A NEW METHOD TO FIND AN OPTIMAL WARPING FUNCTION IN IMAGE STITCHING

Hyung Il Koo, Beom Su Kim, and Nam Ik Cho

School of Electrical Engineering
Seoul National University, Seoul 151-744 Korea
email: hikoo@ispl.snu.ac.kr, bskim@ispl.snu.ac.kr, nicho@snu.ac.kr

ABSTRACT

In image stitching applications, it is very important to find a suitable warping function for the visual quality of a composite (stitching result). In this paper, we present a new mathematical criterion to select an optimal warping function among a set of possible candidates (e.g., parametric family). The proposed criterion can be considered as *a direct view condition* for image stitching , i.e., it is desirable that each part of the composite image looks like its corresponding input image. More specifically, we do not use an explicit modeling of a compositing surface, but, we focus on the differential properties of a warping function. That is we design a cost function so that the Jacobian matrix of a warping function is close to a shape preserving matrix such as rotation and reflection matrices. The proposed cost function can be effectively minimized by using Levenberg-Marquardt algorithm. The experimental results show that the proposed method results in visually pleasing stitched results because the original shape of each image is preserved in the composite.

Index Terms— Image Stitching, Panorama

1. INTRODUCTION

Image stitching is a process to generate a larger and high-resolution composite from multiple images. Recently, it is gaining much attention because it can overcome the limited resolution and the small field-of-view of hand-held cameras and it can also create beautiful high-resolution panoramas [1, 2, 4, 6]. Among several issues to create a visually pleasing panorama such as image registration, blending and so on, we address the issue on how to get a warping function (or in previous literatures, how to select a type of surface and the placement of the surface around the viewing sphere) that transforms a source image to a part of a composite. Finding an appropriate warping function is an important problem because a geometrically and photometrically correct panorama can be perceptually incorrect as shown in the examples of Fig. 2-(b) and Fig. 3-(b).

Panoramas can be constructed if there is no (or little) parallax between images, which occurs when capturing scenes with a fixed optical center [1, 2, 6] or capturing the plane scene [10]. In the case of rotating a camera about its optical center, pictures taken by the camera cover a viewing sphere. Hence the problem to this case is how to project the sphere into flat image with less distortion. A typical approach to this problem is to assume a surface around the viewing sphere and to project the viewing sphere onto the surface by imagining rays from the center of the sphere to the surface. In this step, the surface is usually flat or cylindrical. For example, when only a few images are stitched, one of them is selected as a reference plane and other images are transformed to the reference plane [1, 4]. In *Autostitch* which is a fully-automatic image stitching system [2, 11], a cylindrical surface is used. However, even if we select an appropriate surface (plane, cylinder or cube), there remains a problem about where the surface should be placed. For example, we should determine which image is used for the reference in linear perspective panorama. Also when the cylindrical surface is used, the up-vector of cylinder should be determined. Although there is some heuristic approach to choose the up-vector in [3], there is no universal rule to select a surface and its placement [7]. Hence the most automated approach uses a fixed surface and employs heuristics to place the surface around the viewing sphere. A similar problem arises when capturing a plane scene, which is the second case that image stitching is possible. In this case a plane is a natural choice for a composite. However, which plane results in the most natural composite is still a problem, because images are captured obliquely (it requires much user's effort to capture the scene perfectly perpendicular) to the scene plane in practice, and thus using one of them as a reference plane results in the accumulated distortions as the number of images increase (see Fig. 3-(b)).

It is already noted that the right choice of a surface and its placement depend on the input images: the surface should be selected based on the tradeoff between (1) keeping the local appearance undistorted (a straight line should be straight) and (2) providing a reasonably uniform sampling [4]. Interestingly, similar conditions are already introduced in computer graphics literature (image synthesis) [9]. Because a geometrically correct image is not always perceptually pleasing, the authors of [9] proposed a correcting method of (perceptually) distorted images. For this, they proposed two conditions: *zero-curvature condition* and *direct view condition*, where the former is measured by computing the maximum curvature which can be introduced by warping a straight line and the latter means that it is desirable that each object looks like as if it is at the center of an image (or equivalently, circles should be preserved rather than be warped to be an ellipse).

