Quasi-homography Warps in Image Stitching

Nan Li, Yifang Xu, and Chao Wang

***Abstract*—The naturalness of warps is gaining extensive attentions in image stitching. Recent warps such as SPHP and AANAP, use global similarity warps to mitigate projective distortion (which enlarges regions), however, they necessarily bring in perspective distortion (which generates inconsistencies). In this paper, we propose a novel quasi-homography warp, which effectively balances the perspective distortion against the projective distortion in the non-overlapping region to create a more natural-looking panorama. Our approach formulates the warp as the solution of a bivariate system, where perspective distortion and projective distortion are characterized as slope preservation and scale linearization respectively. Because our proposed warp only relies on a global homography, thus it is totally parameter-free. A comprehensive experiment shows that a quasi-homography warp outperforms some state-of-the-art warps in urban scenes, including homography, AutoStitch and SPHP. A user study demonstrates that it wins most users' favor, comparing to homography and SPHP.**

***Index Terms*—Image stitching, image warping, natural-looking, projective distortion, perspective distortion.**

I. Introduction

IMAGE stitching plays an important role in many multimedia applications, such like panoramic videos [1]–[3], virtual reality [4]–[6]. Conventionally, image stitching is a process of composing multiple images with overlapping fields of views, to produce a wide-view panorama [7], where the first stage is to determine a warp for each image and transform it into a common coordinate system, then the warped images are composed [8]–[12] and blended [13]–[15] into a final mosaic. Evaluations of warping include the alignment quality in the overlapping region and the naturalness quality in the non-overlapping region.

Early warps focus on the alignment quality, which is measured in two different aspects:

- (**global**) the root mean squared error on the set of feature correspondences,
- (**local**) the patch-based mean error along a stitching seam.

Global warps such as similarity or homography warps [16], aim to minimize alignment errors between overlapping pixels via a uniform transformation. Homography is the most frequently used warp, because it is the most flexible planar transformation which preserves all straight lines. For a better global alignment quality, recent spatially-varying warps [17]–[20] use multiple local transformations instead of a single global one to address the large parallax issue in the overlapping region. Some seam-driven warps [21]–[23] address the same problem by pursuing a better local alignment quality in the overlapping region, such that there exists a local region to be seamlessly blended.

N. Li is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China. E-mail: nan@tju.edu.cn.

Y. Xu is with the Center for Combinatorics, Nankai University, Tianjin 300071, China. Email: xyf@mail.nankai.edu.cn.

C. Wang is with the Department of Software, Nankai University, Tianjin 300071, China. Email: wangchao@nankai.edu.cn.

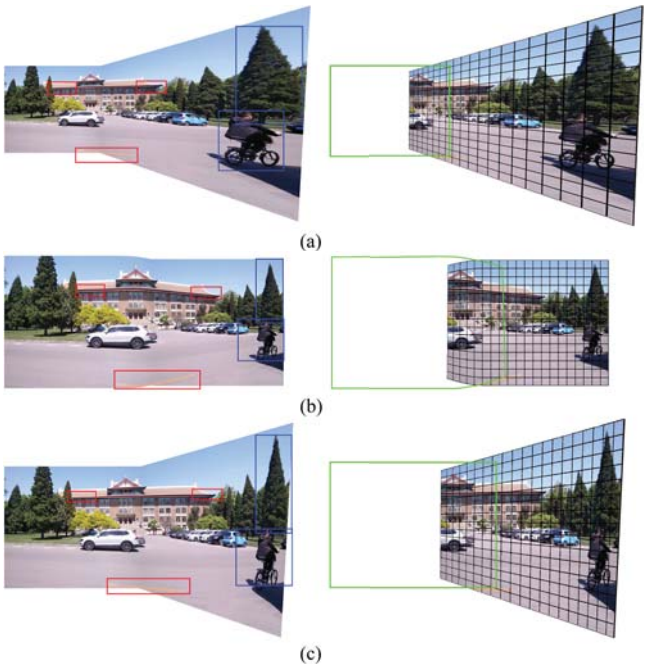

Fig. 1. A distortion comparison among different warps in the non-overlapping region, which use the same homography alignment and the same seam-cutting composition in the overlapping region. In the stitching results, we use blue rectangles to highlight the comparison of projective distortion, and red rectangles to highlight the comparison of perspective distortion. (a) Homography. (b) SPHP. (c) Our warp. Homography is of global consistency but suffers from projective distortion. SPHP is of local consistency but suffers from perspective distortion. Our warp is partially of global and local consistencies and balances projective and perspective distortion.

Recent warps concentrate more on the naturalness quality, which is embodied in two consistency properties:

- (**local**) the region of any object should be consistent with its appearance in the original,
- (**global**) the perspective relation of the same object should be consistent between different images.

Violations of consistencies lead to two types of distortion:

- (**projective**) the region of an object is enlarged, compared to its appearance in the original (see people and trees in Fig. 1(a)),
- (**perspective**) the perspectives of an object in two images are inconsistent with each other (see buildings and signs in Fig. 1(b)).

Similarity warps automatically satisfy the local consistency, since they purely involve translation, rotation and uniformly scaling, but may suffer from perspective distortion. Homography warps conventionally satisfy the global consistency, if a

good alignment quality is guaranteed, but may suffer from projective distortion. Warps such like the shape-preserving half-projective (SPHP) warp [24] and the adaptive as-natural-as-possible (AANAP) warp [25] use a spatial combination of homography and similarity warps to mitigate projective distortion in the non-overlapping region. Other warps address the same problem via a joint optimization of alignment and naturalness qualities, which either constrains the warp resembles a similarity as a whole [26], or constrains the warp preserves detected straight lines [27].

In this paper, we propose a *quasi-homography* warp, which balances the projective distortion against the perspective distortion in the non-overlapping region, to create a more natural-looking mosaic (see Fig. 1(c)). Our proposed warp only relies on a global homography, thus it is totally parameter-free. The rest of the paper is organized as follows. Section II describes some recent works. Section III provides a naturalness analysis of image warps, where Section III-A presents two intuitive tools to demonstrate projective distortion and perspective distortion via mathematical derivations. Section III-B and III-C employ such tools to analyze homography and SPHP warps in aspects of local and global consistencies. Our quasi-homography warp is defined as a solution of a bivariate system in Section IV-B, which is based on a new formulation of homography in Section IV-A. Implementation details (including two-image stitching and multiple-image stitching) and method variations (including orientation rectification and partition refinement) are proposed in Section V. Section VI presents a comparison experiment and a user study, which demonstrate that the quasi-homography warp not only outperforms some state-of-the-art warps in urban scenes, but also wins most users' favor. Finally, conclusions are drawn in Section VII and some mathematical formulas are explained in Appendix.

II. Related Work

In this section, we review some recent works of image warps in aspects of alignment and naturalness qualities respectively. For more fundamental concepts about image stitching, please refer to a comprehensive survey [7] by Szeliski.

