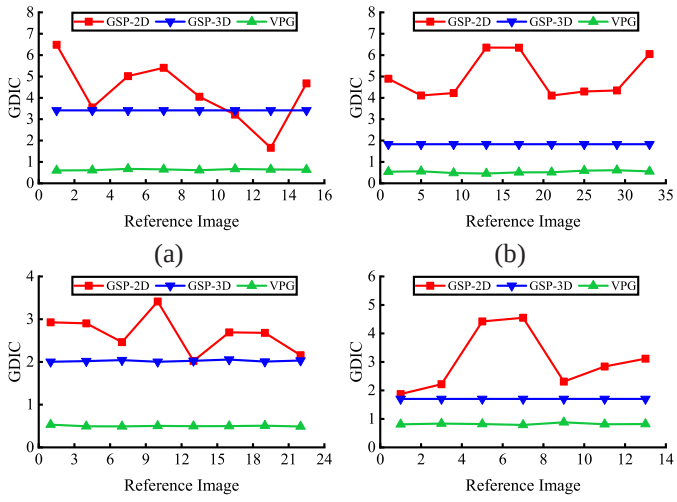
To verify the adaptability of VPG when resorting to ε, we hope to simulate a practical VPG application scenario in which the regularity prior about the scene (Manhattan or non-Manhattan) is unknown and so we tested VPG on the GSP dataset [13]. It consists of 42 image sets and contains nearly all popular images for stitching algorithm evaluation. Since the Manhattan assumption is not necessarily satisfied in GSP dataset, as presented in Figure 10, the associated VP divergences are distributed on both sides of ε_0. Figure 11 and Figure 12 present typical results on GSP dataset. On the one hand, when the ε of a scene is small, VPG manages to improve the panorama naturalness by utilizing the reliable VPs guidance. On the other hand, if the ε is large which indicates the scene is prone to be a non-Manhattan scene, VPG still can produce natural looking panoramas by weakening the extracted VPs guidance. Note that all results are produced by VPG automatically without any manual intervention. It demonstrates that the proposed VPG can be well applied in general (Manhattan or non-Manhattan) scenes. Figure 16 presents more stitching results on GSP dataset.

[3]Thanks to the straighten strategy in [1], Eq. 7 is solvable even we weaken or remove the VP-relevant data term.



(a) (b)

Figure 10. (a) VP divergence distributions of Manhattan scenes and Non-Manhattan scenes. (b) The associated VP divergence distributions of the adopted VPG dataset and GSP dataset.



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

Figure 13. Quantitative evaluations of the influence of reference selection on different stitching methods. (a)-(d) present the results on VPG-03, 05, 08 and 11 respectively.

B. Stability for Reference Selection

Reference selection is an important but challenging issue for image stitching. Many algorithms are sensitive to this step and as a result, they may yield significantly different panoramas when different images are selected as the reference. Selecting the optimal reference is not easy even many methods have been proposed in the past decade [35], [36], [37], [38]. Taking 4 sets of synthetic images as an example, Figure 13 reports the GDIC quantitative results of VPG when different images are selected as the reference during the stitching process. As we can see, results from GSP-2D are severely affected by reference selection since different reference will lead to different images with the zero-rotation, which makes the panorama appearance change significantly. In contrast, the proposed VPG produces much more stable GDIC values no matter which image is chosen as the reference. Note that results from GSP-3D have a similar stability with VPG, but the corresponding GDIC values are much larger which indicate a worse panorama naturalness.

C. Scalability for Higher Alignment Accuracy

Except for naturalness, alignment accuracy is another essential issue that is widely considered when designing a stitching algorithm. Some methods [3] achieve high panorama naturalness at the expense of a decreased alignment accuracy. In previous experiments and analysis, for a fair comparison, VPG follows a similar scheme as in [13] to extract the alignment constraints. Although it inherits the powerful alignment capability provided by APAP [4], the alignment accuracy can be further increased by many recent advanced stitching frameworks like DFW [5] and GCPW [39]. In this section, we show that the proposed VPG is scalable for achieving a higher alignment accuracy. In other words, the naturalness improvement achieved by VPG is compatible with high alignment accuracy. Figure 14 and Figure 15 present two groups of panoramas produced by different methods and report the associated GDIC values and the MSE of alignment accuracy [7], [40], [20], [6]. By combining VPG with DFW and GCPW, the output panoramas not only have more natural looks than GSP, but also have higher alignment accuracy than both GSP and the original VPG.