In order to alleviate the above stated problem, we propose a new approach based on functional minimization. In other words, we do not imagine the surface and its placement, but, we focus on a warping function that results in a natural composite. For this, it is believed that the *direct view condition* which is similar to relatively uniform sampling can be a good criterion. Hence we propose a kind of *direct view condition* which can handle multiple images and automate the selection of warping function. Specifically, the proposed condition can be stated as follows: it is desirable that each part of the composite image looks like its original input image. Also, in order to encode this condition into a sound mathematical formulation, we consider differential properties of a warping function. By considering Jacobian matrix $J \in \Re^{2\times 2}$ of a warping function, we can handle the local distortions quantitatively. The formulation is based on the properties of singular values of J and it is expressed by the sum of a small number of square functions, which enables us to use an effective optimization method based on Levenberg-Marquardt algorithm [1].

The proposed method have been applied to two cases in this paper: a panorama obtained by rotating a camera and a panorama capturing a plane scene. The experimental results show that the proposed method generates a visually better composite because the original shape of each image is preserved in the composite.

2. PROPOSED ALGORITHM

In this section, we will explain the proposed cost function which measures distortions caused by image stitching. For this, we will firstly define the cost function for a single warped image, and then we approximate the cost function to a more mathematically tractable one. Finally we define a cost function which results in a perceptually correct panorama.

2.1. Distortions introduced by Warping

We assume that an image I is transformed by a warping function $T : \Re^2 \to \Re^2$. If we denote $T(x,y) = (\phi(x,y), \psi(x,y))$, the Jacobian $J_T(x,y)$ of the warping function is given by

$$J_T(x,y) = \begin{pmatrix} \frac{\partial}{\partial x}\phi(x,y) & \frac{\partial}{\partial y}\phi(x,y) \\ \frac{\partial}{\partial x}\psi(x,y) & \frac{\partial}{\partial y}\psi(x,y) \end{pmatrix}. \tag{1}$$

Because the Jacobian is an approximation of the warping function to a linear transform (affine transform), we can have the information on the local distortions such as scale change, aspect ratio change and skew. Among many tools for the matrix analysis, SVD (singular value decomposition) of the matrix gives an easy way to estimate the perceptual distortions [8]. The SVD decomposes $J_T(x,y)$ into

$$J_T(x,y) = U(x,y)\,\Sigma(x,y)\,V(x,y)^T \tag{2}$$

where $U(x,y)$ and $V(x,y) \in \Re^{2\times 2}$ are rotations and reflections and $\Sigma(x,y)$ is given by

$$\Sigma(x,y) = \begin{pmatrix} \sigma_T^1(x,y) & 0 \\ 0 & \sigma_T^2(x,y) \end{pmatrix}. \tag{3}$$

Because rotations and reflections do not introduce perceptual distortions, we can see the amount of local distortions is a function of two singular values $\sigma_T^1(x,y)$ and $\sigma_T^2(x,y)$. Actually, $\Sigma(x,y)$ is called a stretch matrix and $\sigma_T^1(x,y)$ and $\sigma_T^2(x,y)$ mean maximum and minimum local stretch respectively:

$$\sigma_T^1(x,y) = \max \frac{|J_T(x,y)\,v|}{|v|} \tag{4}$$

$$\sigma_T^2(x,y) = \min \frac{|J_T(x,y)\,v|}{|v|}. \tag{5}$$

Because the transformed image looks like its source image (also in size) only if $\sigma_T^1(x,y) \simeq 1$ and $\sigma_T^2(x,y) \simeq 1$, we design the cost function so that $\sigma_T^1(x,y)$ and $\sigma_T^2(x,y)$ are close to unity. In other words, the distortion cost $E(T;I)$ which is introduced by transforming I with T is defined as a maximum local distortions:

$$E(T;I) = \max_{(x,y)\in D(I)\ i=1,2} f(\sigma_T^i(x,y)) \tag{6}$$

where $D(I)$ is the domain on which I is defined and $f(\sigma)$ is defined so that same amount of cost is imposed to stretch and contraction:

$$f(\sigma) = \max\left(|\sigma - 1|, \left|\frac{1}{\sigma} - 1\right|\right). \tag{7}$$