A. Warps for Better Alignment

Conventional stitching methods always employ global warps such as similarity, affine and homography, to align images in the overlapping region [16]. Global warps are robust but often not flexible enough to provide accurate alignment. Gao *et al.* [17] proposed a dual-homography warp to address scenes with two dominant planes by a weighted sum of two homographies. Lin *et al.* [18] proposed a smoothly varying affine (SVA) warp to replace a global affine warp with a smoothly affine stitching field, which is more flexible and maintains much of the motion generalization properties of affine or homography. Zaragoza *et al.* [19] proposed an as-projective-as-possible (APAP) warp in a moving DLT framework, which is able to accurately register images that differ by more than a pure rotation. Lou *et al.* [20] proposed a piecewise alignment method, which approximates regions of image with planes by incorporating piecewise local geometric models.

Other methods combine image alignment with seam-cutting approaches [28]–[31], to find a locally registered area which is seamlessly blended instead of aligning the overlapping region globally. Gao *et al.* [21] proposed a seam-driven framework, which searches a homography with minimal seam costs instead of minimal alignment errors on a set of feature correspondences. Zhang and Liu [22] proposed a parallax-tolerant warp, which combines homography and content-preserving warps to locally register images. Lin *et al.* [23] proposed a seam-guided local alignment warp, which iteratively improves the warp by adaptive feature weighting according to the distance to current seams.

B. Warps for Better Naturalness

Many efforts have been devoted to mitigate distortion in the non-overlapping region for creating a natural-looking mosaic. A pioneering work [32] uses spherical or cylindrical warps to produce multi-perspective results to address this problem, but it necessarily curves straight lines.

Recently, some methods take advantage of global similarity (preserves the original perspective) to mitigate projective distortion in the non-overlapping region. Chang *et al.* [24] proposed a SPHP warp that spatially combines a homography warp and a similarity warp, which makes the homography maintain good alignment in the overlapping region while the similarity keep the original perspective in the non-overlapping region. Lin and Pankanti [25] proposed an AANAP warp, which combines a linearized homography warp and a global similarity warp with the smallest rotation angle to create natural-looking mosaics.

Other methods model their warps as mesh deformations via energy minimization, which address naturalness quality issues by enforcing different constraints. Chen *et al.* [26] proposed a global-similarity-prior (GSP) warp, which constrains the warp resembles a similarity as a whole. Zhang *et al.* [27] proposed a warp that produces an orthogonal projection of a wide-baseline scene by constraining it preserves extracted straight lines, and allows perspective corrections via scale preservation.

III. Naturalness Analysis of Image Warps

This section describes a naturalness analysis of image warps. First, the global consistency is characterized as *line-preserving*, where the perspective distortion is illustrated through a mesh-to-mesh transformation, and the local consistency is characterized as *uniformly-scaling*, where the projection distortion is demonstrated via the linearity of a scaling function. Then, we analyze the naturalness of homography and SPHP warps by these tools.

A. Mathematical Setup

Let I and I' denote the *target* image and the *reference* image respectively. A warp $\mathcal{H}$ is a planar transformation [16], which relates pixel coordinates $(x, y) \in I$ to $(x', y') \in I'$, where

$$\begin{cases} x' = f(x, y) \\ y' = g(x, y) \end{cases} . \tag{1}$$

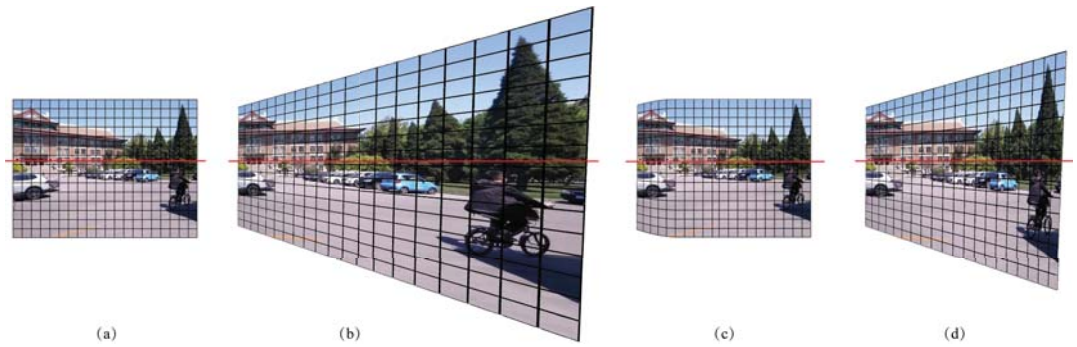

Fig. 2. Perspective distortion v.s. Slope preservation. (a) Mesh of the target image. (b) Warped mesh of homography. (c) Warped mesh of SPHP. (d) Warped mesh of our result. The red line is the special horizontal line that remains horizontal under a homography. Note that homography preserves arbitrary straight lines but incrementally increases areas of meshes along the red line, while SPHP maintains shapes of meshes by a uniform scaling factor but gradually changes the slope of straight lines. Our warp relaxes arbitrary line-preserving to only preserving the slope of the mesh, while relaxes uniformly-scaling everywhere to only uniforming the density of the mesh on the red line, to show a balance of perspective distortion and projective distortion in the non-overlapping region.

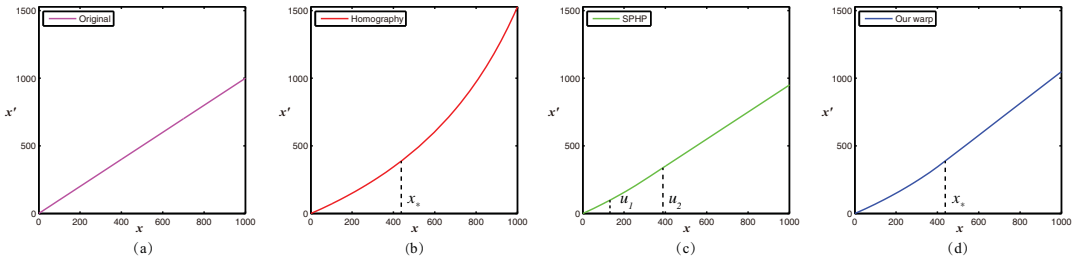

Fig. 3. Projective distortion v.s. Scale linearization. It demonstrates scaling functions of different warps on the special horizontal line that remains horizontal under a homography (marked red in Fig. 2), where x_* corresponds to the closest vertical partition line that isolates $\mathcal{O}$, and u_1, u_2 correspond to two partition lines that divide $\mathbb{R}^2$ into three regions in (13). Note that the scaling function is a rational one in homography (11), while it is a piece-wise function in SPHP which consists of a rational one as same as homography, a quadratic one and a linear one. The scaling function in our warp consists of a rational one as same as homography and a linear one.

