Urban Image Stitching using Planar Perspective Guidance

Joo Ho Lee
Seung-Hwan Baek
Min H. Kim
{jhlee,shwbaek,minhkim}@vclab.kaist.ac.kr

KAIST
School of Computing
291 Daehak-ro, Yuseong-gu
Daejeon, Korea

Abstract

Image stitching methods with spatially-varying homographies have been proposed to overcome partial misalignments caused by global perspective projection; however, local warp operators often fracture the coherence of linear structures, resulting in an inconsistent perspective. In this paper, we propose an image stitching method that warps a source image to a target image by local projective warps using *planar perspective guidance*. We first detect line structures that converge into three vanishing points, yielding line-cluster probability functions for each vanishing point. Then we estimate local homographies that account for planar perspective guidance from the joint probability of planar guidance, in addition to spatial coherence. This allows us to enhance linear perspective structures while warping multiple urban images with grid-like structures. Our results validate the effectiveness of our method over state-of-the-art projective warp methods in terms of planar perspective.

1 Introduction

Suppose two different images of a scene are captured with a certain overlap by moving a camera. These two images result in different perspectives with non-identical pairs of vanishing points. When combining them as a panorama image, it is necessary not only to find out corresponding points in the image pair for registration, but also to correct perspective projections of these two images respectively for projective warps. A straightforward solution is to estimate a global projective warp as a homography from one to the other image using detected features within the *overlapping* region in both images. However, the global projective warp often suffers from partial misalignments due to motion parallax, as the global projective warp assumes that the scene is a single plane at a distance [19]. However, the global projective warp often suffers from partial misalignments, since it only works for scenes in a single plane at a distance [19].

Since the depth information of a scene is unavailable, the estimation of a perfect projective warp for each plane is a severely ill-posed problem, particularly in *non-overlapping* regions. Two different approaches have been proposed to address this problem: First, image warp methods with a set of local homographies, such as dual homography [5] and as-projective-as-possible (APAP) [22], have been proposed to weight spatial coherence between arbitrary pixels and feature correspondences. However, they often result in typical wavy

(a) As-Projective-As-Possible (APAP) warp (b) Our warp with planar perspective guidance

Figure 1: Our image stitching method accounts for *planar perspective* while calculating local projective warp with the help of grid-like structural characteristics of urban scenes. It allows us to overcome wavy artifacts commonly observed by local homography-based warp approaches such as APAP [22].

artifacts in non-overlapping regions. Second, hybrid approaches combined with a projective warp and a similarity transformation have been proposed [3, 4, 13]. These methods apply a projective warp for the overlapping region and transform objects' similarity using shape-preserving transformations such as rotation, scaling, translation and reflection for non-overlapping regions. However, these combined approaches often result in inconsistent perspective and seam artifacts.

In this paper, we propose a novel image warp method that accounts for the *planar perspective probability* as warp guidance to enhance image structures while applying local homographies for projective warp. The proposed method is inspired by planar structure priors used for recent vision applications, such as image inpainting and super-resolution [7, 8], but our method applies the planar probability as guidance to relax the ill-posedness of the projective warp problem for image stitching. Our method first detects line structures that converge into the same vanishing point and obtain clusters of the line segments, yielding planar probabilities that have the same vanishing point pair. We then estimate local homographies that account for not only a planar perspective but also spatial coherence of line structures. We found that the proposed method maintains planar perspective structure of input images without wavy artifacts (see Figure 1), thereby becoming powerful in stitching urban scenes that consist of grid-like structures, targeting rendering applications that provide panoramic views from positions along many streets, such as Google Street View.

2 Related Work

Image stitching has been researched extensively in recent decades. For the sake of brevity, we refer readers to [19] for the foundations of this subject. Recent research in image stitching can be categorized into two groups: (a) adaptive warp and (b) shape-preserving warp. This section reviews these two state-of-the-art approaches.