Table V. Average Runtime of different stitching methods.

Dataset	AANAP	GSP	VPG
VPG Dataset	109.44s	45.82s	47.82s
GSP Dataset	53.63s	23.55s	29.04s

D. Time Efficiency

To evaluate the stitching efficiency of VPG, Table V reports the average runtime of AANAP, GSP, and VPG on 36-set VPG dataset and 42-set GSP dataset respectively. Without any acceleration technology, the proposed VPG method slightly increases the runtime by about 10% when compared with GSP but it significantly improves the panorama naturalness. Besides, VPG is approximately 2 times faster than AANAP with a much better panorama naturalness.

VI. Conclusions and Future Work

In this paper, we propose a vanishing-point-guided stitching method called VPG. VPG successfully exploits the predominance of VPs to achieve a robust estimation for image similarity prior, which finally leads to a more natural looking panorama. Quantitative and qualitative comparisons on synthetic and real images combined with a user study demonstrate VPG's superiority over other state-of-the-art methods. More analyses upon VPG show that although VPG is designed for Manhattan scenes, it possesses good adaptability for general scenes through a degradation mechanism. Meanwhile, due to the introduction of global VPs, VPG outputs stable panoramas that are free from different reference selections. Moreover, VPG is scalable and compatible with other advanced stitching frameworks to achieve a coordination between panorama naturalness and alignment accuracy. We also observed some limitations for VPG. Firstly, VPG is prone to fall back to the non-Manhattan scene when the number of involved images is small (*e.g.*, 2 or 3). Because in such cases, the VP consistency usually is not remarkable for similarity prior estimation. Secondly, VPG may fail when facing dramatic depth variation or large parallax, which could influence the VPs alignment results and cause unnatural artifacts. Note that extremely large parallax is also challenging for most other existing stitching approaches and is the issue that needs to be overcome in the future work.

Figure 11. Qualitative comparisons with AANAP [15] and GSP [13] on 2 GSP sets. (a) results from AANAP. (b) results from GSP-2D. (c) results from GSP-3D. (d) results from VPG. Top row is a indoor scene with 35 input images and the associate $\varepsilon = 0.044 \leq \varepsilon_0$. Bottom row is a outdoor scene with 5 input images and $\varepsilon = 0.048 \leq \varepsilon_0$.