If $\mathcal{H}$ is of global consistency, then it must be *line-preserving*, i.e., a straight line $l = \{(x+z, y+kz)|z \in \mathbb{R}\} \in I$ should be mapped to a straight line $l' = \{(x'+z', y'+k'z')|z' \in \mathbb{R}\} \in I'$. Actually, the calculation of the slope k' provides a criterion to validate line-preserving, i.e., $\mathcal{H}$ is line-preserving, if and only if

$$k' = \frac{g(x+z, y+kz) - g(x,y)}{f(x+z, y+kz) - f(x,y)} \tag{2}$$

is independent of z. Its proof is easy. Given a point $(x, y) \in I$ and a slope k, then they define a straight line $l = \{(x+z, y+kz)|z \in \mathbb{R}\} \in I$. If k' calculated by (2) is a constant, then l is mapped to a straight line $l' = \{(x'+z', y'+k'z')|z' \in \mathbb{R}\} \in I'$, which is defined by $(x', y') \in I'$ and k'. Since k' only depends on (x, y) and k, we denote it by $\text{slope}(x, y, k)$.

Suppose $\mathcal{H}$ is line-preserving and $\mathcal{C}^1$ continuous, then

$$\begin{aligned}\text{slope}(x,y,k) &= \lim_{z\to 0} \frac{g(x+z, y+kz) - g(x,y)}{f(x+z, y+kz) - f(x,y)} \\ &= \frac{g_x(x,y) + k g_y(x,y)}{f_x(x,y) + k f_y(x,y)},\end{aligned} \tag{3}$$

where f_x, f_y, g_x, g_y denote the partial derivatives of f and g. In fact, there exists a mesh-to-mesh transformation that maps all horizontal lines and vertical lines to straight lines with slopes

$$\text{slope}(x,y,0) = \frac{g_x(x,y)}{f_x(x,y)}, \tag{4}$$

$$\text{slope}(x,y,\infty) = \frac{g_y(x,y)}{f_y(x,y)}, \tag{5}$$

which are independent of x and y respectively. Consequently, any point $(x, y) \in I$ can be expressed as the intersection point of a horizontal line and a vertical line, which is corresponding to the point $(x', y') \in I'$ as the intersection point of two lines with slopes $\text{slope}(x, y, 0)$ and $\text{slope}(x, y, \infty)$. In the rest of the paper, we constantly employ this mesh-to-mesh transformation to demonstrate perspective distortion comparisons among different warps (see Fig. 2).

On the other side, if $\mathcal{H}$ is of local consistency, then it must be *uniformly-scaling*. Actually, the local consistency automatically holds for similarity warps, because they purely involve translation, rotation and uniformly scaling. Suppose $\mathcal{H}$ is not only line-preserving but also uniformly-scaling, then a line segment $s = \{(x+z, y+kz)|z \in [z_1, z_2]\} \in I$ should be mapped to a line segment $s' = \{(x'+z', y'+k'z')|z' \in [z'_1, z'_2]\} \in I'$ with a uniform scaling factor. Conversely, the linearity of a scaling function on arbitrary line is a necessary condition of the local consistency.

By assuming cameras are oriented and motions are horizontal, there should exist a horizontal line $l_x = \{(x, y_*)|x \in \mathbb{R}\} \in I$ which remains a horizontal line $l'_x = \{(x', y'_*)|x' \in \mathbb{R}\} \in I'$, if a good alignment is guaranteed. In fact, l_x is roughly located

in the horizontal plane of cameras, and y_* satisfies

$$\text{slope}(x, y_*, 0) = \frac{g_x(x, y_*)}{f_x(x, y_*)} = 0. \tag{6}$$

Given a point $(x_*, y_*) \in l_x$, then for $\forall (x, y_*) \in l_x$, $|f(x, y_*) - f(x_*, y_*)|$ should equal to a uniform scaling factor times $|x - x_*|$. In other words, $f(x, y_*)$ should be linear in x. In the rest of the paper, we constantly employ the linearity to demonstrate projective distortion comparisons among different warps (see Fig. 3).

Some other notations are stated as follows. Let $\mathcal{O}$ denote the overlapping region and $l_y = \{(x_*, y) \,|\, y \in \mathbb{R}\}$ denote a vertical line which divides $\mathbb{R}^2$ into half spaces $R_O = \{(x, y) | x \leq x_*\}$ and $R_Q = \{(x, y) | x_* < x\}$, such that $\mathcal{O} \subset R_O$. Our proposed warp $\mathcal{H}_\dagger$ is a spatial combination of a homography warp $\mathcal{H}_0$ within R_O and a squeezed homography warp $\mathcal{H}_*$ within R_Q, where R'_O and R'_Q are respective half spaces after warping.

B. Naturalness Analysis of Homography

A homography warp $\mathcal{H}_0$ is the most flexible warp for better alignment, which is normally defined as

$$f_0(x, y) = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1}, \tag{7}$$

$$g_0(x, y) = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1}, \tag{8}$$

where h_1-h_8 are eight parameters. It is easy to certify that $\mathcal{H}_0$ is line-preserving, since the slope k' in (2) is independent of z. To illustrate the property more intuitively, we draw a mesh-to-mesh transformation (see Fig. 2(b)), where horizontal lines and vertical lines are mapped to straight lines with slopes

$$\text{slope}(x, y, 0) = \frac{(h_4 h_8 - h_5 h_7) y + (h_4 - h_6 h_7)}{(h_1 h_8 - h_2 h_7) y + (h_1 - h_3 h_7)}, \tag{9}$$

$$\text{slope}(x, y, \infty) = \frac{(h_4 h_8 - h_5 h_7) x + (h_6 h_8 - h_5)}{(h_1 h_8 - h_2 h_7) x + (h_3 h_8 - h_2)}. \tag{10}$$

Under the assumption that cameras are oriented and motions are horizontal, for $l_x = \{(x, y_*) | x \in \mathbb{R}\} \in I$, we derive

$$y_* = \frac{h_6 h_7 - h_4}{h_4 h_8 - h_5 h_7}, \tag{11}$$

by solving the equation (6). For $\forall (x, y_*) \in l_x$,

$$f_0(x, y_*) = \frac{h_1 x + h_2 y_* + h_3}{h_7 x + h_8 y_* + 1}, \tag{12}$$

is non-linear in x when $h_7 \neq 0$ (see Fig. 3(b)), which indicates the invalidation of uniformly-scaling.

In summary, homography warps conventionally satisfy the global consistency if a good alignment is guaranteed, however they usually suffer from projective distortion in the non-overlapping region (see Table I). For example, the people and the tree are enlarged in Fig. 1(a) comparing to the original.

C. Naturalness Analysis of SPHP

To overcome such drawbacks of homography warps, Chang *et al.* [24] proposed a shape-preserving half-projective (SPHP) warp, which is a spatial combination of a homography warp and a similarity warp, to create a natural-looking multi-perspective panorama.