Adaptive Warp Lin et al. [14] introduced the smoothly varying affine (SVA) transformation for image warp. While SVA can relax the difference between local and global transformations, the fundamental difference between affine and projective transformation remains as a limitation. Gao et al. [5] proposed the dual homography (DH) method that determines two different homographies for ground and distant planes from classified features using k-means clustering. However, this often requires manual user interaction to select seed points for clustering in complex scenes that should consist of ground and distant planes. Zaragoza et al. [22] introduced the as-projective-as-possible (APAP) warp method using moving direct linear transformation (DLT). They build locally varying homographies for each grid by taking into account the distance between a pixel to a feature point. The local homographies

near feature points tend to weight locality more, while the local homographies distant from feature points are likely to resemble the global homography. However, the Gaussian weight for estimating local homographies often fracture linear structures, resulting in wavy artifacts. Joo et al. [10] extended the APAP approach by adding line features in addition to original SIFT features. However, since local homographies are calculated using the spatial distance, they still yield traditional wavy artifacts. In order to account for warping/matching error of corresponding lines in the warp estimation, Li et al. [12] and Xiang et al. [20] consider line segments in an image. In contrast, we estimate joint *planar* probabilities of the same vanishing point to preserve a solider planar perspective over the prior work.

In this paper, we mainly attempt to overcome the *wavy artifacts* that commonly occur in locally adaptive warp methods [8, 10, 12, 14, 22] with the help of recognizing grid-like structures of urban scenes. Instead of using a simple *distance*-based weight between arbitrary pixels and feature points, we newly take into account planar perspective guidance in urban scenes, which are obtained from joint probability maps of vanishing points. Our projective warp adjusts local homographies without suffering from wavy artifacts.

Shape-Preserving Warp Owing to motion parallax, projective warp methods result in perspective distortion particularly in non-overlapping regions. Therefore, shape-preserving warp methods with similarity transformation have been proposed to mitigate this fundamental problem in a plausible manner [1, 2, 15]. Chang et al. [3] proposed the shape-preserving-half-projective (SPHP) image stitching method, which provides a smooth transition of interpolation from the DLT region to non-overlapping regions. They apply a similarity transformation in non-overlapping regions such as rotation, scaling, and translation instead of projective warp. Zhang and Liu [24] detect a highly distorted region of a homography warp and determine a proper similarity transformation to attenuate perspective distortion. Lin et al. [13] proposed the adaptive as-natural-as-possible (AANAP) image warp that extrapolates the local homography to the non-overlapping regions using homography linearization. They also compute a global similarity transformation to keep global structures such as a horizon. Chen et al. [4] optimize the warp of mesh grids while preserving conformality using a global rotation prior. These approaches with similarity transformation are oriented to find a *plausible attenuation of projective warp*; they often result in inconsistent perspective particularly in non-overlapping regions.

In contrast, we exploit planar perspective guidance for estimating local homographies in order to preserve linear perspective structures robustly in image stitching. Our objective is to achieve a more accurate linear perspective in local projective warp. To the best of our knowledge, this is the first work to exploit planar perspective guidance for image stitching.

3 Image Warp using Planar Perspective Guidance

We are motivated to estimate more accurate local warps that can provide coherent planar structures, resulting in a sound perspective after stitching. We attempt to achieve this goal by estimating planar perspective guidance from joint probabilities of detected vanishing points.

3.1 Global and Local Projective Warp

Homography This section briefly reviews the geometric perspective of a global homography and its fundamental limitation. A global homography is used for projective warp. Suppose we have a corner of a building captured by two different camera poses resulting in two

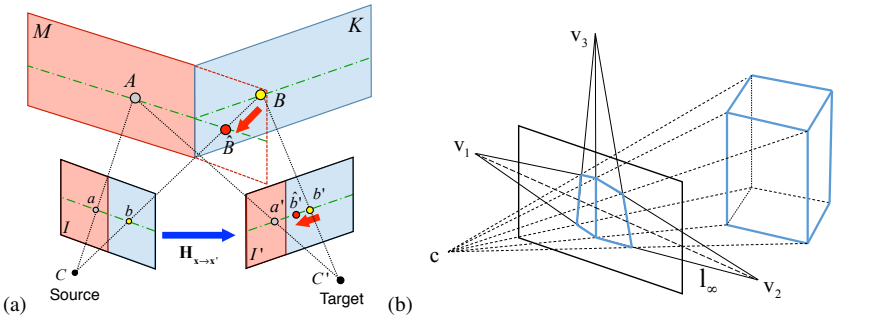

Figure 2: (a) A homography $\mathbf{H}_{\mathbf{x}\to\mathbf{x}'}$ warps a point $\mathbf{x}$ in a source image I to a point $\mathbf{x}'$ in a target image I'. The homography $\mathbf{H}_{\mathbf{x}\to\mathbf{x}'}$ is only valid when the point $\mathbf{x}$ sits on a object plane M such that $\mathbf{x}$ is a on the image I. When the homography $\mathbf{H}_{\mathbf{x}\to\mathbf{x}'}$ is applied to a point b that is a projected point of the point B on the different object plane K, the point b is warped to a point $\hat{b}'$, rather than b' in the target image I' due to the perspective difference between the two object planes. The red arrow indicates a homography error. Points C and C' indicate the center of projection. (b) Three vanishing points (v_1, v_2 and v_3) can be obtained from line clusters, where connecting the two horizontal vanishing points forms a vanishing line at infinity.