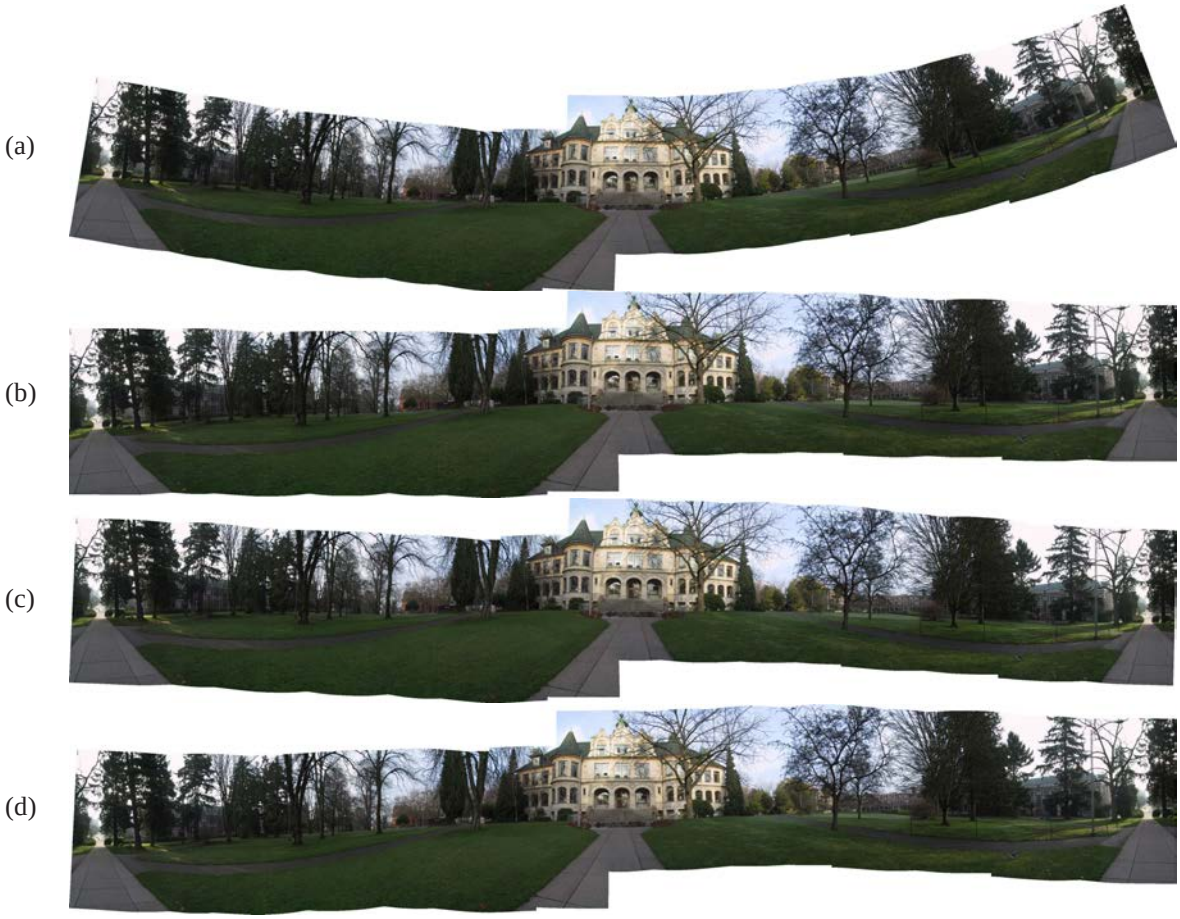
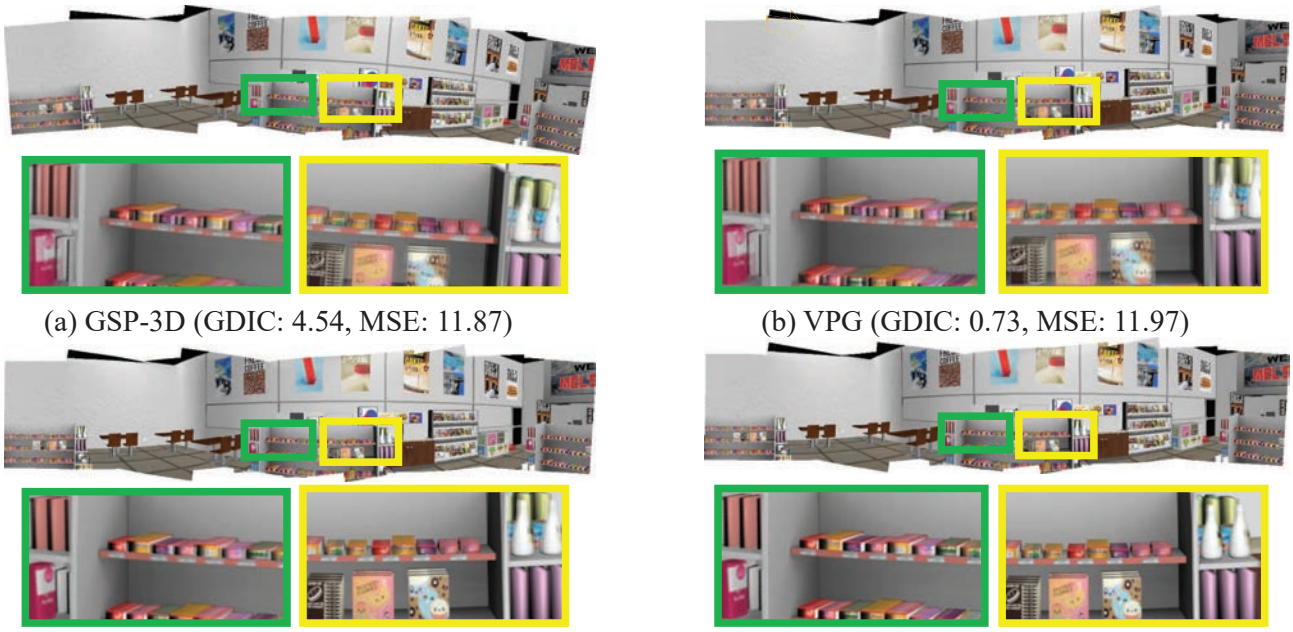