Specifically, after adopting the change of coordinates, SPHP divides $\mathbb{R}^2$ into three regions. 1. $R_H = \{(u, v) | u \leq u_1\}$, where a homography warp is applied to achieve a good alignment. 2. $R_S = \{(u, v) \,|\, u_2 \leq u\}$, where a similarity warp is applied to mitigate projective distortion. 3. $R_T = \{(u, v) \,|\, u_1 < u < u_2\}$, a buffer region where a warp is applied to gradually change a homography warp to a similarity warp. Consequently, a SPHP warp $\mathcal{W}$ is defined as

$$w(u, v) = \begin{cases} H(u, v), & \text{if } (u, v) \in R_H \\ T(u, v), & \text{if } (u, v) \in R_T \\ S(u, v), & \text{if } (u, v) \in R_S \end{cases}, \tag{13}$$

where u_1 and u_2 are parameters, such that $\mathcal{W}$ can approach a similarity warp as much as possible. Note that, the change of coordinates plays an important role in SPHP, since a similarity simply combines a homography via a single partition line.

Both homography and similarity are line-preserving, thus $\mathcal{W}$ is certainly of global consistency in R_H and R_S respectively. However, $\mathcal{W}$ may suffer from line-bending within R_T, because of its non-linearity. Moreover, perspectives of R_H and R_S may contradict each other. For example, parallels remain parallels in R_S, while they do not in R_H (see Fig. 2(c)). $\mathcal{W}$ is certainly of local consistency in R_S, because a similarity warp is applied (see Fig. 3(c)).

In summary, SPHP warps achieve the alignment quality as good as homography warps in R_H, and the local consistency as good as similarity warps in R_S. However, SPHP warps may suffer from line-bending in R_T and perspective distortion between R_H and R_S (see Table I). Note that, the non-linearity of $T(u, v)$ in R_T merely blends certain lines in theory, but it is still possible to preserve visible straight lines in practice. Many results in [24] justify that SPHP is capable of doing so. Unfortunately, it will get worse for urban scenes, which are filled with visible lines and visible parallels (see the sign in Fig. 1(b)). It is also worth noting that, SPHP creates a multi-perspective panorama, thus different perspectives may contradict each other (see buildings in Fig. 1(b)).

These naturalness analysis of homography and SPHP warps motivate us to construct a warp, which achieves a good balance between the perspective distortion and the projective distortion in the non-overlapping region, via relaxing the local and global consistencies such that they are both partially satisfied.

TABLE I
NATURALNESS ANALYSIS OF DIFFERENT WARPS.

Methods	Naturalness quality	
	Local consistency	Global consistency
Homography	invalid	perfect
SPHP [24]	perfect	invalid
Quasi-homography	partial	partial

IV. PROPOSED WARPS

This section presents how to construct a warp for balancing perspective distortion against projective distortion in the non-overlapping region. First, we propose a different formulation

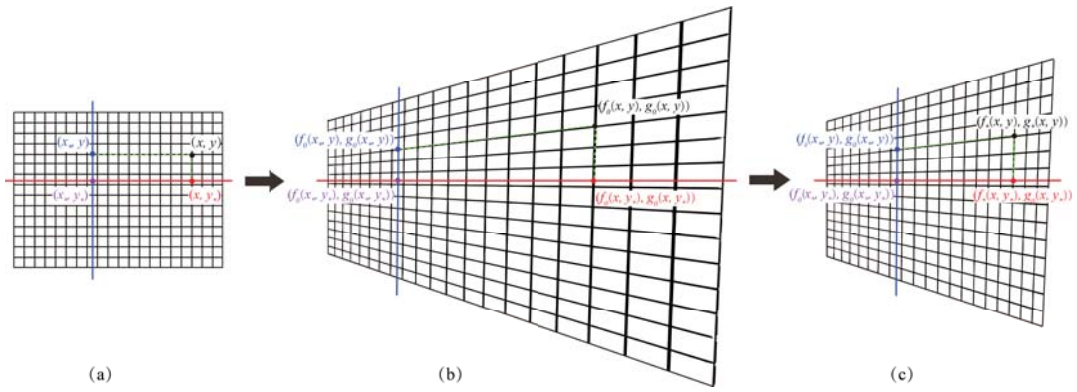

Fig. 4. Quasi-homography v.s. Homography. (a) Target image. (b) Reformulation of homography. (c) Derivation of quasi-homography. In the target image, any point (x, y) can be expressed as the intersection point of a horizontal line and a vertical line, which is corresponding to a point (x', y') under a homography as the intersection point of two lines with slopes $\text{slope}(x, y, 0)$ and $\text{slope}(x, y, \infty)$. The location of (x', y') can be controlled by the density on a horizontal line (marked in red) and a vertical line (marked in blue). Note that quasi-homography linearizes the density on the red line but without changing the density on the blue line, such that it combines the homography by a single partition line and squeezes the meshes of homography without varying the shape.

of the homography warp to characterize the global consistency as slope preservation while the local consistency as scale linearization respectively. Then, we describe how to adopt this formulation to present a *quasi-homography* warp, which squeezes the mesh of the corresponding homography warp but without varying its shape.

A. Review of Homography

Given eight parameters h_1-h_8, we formulate a homography warp $\mathcal{H}_0$ in another way, as the solution of a bivariate system

$$\frac{y' - g_0(x_*, y)}{x' - f_0(x_*, y)} = \text{slope}(x, y, 0), \tag{14}$$

$$\frac{y' - g_0(x, y_*)}{x' - f_0(x, y_*)} = \text{slope}(x, y, \infty), \tag{15}$$

where (x_*, y) and (x, y_*) are projections of a point (x, y) onto $l_y = \{(x_*, y) \,|\, y \in \mathbb{R}\}$ and $l_x = \{(x, y_*) \,|\, x \in \mathbb{R}\}$ respectively (see Fig. 4(a)). Besides, equations of f_0, g_0 and $\text{slope}(x, y, 0)$, $\text{slope}(x, y, \infty)$ are given in (7,8,9,10).

Our formulation (14,15) is equivalent to (7,8). In fact, it is easy to check that (7,8) is a solution of (14,15). Furthermore, it is the unique solution, because the Jacobian is invertible if and only if $\text{slope}(x, y, 0) \neq \text{slope}(x, y, \infty)$. Then, comparing with (7,8), our formulation (14,15) characterizes the global consistency as slope preservation while the local consistency as scale linearization respectively. Intuitively, $\text{slope}(x, y, 0)$ and $\text{slope}(x, y, \infty)$ formulate the shape of the mesh, while $f_0(x, y_*)$ and $g_0(x_*, y)$ formulate the density of the mesh (see Fig. 4(b)). It should be noticed that we made no assumptions on x_* or y_* in the above analysis. In the next subsection, we will assume that l_y isolates the overlapping region $\mathcal{O}$ and l_x remains horizontal under $\mathcal{H}_0$, for stitching multiple images captured by oriented cameras via horizontal motions.