different images I and I', schematically depicted in Figure 2. Let a plane of the building M be defined as $(\mathbf{n}^\top, d)^\top \in \mathbb{R}^{4\times1}$ in the 3D world coordinates, where $\mathbf{n}$ is the surface normal and d is the distance from the camera. In homogeneous 2D coordinates, a point $\mathbf{x} \in \mathbb{R}^{3\times1}$ captured in an image I can be unprojected into a point $\mathbf{X} = (\mathbf{x}^\top, -\mathbf{n}^\top\mathbf{x}/d)^\top \in \mathbb{R}^{4\times1}$ in the object plane M. The object point $\mathbf{X}$ in the world coordinates can then be transformed to the other image I' via a rigid body transformation of rotation and translation. We denote the transformation of the point $\mathbf{X}$ to the point $\mathbf{x}'$ in image I' as $\mathbf{M}' = [\mathbf{R}|\mathbf{t}] \in \mathbb{R}^{3\times4}$, where $\mathbf{R} \in \mathbb{R}^{3\times3}$ is rotation and $\mathbf{t} \in \mathbb{R}^{3\times1}$ is translation: $\mathbf{x}' = \mathbf{M}'\mathbf{X}$. Finally a homography $\mathbf{H}$, a transformation from the point $\mathbf{x}$ in the image I to the point $\mathbf{x}'$ in the image I', can be calculated via 3D world coordinates $\mathbf{X}$ as follows [8]:

$$\begin{aligned} \mathbf{x}' &= \mathbf{R}\mathbf{x} - \mathbf{t}\mathbf{n}^\top\mathbf{x}/d = (\mathbf{R} - \mathbf{t}\mathbf{n}^\top/d)\mathbf{x}, \\ \mathbf{x}' &= \mathbf{H}\mathbf{x}, \quad \text{where} \quad \mathbf{H} = (\mathbf{R} - \mathbf{t}\mathbf{n}^\top/d). \end{aligned} \tag{1}$$

To this end, any point $\mathbf{x}$ on an object plane M in the image I can be transformed into a point $\mathbf{x}'$ in the image I' by the homography $\mathbf{H} \in \mathbb{R}^{3\times3}$: $\mathbf{x}' = \mathbf{H}\mathbf{x}$. While the homography is valid for transforming a point $\mathbf{x}$ from the specific object plane M to the other image I', it is invalid for another object plane K on the scene as shown in Figure 2(a).

Homography Estimation In order to obtain a global homography for given correspondence between images, we utilize the traditional direct linear transformation (DLT) [8]. A homography between two images I and I' defines a linear transformation between a pair of matching correspondences. A projective warp maps a point $\mathbf{x}$ in the source image I to the other point $\mathbf{x}'$ in the target image I' using $\mathbf{x}' = \mathbf{H}\mathbf{x}$:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \simeq \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^\top\mathbf{x} \\ \mathbf{h}_2^\top\mathbf{x} \\ \mathbf{h}_3^\top\mathbf{x} \end{bmatrix},$$

where $\simeq$ denotes equality up to scale, $\mathbf{h}_i^\top \in \mathbb{R}^{1\times3}$ is the i-th row of $\mathbf{H} \in \mathbb{R}^{3\times3}$. Since the directions of $\mathbf{x}'$ and $\mathbf{H}\mathbf{x}$ are the same but may differ in mangitude, we obtain a linear equation

by the cross product $\mathbf{O}_{3\times1} = \mathbf{x}' \times \mathbf{H}\mathbf{x}$. The equation is then rewritten in terms of the vector form of homography $\mathbf{h} \in \mathbb{R}^{9\times1}$,

$$\mathbf{O}_{3\times1} = \begin{bmatrix} \mathbf{O}_{1\times3} & -\mathbf{x}^\top & y'\mathbf{x}^\top \\ \mathbf{x}^\top & \mathbf{O}_{1\times3} & -x'\mathbf{x}^\top \\ -y'\mathbf{x}^\top & x'\mathbf{x}^\top & \mathbf{O}_{1\times3} \end{bmatrix} \mathbf{h}, \; \mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix}.$$