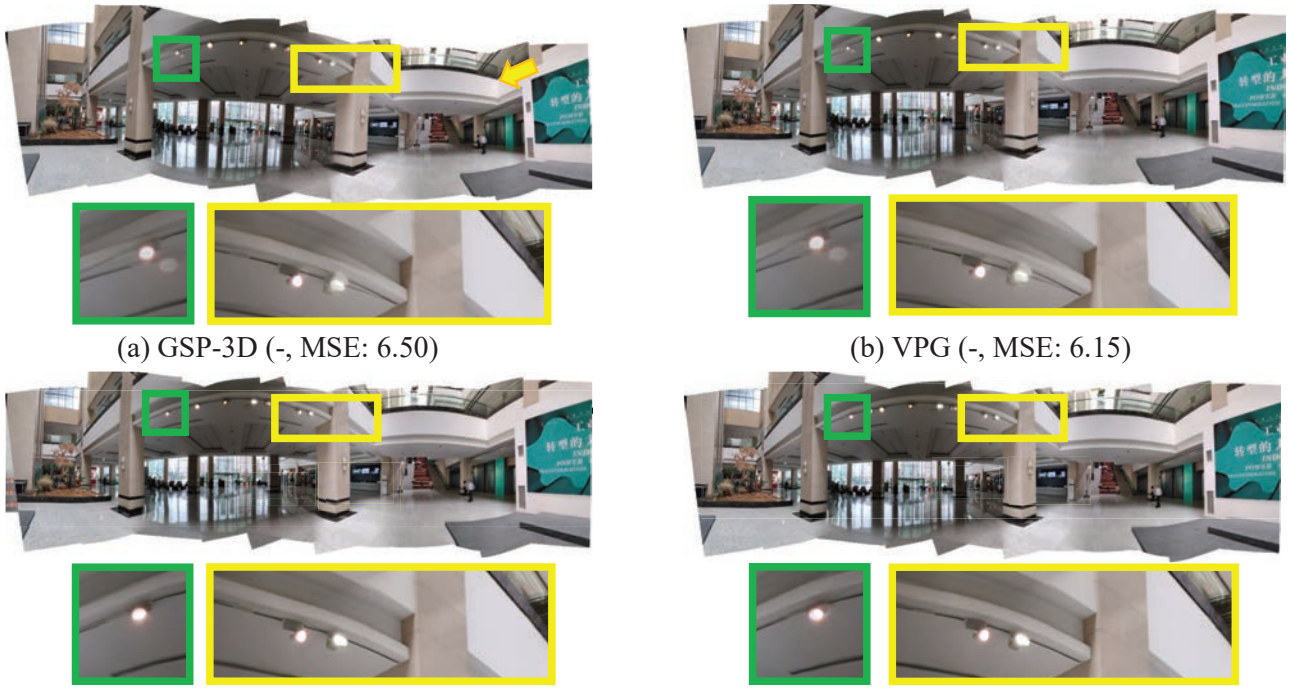
Figure 12. Qualitative comparisons with AANAP [15] and GSP [13] on one GSP set. (a) results from AANAP. (b) results from GSP-2D. (c) results from GSP-3D. (d) results from our VPG. It is a scene with 15 input images and $\varepsilon = 0.161 > \varepsilon_0$.



(a) GSP-3D (GDIC: 4.54, MSE: 11.87) (b) VPG (GDIC: 0.73, MSE: 11.97)

(c) VPG+DFW (GDIC: 1.00, MSE: 10.89) (d) VPG+GCPW (GDIC: 1.09, MSE: 10.19)

Figure 14. Qualitative comparisons between GSP-3D [13], the original VPG, VPG+DFW and VPG+GCPW on VPG synthetic set 01. By combining VPG with DFW and GCPW, the alignment accuracy gets improved significantly without any obvious loss of panorama naturalness.



(a) GSP-3D (-, MSE: 6.50) (b) VPG (-, MSE: 6.15)

(c) VPG+DFW (-, MSE: 5.93) (d) VPG+GCPW (-, MSE: 5.49)

Figure 15. Qualitative comparisons between GSP-3D [13], the original VPG, VPG+DFW and VPG+GCPW on VPG real set 13. By combining VPG with DFW and GCPW, the alignment accuracy gets improved significantly without any obvious loss of panorama naturalness. Note that the GDIC values are unavilable for real image data.

Table IV. Basic information about the 12 panoramas (a)-(l) presented in Figure 16. Note that the ε values from (a)-(h) are less than ε_0 while values from (i)-(l) are larger than ε_0.