B. Quasi-homography

Our proposed warp makes use of the formulation (14,15) to balance perspective distortion against projective distortion in the non-overlapping region. First, we divide $\mathbb{R}^2$ by the vertical line $l_y = \{(x_*, y) \,|\, y \in \mathbb{R}\}$ into half spaces $R_O = \{(x, y) | x \leq x_*\}$ and $R_Q = \{(x, y) | x_* < x\}$, where the overlapping region $\mathcal{O} \subset R_O$. Then, we formulate our warp $\mathcal{H}_\dagger$ as the solution of a bivariate system

$$\frac{y' - g_0(x_*, y)}{x' - f_0(x_*, y)} = \text{slope}(x, y, 0), \tag{16}$$

$$\frac{y' - g_0(x, y_*)}{x' - f_\dagger(x, y_*)} = \text{slope}(x, y, \infty), \tag{17}$$

where y_* satisfies (11) and $f_\dagger(x, y_*)$ is defined as

$$f_\dagger(x, y_*) = \begin{cases} f_0(x, y_*), & \text{if } (x, y_*) \in R_O, \\ f_*(x, y_*), & \text{if } (x, y_*) \in R_Q, \end{cases} \tag{18}$$

$$f_*(x, y_*) = f_0(x_*, y_*) + f_0'(x_*, y_*)(x - x_*), \tag{19}$$

on the horizontal line $l_x = \{(x, y_*) \,|\, x \in \mathbb{R}\}$. In fact, $f_*(x, y_*)$ is the first-order truncation of the Taylor's series for $f_0(x, y_*)$ at $x = x_*$, which successfully makes $f_\dagger(x, y_*)$ piece-wise $\mathcal{C}^1$ continuous and linear in x within R_Q.

Because the Jacobian of (16,17) is invertible, it possesses a unique solution $\mathcal{H}_\dagger$ as

$$\mathcal{H}_\dagger = \begin{cases} \mathcal{H}_0, & \text{if } (x, y) \in R_O \\ \mathcal{H}_*, & \text{if } (x, y) \in R_Q \end{cases}, \tag{20}$$

where

$$x' = f_\dagger(x, y) = \begin{cases} f_0(x, y), & \text{if } (x, y) \in R_O, \\ f_*(x, y), & \text{if } (x, y) \in R_Q, \end{cases} \tag{21}$$

$$y' = g_\dagger(x, y) = \begin{cases} g_0(x, y), & \text{if } (x, y) \in R_O, \\ g_*(x, y), & \text{if } (x, y) \in R_Q, \end{cases} \tag{22}$$

where $f_*(x, y)$ and $g_*(x, y)$ are rational functions in variables x and y, whose coefficients are polynomial functions in h_1-h_8 and x_*. The detailed derivations are presented in Appendix. In fact, the warp $\mathcal{H}_\dagger$ just squeezes the meshes of homography in the horizontal direction but without varying its shape (see Fig. 4(c)). In this sense, we call $\mathcal{H}_\dagger$ a *quasi-homography* warp that corresponds to a homography warp $\mathcal{H}_0$. A quasi-homography warp maintains good alignment in R_O as a homography warp,

and it mitigates perspective distortion and projective distortion simultaneously via slope preservation and scale linearization in R_Q. Intuitively, $\mathcal{H}_\dagger$ relaxes arbitrary line-preserving to only preserving the shape of the mesh (see Fig. 2(d)), while relaxes uniformly-scaling everywhere to only uniforming the density of the mesh on l_x in R_Q (see Fig. 3(d)).

On the other hand, since $\mathcal{H}_\dagger$ just squeezes the mesh of $\mathcal{H}_0$ but without varying its shape, $\mathcal{H}_\dagger$ is an injection if $\mathcal{H}_0$ is an injection. Given $(x', y') \in R'_O$, then $(x, y) \in R_O$ is determined by $\mathcal{H}_0^{-1}$. Given $(x', y') \in R'_Q$, then $(x, y) \in R_Q$ is determined by solving (16,17) (regard x, y as unknowns)

$$x = \text{RootOf}(m_1 x^2 + m_2 x + m_3), \tag{23}$$

$$y = \frac{(h_6h_7 - h_4)x' + (h_1 - h_3h_7)y' + (h_3h_4 - h_1h_6)}{(h_4h_8 - h_5h_7)x' + (h_2h_7 - h_1h_8)y' + (h_1h_5 - h_2h_4)}, \tag{24}$$

where m_1-m_3 are polynomial functions in x', y', x_*, and h_1-h_8. The detailed derivations are presented in Appendix.

Note that, though both SPHP and quasi-homography warps adopt a spatial combination of a homography warp and another warp to create more natural-looking mosaics, their motivations and frameworks are different. SPHP focuses on the local consistency, to create a natural-looking multi-perspective panorama. Quasi-homography concentrates on balancing global and local consistencies, to generate a natural-looking single-perspective panorama. SPHP introduces a change of coordinates such that a similarity combines a homography via a single partition line, and a buffer region such that a homography gradually changes into a similarity. Quasi-homography reorganizes homography's point correspondences via solving the bivariate system (16,17), where the shape is preserved and the size is squeezed.

It is worth noting that the construction of quasi-homography makes no assumptions on the special horizontal line l_x and the vertical partition line l_y. For stitching multiple images captured by oriented cameras via horizontal motions, the horizontal line that remains horizontal best measures the projective distortion. Therefore, quasi-homography can preserve horizontal lines or nearly-horizontal lines better than SPHP (see result comparisons in Section VI-A), and ordinary users prefer such stitching results in urban scenes (see user study in Section VI-B).

In summary, quasi-homography warps achieve a good alignment quality as homography warps in R_O, while partially possess the local consistency and the global consistency in R_Q, such that perspective distortion and projective distortion are balanced (see Table I). Note that the warp may still suffer from diagonal line-bending and vertical region-enlarging within R_Q, because line-preserving and uniformly-scaling are relaxed to partially valid. Please see more details in Section VI-C.

V. Implementation

In this section, we first present more implementation details of our quasi-homography in two-image stitching and multiple-image stitching, then we propose two variations of the method including orientation rectification and partition refinement.

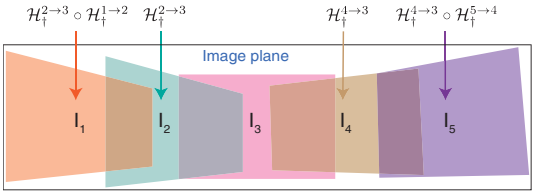

Fig. 5. A sketch for stitching multiple images.