Note that there are two independent linear equations for a pair of matching points in this equation. For given N feature correspondences $\{\mathbf{x}_i, \mathbf{x}'_i\}_{i=1}^N$ for training, we estimate a homography $\hat{\mathbf{h}}$ to predict a projective warp for an arbitrary position $\mathbf{x}_*$ by solving the linear least-square system by singular value decomposition (SVD):

$$\hat{\mathbf{h}} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{h}\|^2 \quad \text{s.t. } \|\mathbf{h}\| = 1. \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{2N\times9}$ is a matrix that stacks all linearly independent equations of N matching pairs. Since we have the estimated $\mathbf{H}$ (reshaped $\hat{\mathbf{h}}$), we can warp a pixel at an arbitrary position $\mathbf{x}_*$ in the source image I to the position $\mathbf{x}'_*$ in the target image I'.

Combining Global and Local Warp As shown in Figure 2(a), a single global homography is insufficient to represent the geometric variety of every surface in the real world. Therefore, the use of locally adaptive homographies has been proposed by previous works [5, 10, 14, 22]. Zaragoza et al. [22] proposed using a spatial weight to build locally varying homographies instead of using a single global homography. When estimating each local homography for an arbitrary position $\mathbf{x}_*$ using Equation (2), they account for the *spatial distance* between the arbitrary position $\mathbf{x}_*$ and the feature correspondence $\mathbf{x}_i$ with weight $\mathbf{W}^s_*$.

$\mathbf{W}^s_* \in \mathbb{R}^{2N\times2N}$ is a diagonal form of spatial weights that can be used as $\mathbf{W}^s_*\mathbf{A}$ when solving the linear system in Equation (2): $\mathbf{W}^s_* = \operatorname{diag}\left(\begin{bmatrix} w^1_* & w^1_* & \cdots & w^N_* & w^N_* \end{bmatrix}\right)$, where w^i_* is the Gaussian-weighted distance between the arbitrary position $\mathbf{x}_*$ and the i-th feature correspondence $\mathbf{x}_i$: $w^i_* = \max\left(\exp\left(-\|\mathbf{x}_i - \mathbf{x}_*\|^2/\sigma^2\right), \gamma\right)$. In particular, the homography estimation could fail when the distance is large in the extrapolated region. To prevent the numerical problem, the Gaussian weight is clamped with a small value $\gamma \in [0,1]$. When the parameter γ increases, the global homography is weighted more than the local homography. The parameter σ varies depending on the image resolution. However, the spatially-weighted estimation of local homographies does not take any image structures into account while warping multiple planes in the scene, resulting in typical *wavy artifacts* in warped images (Figure 1).

3.2 Planar Perspective Guidance

In order to overcome the limitation of the spatially-weighted local homographies, we are motivated to account for grid-like image structures in typical urban scenes, while estimating projective warp as local homographies.

Vanishing Points When linear perspective is preserved in a landscape, the extensions of line structures in the scene converge to a vanishing point (VP), located outside the canvas, as shown in Figure 2(b). Since vanishing points can help us understand geometric structures of a scene [16, 18], VPs have been utilized in many vision applications [2, 7, 8]. There are several methods available for estimating VPs. Ikeuchi et al. [9] estimate VPs using line clustering, and Lezama et al. [11] approximate VPs through line fitting in the primal-and-dual space, and Zhai et al. [23] trained deep neural networks to detect VPs automatically.

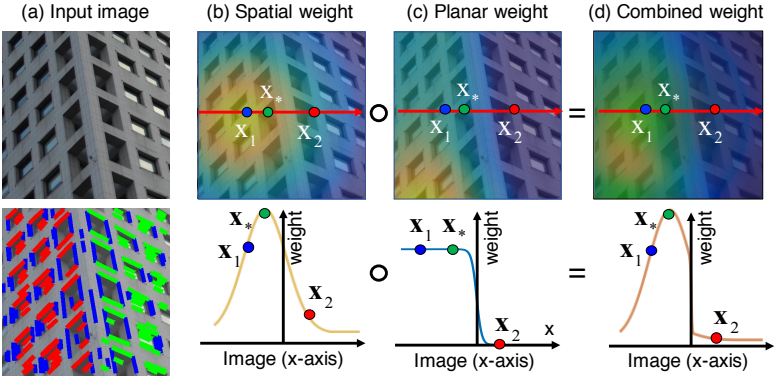

Figure 3: Column (a) shows an input image and detected line structures. Columns (b), (c) and (d) depict spatial, planar and combined weights respectively, where the combined weights are obtained by element-wise multiplication of the spatial and planar weights. The first-row images overlay color-indexed weight maps on the input image. Yellowish colors indicate higher weights than bluish colors. The second-row present three functions of spatial, planar and combined weights along an intersection line of the image. In addition to the traditional spatial weight (b), we introduce the use of planar weight (c) to combine both of the weights together to an plane-aware weight.