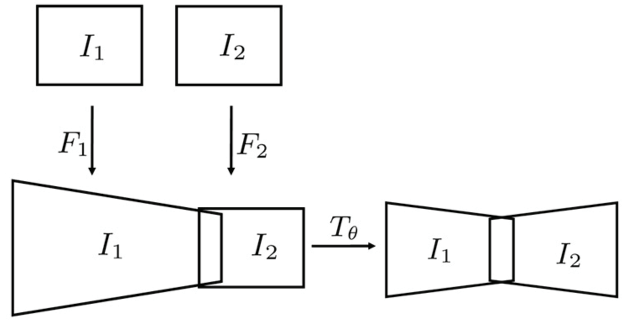

Fig. 1. Illustration of the notations of the proposed algorithm. I_1 and I_2 denote input images. They are transformed by F_1 and F_2 and aligned in some intermediate surface. In that surface, some image may seem too stretched or contracted. In order to correct them, we find an optimal mapping T_θ that each transformed image looks like the input picture.

Although the cost function is intuitive and well-defined (it minimizes a maximum singular value derivation), we approximate the cost function to more computationally efficient one. First, we substitute a set of a few control points $C(I)$ for the domain of an image I. Second we approximate l^∞-norm to l^2-norm due to the simplicity of minimization and the satisfactory performance of l^2-norm. Such approximations result in a simplified cost $E_{sim}(T;I)$:

$$E_{sim}(T;I) = \frac{1}{2|C(I)|} \sum_{(x,y)\in C(I)} f_{sim}(\sigma_T^1(x,y)) + f_{sim}(\sigma_T^2(x,y)) \tag{8}$$

where $|\cdot|$ is the cardinality of a set and $f_{sim}(\sigma)$ is defined to measure the deviation from one, i.e.,

$$f_{sim}(\sigma) = (\sigma - 1)^2 + \left(\frac{1}{\sigma} - 1\right)^2. \tag{9}$$

2.2. Multiple Images Case

For applying the proposed distortion measure to multiple image stitching problem, let us assume that n images $(I_1, I_2, \cdots, I_n)$ are given and they are registered (stitched) on some intermediate surface $\mathcal{S}$. It can be done by using conventional image registration methods such as [1, 2, 4]. Specifically, we have used the method in [2] for the construction of a single viewpoint panorama such as Fig. 2 and the method in [1] has been used for the registration of planar scene such as Fig. 3. We denote a warping function that transforms I_i to an intermediate surface as $F_i : \Re^2 \to \Re^2$. After transforming each image by F_i, we can get a geometrically correct composite on $\mathcal{S}$. However, the perceptual distortions of image on $\mathcal{S}$ depends on the choice of $\mathcal{S}$ and the composite usually suffers from distortions. In order to correct such distortions, the image on $\mathcal{S}$ should be warped by some warping function. We denote a parametric family of warping functions as $T_\theta : \Re^2 \to \Re^2$ where θ is a set of parameters that controls the transform. Then, we can get the warping function that transforms a domain of I_i to the final composite plane by concatenating T_θ and F_i. This procedure is illustrated in Fig. 1.

We assume that the amount of overall distortions caused by image stitching is the sum of distortions caused by transforming each

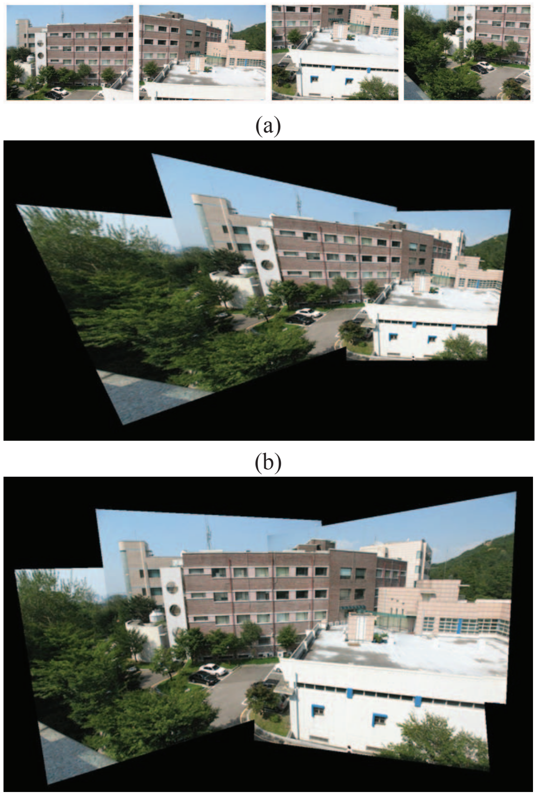
(a)
(b)
(c)