No.	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
Image Numbers	2	4	5	6	11	3	5	21	10	5	7	15
ε	0.0204	0.0019	0.0068	0.0005	0.0740	0.0002	0.0114	0.0781	0.1917	0.1291	0.1464	0.2022

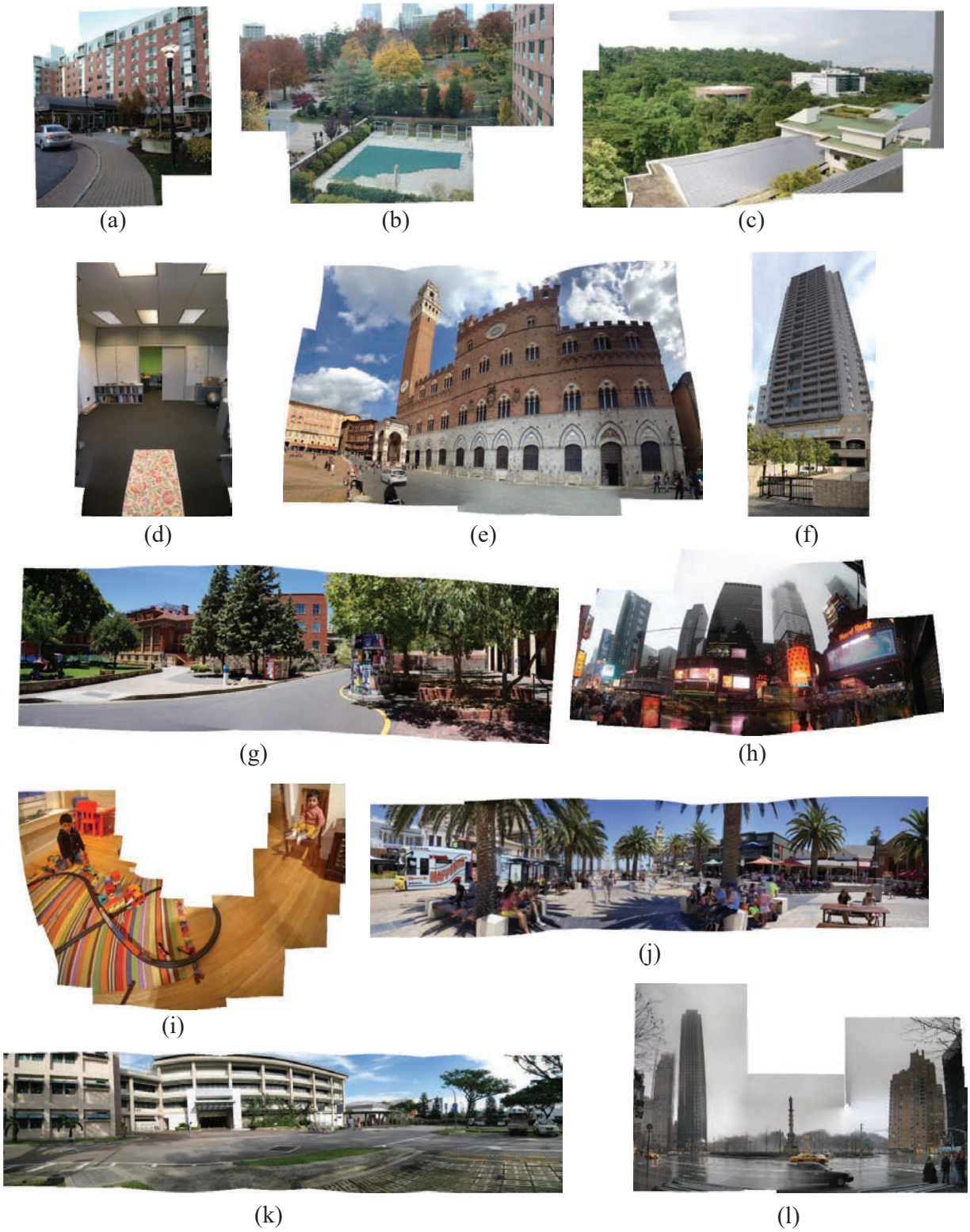


Figure 16. More stitching results on GSP dataset.