A. Two-image Stitching

Given a pair of two images, which are captured by oriented cameras via horizontal motions, if a homography warp $\mathcal{H}_0$ can be estimated with a good alignment quality in the overlapping region, then a quasi-homography warp $\mathcal{H}_\dagger$ can be calculated, which smoothly extrapolates from $\mathcal{H}_0$ in R_O into $\mathcal{H}_*$ in R_Q. A brief algorithm is given in Algorithm 1.

Algorithm 1 Two-image stitching using quasi-homography.

Input: two images taken by oriented cameras via horizontal motions.

Output: one horizontally stitched image.

1) Use SIFT [33] to extract and match features;
2) Use RANSAC [34] to estimate a global homography $\mathcal{H}_0$;
3) Calculate a quasi-homography warp:
   a) Calculate a forward map $\mathcal{H}_\dagger$ (21,22) to get the canvas;
   b) Calculate a backward map $\mathcal{H}_\dagger^{-1}$ (23,24) for filling the canvas by bilinear interpolations;
4) Use seam-cutting [28] to blend the overlapping region.

B. Multiple-image Stitching

Given a sequence of multiple images, which are captured by oriented cameras via horizontal motions, our warping method consists of three stages. In the first stage, we pick a reference image as a standard perspective, such that other images should be consistent with it. Then we estimate a homography warp for each image and transform them in the coordinate system of the reference image via bundle adjustment as in [35] and calculate pairwise quasi-homography warps of adjacent images. Finally, we concatenate other images to the reference one by a chained composite map of pairwise quasi-homgraphy warps.

Fig. 5 illustrates an example of the concatenation procedure for stitching five images. First we select I_3 as the reference one such that perspectives of other four images should agree with it. Then we estimate homography warps $\mathcal{H}_0^{1\to3}$, $\mathcal{H}_0^{2\to3}$, $\mathcal{H}_0^{4\to3}$, $\mathcal{H}_0^{5\to3}$ via bundle adjustment [35] and calculate pairwise quasi-homography warps $\mathcal{H}_\dagger^{1\to2}$, $\mathcal{H}_\dagger^{2\to3}$, $\mathcal{H}_\dagger^{4\to3}$, $\mathcal{H}_\dagger^{5\to4}$. Finally, we concatenate I_1 and I_5 to I_3 by

$$\begin{cases} \mathcal{H}_\dagger^{1\to3} = \mathcal{H}_\dagger^{2\to3} \circ \mathcal{H}_\dagger^{1\to2}, \\ \mathcal{H}_\dagger^{5\to3} = \mathcal{H}_\dagger^{4\to3} \circ \mathcal{H}_\dagger^{5\to4}. \end{cases} \tag{25}$$

Therefore, the concatenation warp for every image is a chained composite map of pairwise quasi-homography warps.

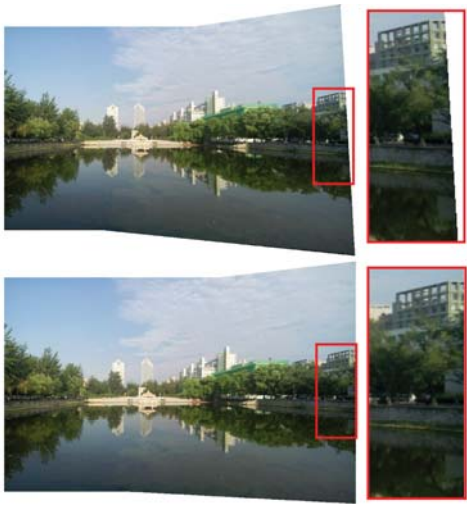

Fig. 6. An example of orientation rectification.

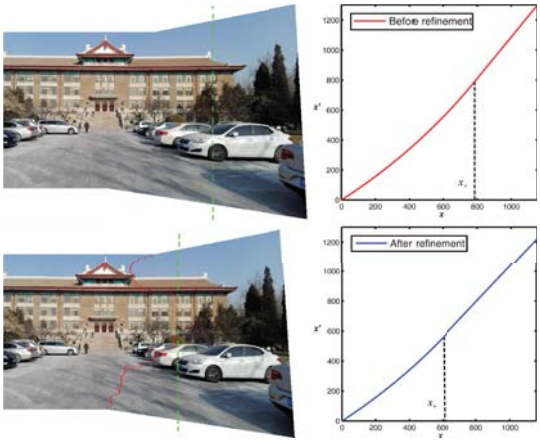

Fig. 7. An example of partition refinement.

C. Orientation Rectification

In urban scenes, users accustom to taking pictures by oriented cameras via horizontal motions, hence any vertical line in the target image is expected to be transformed to a vertical line in the warped result. However, it inevitably sacrifices the alignment quality in the overlapping region.

In order to achieve orientation rectification, we incorporate an extra constraint in the homography estimation, which constrains that the external vertical boundary of the target image preserves vertical in the warped result. Then for a homography warp $\mathcal{H}_0$ as (7,8), it should satisfy

$$f_0(w,0) = f_0(w,h) \;\Leftrightarrow\; h_8 = \frac{h_2(h_7 w + h_9)}{h_1 w + h_3}, \tag{26}$$

where w and h are the width and the height of I respectively. A global homography is then estimated by solving

$$\min \sum_{i=1}^{N} \|\mathbf{a}_i \mathbf{h}\|^2 \text{ s.t. } \|\mathbf{h}\| = 1, h_8 = \frac{h_2(h_7 w + h_9)}{h_1 w + h_3}. \tag{27}$$

Because the quasi-homography warp just squeezes the mesh of a homography warp but without varying its shape, the external vertical boundary still preserves vertical (see Fig. 6).

D. Partition Refinement

In our analysis of quasi-homography warps in Section IV-B, the uniform scaling factor on the special horizontal line l_x in R_Q depends on the linearized scaling function (19). Moreover, it depends on the determination of the partition point (x_*, y_*). In fact, the factor is more accurate if (x_*, y_*) is better aligned.

Hence, we replace the partition line l_y closest to the border of the overlapping region and the non-overlapping region, by it closest to the external border of the seam for further refinement (see Fig. 7).

VI. Experiments

We experimented our proposed method on a range of images captured through both rear and front cameras in urban scenes. In our experiments, we employ SIFT [33] to extract and match features, RANSAC [34] to estimate a global homography, and seam-cutting [28] to blend the overlapping region. Codes are implemented in OpenCV 2.4.9 and generally take 1s to 2s on a desktop PC with Intel i5 3.0GHz CPU and 8GB memory to stitching two images with 800×600 resolution by Algorithm 1, where the calculation of the quasi-homography warp only takes 0.1s (including the forward map and the backward map). We used the codes of AutoStitch[1] and SPHP[2] from the authors' homepage in the experiment.