Fig. 2. Image stitching result on images which capture a scene by rotating a camera (a) Four inputs of a single viewpoint panorama (b) Image stitching result using one of inputs as a reference (c) Image stitching result using the proposed method

image by $T_\theta \circ F_i$. Therefore, we define a cost function $E(\theta)$ and find the optimal parameter $\hat{\theta}$ by minimizing the function, i.e.,

$$\hat{\theta} = \arg\min E(\theta) \tag{10}$$

where $E(\theta)$ is defined as

$$E(\theta) = \sum_{i=1}^{n} E_{sim}(T_\theta \circ F_i; I_i). \tag{11}$$

$E(\theta)$ consists of $2 \times \sum_{i=1}^{n} |C(I_i)|$ square functions and the optimization can be done using Levenberg-Marquardt algorithm [1]. Finally, by warping each image I_i $(i = 1, 2, \cdots, n)$ using $T_{\hat{\theta}} \circ F_i$, we can get a composite.

3. EXPERIMENTAL RESULTS

Although the proposed *direction view condition* can be applied to any kind of a differentiable warping function $T_\theta \circ F_i$, it is another problem to achieve a perceptual balance between the *zero-curvature condition* and the proposed *direct view condition* [9]. Even if we have a quantitative measure of *zero-curvature condition*, (for example, a maximum curvature of a warped straight line can be an answer), how to combine semantically different two cost function still remains as a problem. Hence we have applied the proposed method to simple cases where *zero curvature condition* is perfectly satisfied. In other words, we consider the case that T_θ and F_i are 2-D planar homography. Such a situation occurs in two cases: a linear perspective panorama and stitching images which capture a planar scene.

3.1. Implementation Details

After capturing the scene, we use the software in [15] to remove radial distortions. We denote radial distortion free images as $I_1, I_2, \cdots, I_n$. Then we compute the pairwise homography between images. For this a SIFT descriptor and RANSAC (RANdom SAmple Consensus) algorithm are used. Then the geometrical error minimization using Levenberg-Marquardt algorithm is followed [1, 14, 13]. In selecting an intermediate plane $\mathcal{S}$, we use the first image as a reference. Therefore, F_1 is an identity and F_i is defined as a homography that transforms I_i to the plane of I_1. The family of T_θ is defined as

$$T_\theta(x, y) = \left(\frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + h_9}, \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + h_9} \right) \tag{12}$$

where $\theta = (h_1, h_2, \cdots, h_9) \in \Re^9$.

In experiments, we use a set of four corner points of I as $C(I)$ so that $E(\theta)$ consists of $8 \times n$ square functions. Levenberg-Marquardt algorithm implementation in [12] is used with a closed form solution of singular values, i.e.,

$$\sigma_T^1(x, y) = \sqrt{\frac{1}{2}\left((E+G) + \sqrt{(E-G)^2 + 4F^2}\right)} \tag{13}$$

$$\sigma_T^2(x, y) = \sqrt{\frac{1}{2}\left((E+G) - \sqrt{(E-G)^2 + 4F^2}\right)} \tag{14}$$

where $E = \phi_x^2 + \psi_x^2$, $F = \phi_x \phi_y + \psi_x \psi_y$ and $G = \phi_y^2 + \psi_y^2$ [9].