In this work, we estimate VPs to represent the parameters of an arbitrary plane M_* in the scene, by employing a standard voting approach [6]. We first extract lines by using a line segment detector (LSD) to obtain line clusters and then employ random sample consensus (RANSAC) for estimating VPs. We detect up to three VPs in an image, assuming that the input image includes grid-like structures which is typical in urban landscapes. See the first column (a) in Figure 3 for an example.

Coplanarity Probability Since we have three VPs, we can parameterize an arbitrary plane M_* in an image using three perspective basis planes, where each basis plane shares two distinct VPs, following [8]. Under a projective transformation, when line clusters on the plane M_* are extended to infinity, they are mapped to a horizontal fixed line $\mathbf{l}_\infty^{M_*}$ that connects the two distinct VPs [v_1 and v_2 in Figure 2(b)], which is also known as the vanishing line:

$$\mathbf{l}_\infty^{M_*} = [l_1^{M_*}\ l_2^{M_*}\ l_3^{M_*}]^\top. \tag{3}$$

Here $\mathbf{l}_\infty^{M_*}$ is also homogeneous and has two degrees of freedom. For instance, when we want to parameterize a perspective basis plane M_1, we need at least two detected line clusters, $l_1^{M_1}$ and $l_2^{M_1}$, as input. Note that instead of computing affine rectification from the perspective parameterization to a Euclidean plane [6], we directly use the probability of projective plane parameters, which estimates which plane the pixel $\mathbf{x}_*$ resides in.

Since we estimate three line clusters from each vanishing point, we iteratively diffuse the line clusters $\{l_i^{M_*}\}_{i=1}^3$ by applying the Gaussian filter and then obtain each line probability of the i-th line clusters for plane M^*, $P(l_i^{M_*}|\mathbf{x}_*)$, on the given arbitrary pixel $\mathbf{x}_*$. Finally, we can estimate the plane probability that includes the vanishing line of two VPs. Each plane probability that the arbitrary pixel $\mathbf{x}_*$ resides in is calculated from the joint probability of two

different line probabilities:

$$P(M_1|\mathbf{x}_*) = P(l_1^{M_*}|\mathbf{x}_*) \cdot P(l_2^{M_*}|\mathbf{x}_*),$$
$$P(M_2|\mathbf{x}_*) = P(l_2^{M_*}|\mathbf{x}_*) \cdot P(l_3^{M_*}|\mathbf{x}_*),$$
$$P(M_3|\mathbf{x}_*) = P(l_3^{M_*}|\mathbf{x}_*) \cdot P(l_1^{M_*}|\mathbf{x}_*).$$

Figure 3 shows an example of the probability distributions of planar perspective of an image.

Estimating Planar-Perspective Warp In addition to the spatial weight $\mathbf{W}_*^s$ using the distance between pixel $\mathbf{x}_*$ and $\mathbf{x}_i$ in the source image, we utilize the planar perspective distribution as a probability function for estimating local homography for pixel $\mathbf{x}_*$. $\mathbf{W}_*^p \in \mathbb{R}^{2N\times 2N}$ is also a diagonal form of planar perspective weights that can be used with the corresponding matrix $\mathbf{A}$ of matched feature pairs for estimating local homographies: $\mathbf{W}_*^p = \text{diag}\left(\begin{bmatrix} u_*^1 & u_*^1 & \cdots & u_*^N & u_*^N \end{bmatrix}\right)$, where u_*^i is a normalized joint planar probability distribution of three projective basis planes at arbitrary pixel $\mathbf{x}_*$ and the i-th feature correspondence $\mathbf{x}_i$.

$$u_*^i = \left\{ \left(\sum_{j=1}^{3} P(M_j|\mathbf{x}_*) \cdot P(M_j|\mathbf{x}_i) \right) + \lambda^2 \right\}^{\alpha}, \tag{4}$$

where λ is the minimum probability of the fronto-parallel plane for far distant objects without any line structures such as sky in order to avoid the numerical problem (λ is a fixed value 10^{-5}), $\alpha \in [0,1]$ is for handling the balance between the spatial and the planar weight.