REFERENCES

[1] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, "Smoothly varying affine stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[4] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving dlt," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[5] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-feature warping-based motion model estimation," in *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[6] K. Lin, N. Jiang, S. Liu, L.-F. Cheong, M. Do, and J. Lu, "Direct photometric alignment by mesh deformation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[7] K. Chen, J. Tu, B. Xiang, L. Li, and J. Yao, "Multiple combined constraints for image stitching," in *IEEE International Conference on Image Processing (ICIP)*, 2018.

[8] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3d video stabilization," in *ACM Transactions on Graphics*, vol. 28, no. 3, 2009, p. 44.

[9] W. Yan and C. Hou, "Reducing perspective distortion for stereoscopic image stitching," in *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2016.

[10] H. Wang, Y. Zhou, X. Wang, and L. Fang, "A natural shape-preserving stereoscopic image stitching," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[11] F. Zhang and F. Liu, "Casual stereoscopic panorama stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[12] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[13] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *European Conference on Computer Vision (ECCV)*, 2016.

[14] N. Li, Y. Xu, and C. Wang, "Quasi-homography warps in image stitching," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1365–1375, 2018.

[15] C.-C. Lin, S. U. Pankanti, K. Natesan Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[17] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[18] H. Guo, S. Liu, T. He, S. Zhu, B. Zeng, and M. Gabbouj, "Joint video stitching and stabilization from moving cameras," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5491–5503, 2016.

[19] T.-Z. Xiang, G.-S. Xia, X. Bai, and L. Zhang, "Image stitching by line-guided local warping with global similarity constraint," *Pattern Recognition*, vol. 83, pp. 481–497, 2018.

[20] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, "Seagull: Seam-guided local alignment for parallax-tolerant image stitching," in *European Conference on Computer Vision (ECCV)*, 2016.

[21] Q. Chai and S. Liu, "Shape-optimizing hybrid warping for image stitching," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2016.

[22] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, "Multi-viewpoint panorama construction with wide-baseline images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3099–3111, 2016.

[23] X. Lu, J. Yaoy, H. Li, and Y. Liu, "2-line exhaustive searching for real-time vanishing point estimation in manhattan world," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017.

[24] D. C. Lee, M. Hebert, and T. Kanade, "Geometric reasoning for single image structure recovery," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[25] J.-K. Lee and K.-J. Yoon, "Real-time joint estimation of camera orientation and vanishing points," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[26] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Transactions on graphics*, vol. 33, no. 4, p. 129, 2014.

[27] C. Zou, A. Colburn, Q. Shan, and D. Hoiem, "Layoutnet: Reconstructing the 3d room layout from a single rgb image," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[28] S. Lee, J. Kim, J. Shin Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. Seok Hong, S.-H. Han, and I. So Kweon, "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition," in *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[29] Y. H. Lee, C. Nam, K. Y. Lee, Y. S. Li, S. Y. Yeon, and N. L. Doh, "Vpass: Algorithmic compass using vanishing points in indoor environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009.

[30] F. Camposeco and M. Pollefeys, "Using vanishing points to improve visual-inertial odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[31] H. Li, J. Yao, J.-C. Bazin, X. Lu, Y. Xing, and K. Liu, "A monocular slam system leveraging structural regularity in manhattan world," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[32] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2010.

[33] G. Cheung, L. Yang, Z. Tan, and Z. Huang, "A content-aware metric for stitched panoramic image quality assessment," in *IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2017.

[34] S. Ling, G. Cheung, and P. Le Callet, "No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2018.

[35] T. E. Choe, I. Cohen, M. Lee, and G. Medioni, "Optimal global mosaic generation from retinal images," in *International Conference on Pattern Recognition (ICPR)*, 2006.

[36] R. Szeliski *et al.*, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2007.

[37] R. Xie, J. Tu, J. Yao, M. Xia, and S. Li, "A robust projection plane selection strategy for uav image stitching," *International Journal of Remote Sensing*, pp. 1–21, 2018.

[38] M. Xia, J. Yao, R. Xie, L. Li, and W. Zhang, "Globally consistent alignment for planar mosaicking via topology analysis," *Pattern Recognition*, vol. 66, pp. 239–252, 2017.

[39] K. Chen, J. Tu, J. Yao, and J. Li, "Generalized content-preserving warp: Direct photometric alignment beyond color consistency," *IEEE Access*, vol. 6, pp. 69 835–69 849, 2018.

[40] K. Chen, J. Yao, B. Xiang, and J. Tu, "Video stitching with extended-meshflow," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3049–3054.