A. Result Comparisons

We compared our quasi-homography warp to state-of-the-art warps in urban scenes, including homography, AutoStitch and SPHP. Because our method focuses on the naturalness quality in the non-overlapping region, we only compare with methods using global homography alignment in the overlapping region, while not comparing to methods using spatially-varying warps. Nevertheless, some urban scenes with repetitive structures still cause alignment issues [36], which may limit the application of our method. Therefore, we use a more robust feature matching method RepMatch [37], and a more robust RANSAC solution USAC [38] for estimating a global homography, to generalize our proposed method in urban scenes. Non-planar scenes may cause outlier removal issues [17], but fortunately, [39] justifies that a simple RANSAC-driven homography still works reasonably well even for such cases.

In order to highlight the comparison of the naturalness quality in the non-overlapping region, for homography, SPHP and quasi-homography, we use the same homography alignment and the same seam-cutting composition in the overlapping region.

Fig. 8 illustrates a naturalness comparison for stitching two and three images from data sets of DHW [17] and SPHP [24]. Homography preserves straight lines, but it enlarges the regions of cars and people. SPHP preserves respective perspectives, but it causes contradictions in the ground and

[1] http://matthewalunbrown.com/autostitch/autostitch.html
[2] http://www.cmlab.csie.ntu.edu.tw/∼frank/

wires. AutoStitch uses a spherical projection to produce a multi-perspective stitching result. Quasi-homography uses a planar projection to produce a single-perspective stitching result, which appears as oriented line-preserving and uniformly-scaling. More results from other data sets including DHW [17], SPHP [24], GSP [26] and APAP [35] are available in the supplementary material.

Fig. 9 illustrates a naturalness comparison for stitching two sequences of ten and nine images. Homography stretches cars, trees and people. AutoStitch presents a nonlinear-view stitching result. Quasi-homography creates a natural-looking linear-view stitching result. More results for stitching long sequences of images are available in the supplementary material.

B. User Study

To investigate whether quasi-homography is more preferred by users in urban scenes, we conduct a user study to compare our results to homography and SPHP. We invite 17 participants to rank 20 unannotated groups of stitching results, including 5 groups from the front cameras and 15 groups from rear ones. For each group, we adopt the same homography alignment and the same seam-cutting composition, and all parameters are set to produce optimal final results. In our study, each participant ranks three unannotated stitching results in each group, and a score is recorded by assigning weights 4, 2 and 1 to Rank 1, 2 and 3. Twenty groups of stitching results are available in the supplementary material.

Table II shows a summary of rank votes and total scores for three warps, and the histogram of three scores is shown in Fig. 10 in three aspects. This user study demonstrates that stitching results of quasi-homography warps win most users' favor in urban scenes.

TABLE II
SCORE RESULTS OF USER STUDY.

Methods	Results			
	Rank 1	Rank 2	Rank 3	Total score
Homography	69	210	61	757
SPHP [24]	19	56	265	453
Quasi-homography	252	74	14	1170

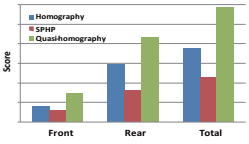

Fig. 10. Histogram of scores.

C. Failure Cases

Experiments show that quasi-homography warps usually balance the projective distortion against the perspective distortion in the non-overlapping region, but there still exist some limitations. For example, diagonal lines may not stay straight anymore and regions of objects may suffer from vertical stretches (especially for stitching images from different planes). Two of failure examples are shown in Fig. 11.

VII. CONCLUSION

In this paper, we propose a quasi-homography warp, which balances the perspective distortion against the projective distortion in the non-overlapping region, to create natural-looking single-perspective panoramas. Experiments show that stitching results of quasi-homography outperform some state-of-the-art warps under urban scenes, including homography, AutoStitch and SPHP. A user study demonstrates that quasi-homography wins most users' favor as well, comparing to homography and SPHP.

Fig. 11. Failure examples. (a) Diagonal lines are bent. (b) People are vertically stretched.

Future works include generalizing quasi-homography warps into spatially-varying warping frameworks like in [26], [27] to improve alignment qualities as well in the overlapping region, and incorporating our method in SPHP to create more natural-looking multi-perspective panoramas. Multimedia applications that are relevant to image stitching could also be considered, such as feature selection [40], video composition [41], cross-media stitching [42] and action recognition [43].

APPENDIX

The forward map (21,22) and the backward map (23,24) of a quasi-homography warp (16,17) are solved by the computer algebra system *Maple* via commands

$$\text{solve}\left(\{(16),(17)\},\{x',y'\}\right), \tag{28}$$

where x', y' are unknowns and x, y, x_*, y_*, h_1-h_8 are parameters,

$$\text{solve}\left(\{(16),(17)\},\{x,y\}\right) \tag{29}$$

where x, y are unknowns and x', y', x_*, y_*, h_1-h_8 are parameters.

Because the analytic solutions of (28,29) contain over one thousand monomials, we omitted their complicated expressions here. A maple worksheet is available to download at the page http://cam.tju.edu.cn/%7enan/QH.html for readers to verify the correctness. Actually, if the parameters h_1-h_8 and a pair of x, y or x', y' are given, we plug their values into (21,22) or (23,24) then solve the forward or the backward map directly, without using these analytic solutions. A symbolic proof for the line-preserving property of homography warps and the equivalence of two homography formulations are included in the worksheet as well.

REFERENCES

[1] S. Tzavidas and A. K. Katsaggelos, "A multicamera setup for generating stereo panoramic video," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 880–890, 2005.
[2] X. Sun, J. Foote, D. Kimber, and B. S. Manjunath, "Region of interest extraction and virtual camera control based on panoramic video capturing," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 981–990, 2005.
[3] V. R. Gaddam, M. Riegler, R. Eg, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1819–1831, 2016.
[4] H. Y. Shum, K. T. Ng, and S. C. Chan, "A virtual reality system using the concentric mosaic: construction, rendering, and data compression," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 85–95, 2005.