Finally, a local homography can be estimated by taking into account the spatial weight $\mathbf{W}_*^s$ and the planar weight $\mathbf{W}_*^p$ between arbitrary pixel $\mathbf{x}_*$ and the i-th feature correspondence $\mathbf{x}_i$:

$$\hat{\mathbf{h}}_* = \underset{\mathbf{h}}{\text{argmin}} \|\mathbf{W}_*^p \mathbf{W}_*^s \mathbf{A}\mathbf{h}\|^2. \tag{5}$$

As shown in Figure 3, while estimating a local homography at arbitrary pixel position $\mathbf{x}_*$, the spatial weight $\mathbf{W}_*^s$ accounts for the feature correspondence at the shorter distance ($\mathbf{x}_1$) more than the distant features ($\mathbf{x}_2$), and the planar weight $\mathbf{W}_*^p$ accounts for the feature correspondence ($\mathbf{x}_1$) in the plane of the similar orientation more than the different one ($\mathbf{x}_2$).

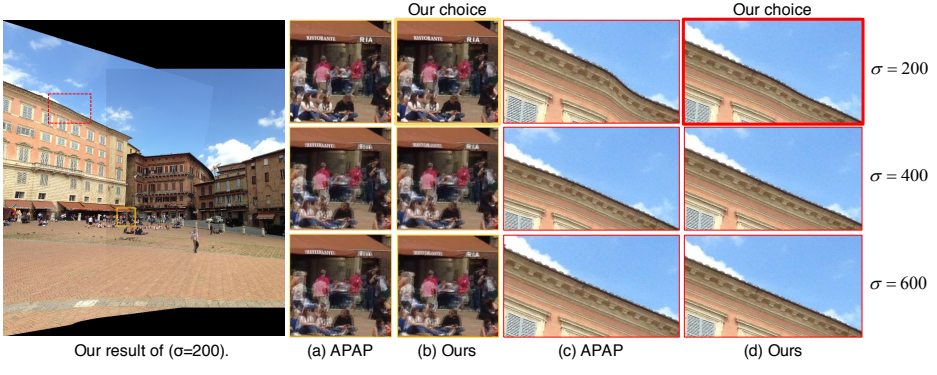

Figure 4: The impact of planar weights while varying a feature influence σ. We set the parameter σ to 200 for results in this paper by accounting for planar perspective and overall alignment. APAP results that are calculated without planar weights are in (a) and (c), which suffers from wavy artifacts and misalignments. Our method with planar weights reduces these artifacts by attenuating the parallax effect on non-coplanar features. (b) and (d) validate the robustness of our proposed method against changing parameters.

4 Results

We implement our planar-perspective warp method in MATLAB. We use the VLFeat library to detect and match SIFT features from input images. We demonstrate how our method preserves the planar perspective, compared to other projective warp methods.

Impacts of Parameters Figure 4 demonstrates the effectiveness of our planar weight while varying the parameter σ for the spatial weight $\mathbf{W}_*^s$, compared with results by the APAP method [22] that does not account for planar perspective. The steepness of the Gaussian function controlled by the parameter σ has significant impacts on wavy artifacts and overall alignment, as shown in (c). Our planar weight is effective for mitigating the drawback of local homographies methods by attenuating the parallax effect in perspective projection of different object planes in the scene.