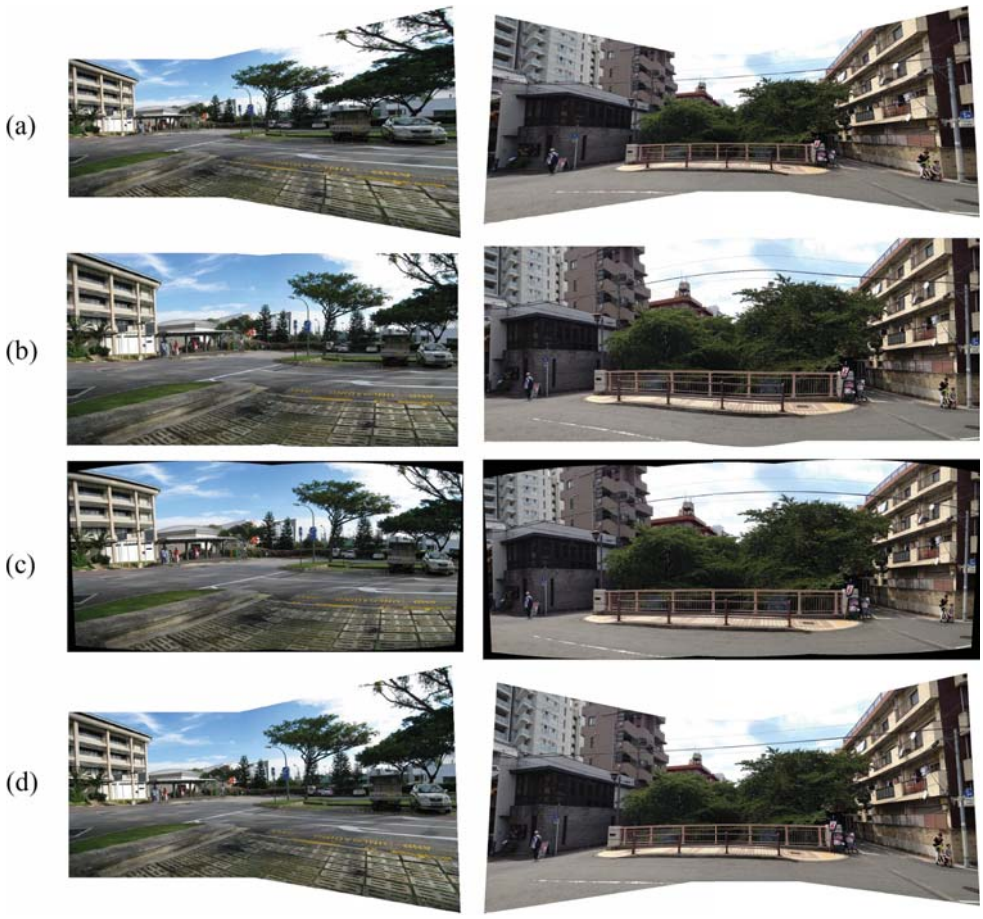

Fig. 8. A naturalness comparison of different warps for stitching two and three images. (a) Homography. (b) SPHP. (c) AutoStitch. (d) Our warp.

Fig. 9. A naturalness comparison of different warps for stitching two sequences of ten and nine images. (a) Homography. (b) AutoStitch. (c) Our warp. SPHP fails to produce a panorama.

[5] W. K. Tang, T. T. Wong, and P. A. Heng, "A system for real-time panorama generation and display in tele-immersive applications," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 280–292, 2005.

[6] Q. Zhao, L. Wan, W. Feng, and J. Zhang, "Cube2Video: Navigate between cubic panoramas in real-time," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1745–1754, 2013.

[7] R. Szeliski, "Image alignment and stitching: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2006.

[8] S. Peleg, "Elimination of seams from photomosaics," *Comput. Graph. Image Process.*, vol. 16, no. 1, pp. 90–94, 1981.

[9] M.-L. Duplaquet, "Building large image mosaics with invisible seam lines," in *Proc. SPIE Visual Information Processing VII*, 1998, pp. 369–377.

[10] J. Davis, "Mosaics of scenes with moving objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, June. 1998, pp. 354–360.

[11] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proc. ACM SIGGRAPH*, 2001, pp. 341–346.

[12] A. Mills and G. Dudek, "Image stitching with dynamic elements," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1593–1602, 2009.

[13] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graphics*, vol. 2, no. 4, pp. 217–236, 1983.

[14] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 313–318, 2003.

[15] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in *Proc. Eur. Conf. Comput. Vis.*, May 2004, pp. 377–389.

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

[17] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 49–56.

[18] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, "Smoothly varying affine stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 345–352.

[19] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 2339–2346.

[20] Z. Lou and T. Gevers, "Image alignment by piecewise planar region matching," *IEEE Transactions on Multimedia*, vol. 16, no. 7, pp. 2052–2061, 2014.

[21] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, "Seam-driven image stitching," *Eurographics*, pp. 45–48, 2013.

[22] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, May 2014, pp. 3262–3269.

[23] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, "SEAGULL: Seam-guided local alignment for parallax-tolerant image stitching," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016.

[24] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, May 2014, pp. 3254–3261.

[25] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 1155–1163.

[26] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 186–201.

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

[28] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[29] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 294–302, 2004.

[30] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 277–286, 2003.

[31] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, Jun. 2006, pp. 2498–2505.

[32] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.

[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[34] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[35] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, 2014.

[36] M. Kushnir and I. Shimshoni, "Epipolar geometry estimation for urban scenes with repetitive structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2381–2395, 2014.

[37] W.-Y. Lin, S. Liu, N. Jiang, M. N. Do, P. Tan, and J. Lu, "RepMatch: Robust feature matching and pose for reconstructing modern cities," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 562–579.

[38] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "USAC: a universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, 2013.

[39] Q.-H. Tran, T.-J. Chin, G. Carneiro, M. S. Brown, and D. Suter, "In defence of RANSAC for outlier rejection in deformable registration," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 274–287.

[40] Y. Yang, Z. Ma, A. G. Hauptmann, and N. Sebe, "Feature selection for multimedia analysis by sharing information among multiple tasks," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 661–669, 2013.

[41] Q. Chen, M. Wang, Z. Huang, Y. Hua, Z. Song, and S. Yan, "VideoPuzzle: Descriptive one-shot video composition," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 521–534, Apr. 2013.

[42] Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, and D. Xu, "Image classification by cross-media active learning with privileged information," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2494–2502, Dec. 2016.

[43] Y. Li, P. Li, D. Lei, Y. Shi, and L. Tan, "Investigating image stitching for action recognition," *Multimedia Tools and Applications*, Aug. 2017. [Online]. Available: https://doi.org/10.1007/s11042-017-5072-4

Nan Li received the Ph.D. degree in applied mathematics from University of Chinese Academy of Science, Beijing, China, in 2013. He is currently an Assistant Professor with the Center of Applied Mathematics, Tianjin University, Tianjin, China. His research interests include computer vision, machine learning and computational algebra geometry.

Yifang Xu received the B.Sc. degree in mathematics from Sichuan University, Chengdu, China, in 2013. Since 2013, she has been with Center for Combinatorics, Nankai University, Tianjin, China, where she is currently working toward the Ph.D. degree in applied mathematics. Her research interests include computer vision, object detection and combinatorics.

Chao Wang received the M.Sc. degree and the Ph.D. degree from Nankai Uiversity, Tianjin, China, in 1997 and 2003, respectively. He had been a postdoctoral in Tsinghua University from 2003 to 2005 before he entered College of Software, Nankai University. He is currently an associate professor. His research interests include computer vision, theoretical computer science and combinatorics. He has served as a reviewer for several journals. He has been an ACM member since 2008.