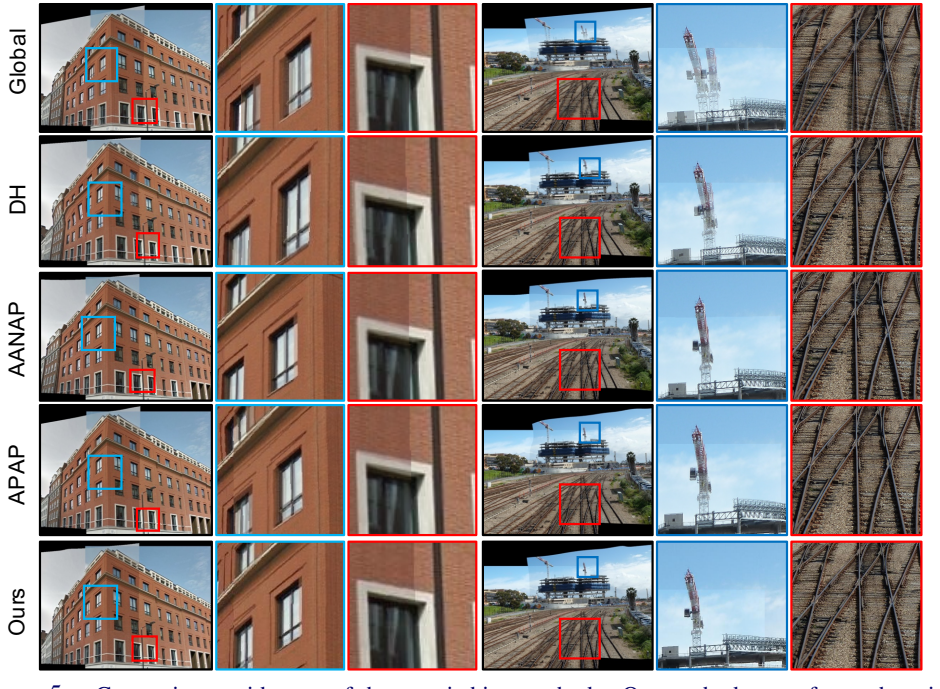

Figure 5: Comparisons with state-of-the-art stitching methods. Our method outperforms them in terms of warping planar building structures with less distortion, resulting in solid urban structures.

Comparison We compare our method with other projective warp-based image stitching methods, a global homography (the result of the DLT), a dual homography (DH) [5], APAP [22], and AANAP [13] in Figure 5. We use the common image datasets from previous works [3, 4, 5, 13, 17, 22]. For more results of our method compared with the state-of-the-art methods, refer to the supplemental material. For implementing the dual homography method, we carefully choose seed points to make a point cluster on every plane. In both of the homography planes, the DH method accurately estimates the homography of planes when there are two dominant homographies in the scene. The APAP method presents good alignments in the overlapping region, but often suffers from misalignments, resulting in wavy artifacts. The result of AANAP often shows the limitation of homography linearization on

a projective planar region. Most algorithms suffer from misalignments near the corner and severe distortion while propagating local homographies on the planar projective region. In summary, by detecting object planes in a scene using projective geometry, our method can estimate more accurate local homographies for planar structures than state-of-the-art projective warp methods.

5 Discussion and Future Work

Even though our planar-perspective warp method overcomes partial misalignments caused by perspective projection for urban scenes, there are several limitations.

We design our algorithm for urban scenes, assuming that an urban scene includes enough grid-like line structures to estimate vanishing points. Our method performs consistently for general urban scenes that include grid-like structures such as windows. Quality of our results depends on image structures of input images and also accuracy of our vanishing point detection method as shown in Figure 6. As shown in Figure 6(c), a planar probability is not uniform because of sparsity of horizontal lines in Figure 6(b). Even though our method can reduce wavy artifacts thanks to an isotropic distance weight, it still can suffer from wavy artifacts in certain scenes, due to a simple Gaussian weighted propagation of the weight. We could additionally introduce a smoothness term over a homography field or apply contents-aware warping approaches such as local similarity optimization [2, 4, 5, 15, 24] to balance both conformity and planar perspective. This remains as future work.

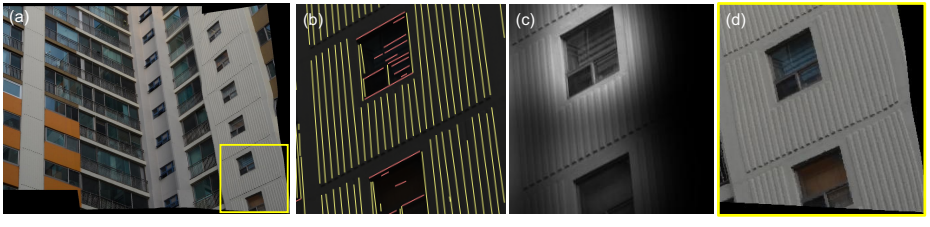

Figure 6: The impact on planar probability map. (a) shows our image stitching result. (b) presents detected line segments. (c) is a planar probability map computed from joint probabilities. (d) shows a closeup view of our result with wavy artifacts due to the non-uniformity of the planar probability map.

In addition, urban scenes with many natural objects like trees or dynamic free-formed shapes in the architecture cannot be handled properly. In such cases, our method performs similarly to the APAP method with wavy artifacts, since the vanishing point probabilities cannot be estimated accurately. An urban scene with too many planes could be also problematic due to depth ambiguity while estimating planar perspective guidance. It could be solved by adopting a depth-based approach of DLT, such as the bundle adjusted MDLT [21], for the overlapping region and 3D scene self-understanding to infer depth in the non-overlapping region.

6 Conclusion

We have exploited planar perspective guidance for image stitching. The proposed method performs 3D scene understanding using vanishing points and estimates an accurate projective warp in a non-overlapping region while preserving perspective alignments in overlap-

ping regions. To validate the proposed method, we demonstrate the effectiveness of planar perspective guidance over the state-of-the-art methods in terms of robust alignment and planar-structure preservation.

Acknowledgements

Min H. Kim, the corresponding author, gratefully acknowledges grants from National Research Foundation of Korea (2016R1A2B2013031, 2013M3A6A6073718) and additional support by Samsung Electronics (SRFC-IT1402-02), Cross-Ministry Giga KOREA Project (GK17P0200), KOCCA in MCST of Korea, and an ICT R&D program of MSIP/IITP of Korea (R7116-16-1035).

References

[1] Robert Carroll, Maneesh Agrawal, and Aseem Agarwala. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph. (TOG)*, 28(3):43:1–43:9, 2009.

[2] Robert Carroll, Aseem Agarwala, and Maneesh Agrawala. Image warps for artistic perspective manipulation. *ACM Trans. Graph. (TOG)*, 29(4):127:1–127:9, 2010.

[3] Chen-Han Chang, Yoichi Sato, and Yung-Yu Chuang. Shape-preserving half-projective warps for image stitching. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 3254–3261, 2014.

[4] Yu-Sheng Chen and Yung-Yu Chuang. Natural image stitching with the global similarity prior. In *Proc. European. Conf. Comput. Vision (ECCV)*, pages 186–201, 2016.

[5] Junhong Gao, Seon Joo Kim, and Michael S. Brown. Constructing image panoramas using dual-homography warping. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 49–56, 2011.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. June 2004.

[7] J. B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015.

[8] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Image completion using planar structure guidance. volume 33, pages 129:1–10, 2014.

[9] K. Ikeuchi, P. Vasseur, C. Demonceaux, Inso Kweon, Yongduek Seo, J.-C Bazin, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in manhattan world. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 638–645, 2012.

[10] K. Joo, N. Kim, T. H. Oh, and I. S. Kweon. Line meets as-projective-as-possible image stitching with moving dlt. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 1175–1179, 2015.

[11] José Lezama, Rafael Grompone von Gioi, Gregory Randall, and Jean-Michel Morel. Finding vanishing points via point alignments in image primal and dual domains. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 509–515, 2014.

[12] Shiwei Li, Lu Yuan, Jian Sun 0001, and Long Quan. Dual-feature warping-based motion model estimation. In *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, pages 4283–4291, 2015.

[13] Chung-Ching Lin, Sharathchandra Pankanti, Karthikeyan Natesan Ramamurthy, and Aleksandr Y. Aravkin. Adaptive as-natural-as-possible image stitching. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 1155–1163, 2015.

[14] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong Fah Cheong. Smoothly varying affine stitching. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 345–352, 2011. ISBN 978-1-4577-0394-2.

[15] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph. (TOG)*, 28(3):44:1–44:9, 2009.

[16] B. Micusik, H. Wildenauer, and J. Kosecka. Detection and matching of rectilinear structures. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 1–7, June 2008.

[17] Yoshikuni Nomura, Li Zhang, and Shree K. Nayar. Scene collages and flexible camera arrays. In *Proc. European. Conf. Rendering Techniques*, EGSR, pages 127–138, 2007.

[18] A. G. Schwing, T. Hazan, M. Pollefeys, and R. Urtasun. Efficient structured prediction for 3d indoor scene understanding. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 2815–2822, 2012.

[19] Richard Szeliski. Image alignment and stitching: A tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, 2006.

[20] Tianzhu Xiang, Gui-Song Xia, Xiang Bai, and Liangpei Zhang. Image stitching by line-guided local warping with global similarity constraint. *CoRR*, abs/1702.07935, 2017.

[21] J. Zaragoza, T. J. Chin, Q. H. Tran, M. S. Brown, and D. Suter. As-projective-as-possible image stitching with moving DLT. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(7):1285–1298, 2014.

[22] Julio Zaragoza, Tat-Jun Chin, Michael S. Brown, and David Suter. As-projective-as-possible image stitching with moving DLT. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 2339–2346, 2013.

[23] Menghua Zhai, Scott Workman, and Nathan Jacobs. Detecting vanishing points using global image context in a non-manhattanworld. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 5657–5665, 2016.

[24] Fan Zhang and Feng Liu. Parallax-tolerant image stitching. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 3262–3269, 2014.