3.2. Experimental Results

Experimental results on the images taken by rotating a camera about its optical center is shown in Fig. 2. As can be seen in Fig. 2-(a),(b), none of four inputs can be a good reference plane (if any, it requires user interaction). The proposed method finds an optimal view of the scene automatically and results in Fig. 2-(c). Another example is shown in Fig. 3. It is a result on images capturing a planar scene. If there is an image which is taken perfectly perpendicular to the scene, it is a good choice to use the image as a reference, although it requires much user attention in taking the pictures and also the user interaction in image stitching. However, the proposed method results in a visually pleasing result with less attention in taking pictures (it is sufficient to capture the plane scene approximately perpendicularly) and no user interaction is required in image stitching.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we consider the problem of image stitching as a problem to find a good warping function. In order to find an optimal function for this purpose, we have presented a new *direct view condition*, which is derived from the observation that the best view of the object in a composite is to view the object as if they appear in a source image. Precisely, the condition is formulated as a minimization problem which consists of the sum of small number of square functions, which can be effectively minimized using Levenberg-Marquardt al-

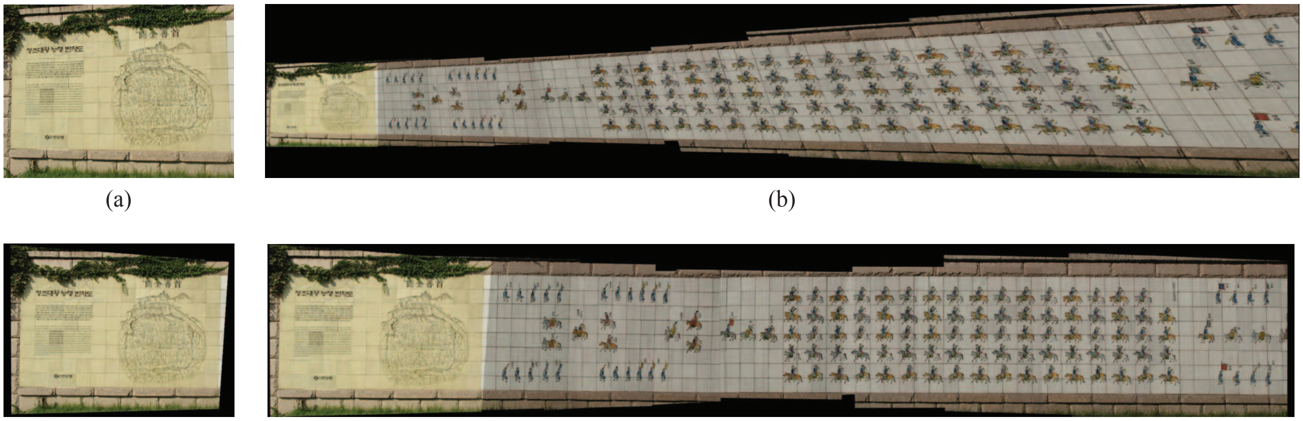

Fig. 3. Stitching result of images which capture a planar scene (a) Leftmost image of inputs (it is slightly oblique to the scene). (b) Image stitching result using the leftmost image as a reference (c) Automatically warped leftmost image so that all other images can be seen less distorted (d) Image stitching result using the proposed method.

gorithm. The experimental results show that the proposed criterion provides a better view.

However the warping function T_θ in this paper is quite limited (2D homography), whereas the family of T_θ can be much wider class. We think a more natural composite can be obtained in more complex cases by combining *the zero curvature condition* with the proposed condition.

5. REFERENCES

[1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* Cambridge University Press, 2000

[2] M. Brown and D. Lowe, "Recognising panoramas," *in Proc. ICCV ,* vol. 2, pp. 1218-1225, Oct. 2003.

[3] M. Brown and D. Lowe, "Automatic Panoramic Image Stitching using Invariant Features," *International Journal of Computer Vision,* 74(1), pp. 59-73, 2007

[4] R. Szeliski, "Image alignment and stitching: A Tutorial," Technical Rerpot MSR-TR-2004-92, Microsoft Research, Dec. 2004.

[5] R. Szeliski and H. Shum, "Creating full view panoramic image mosaics and texture-mapped modles," *in Proc. SIGGRAPH ,* pp. 251-258, 1997.

[6] M. Brown, R. I. Hartley, and D. Nister, "Minimal Solutions for Panoramic Stitching," *in Proc. CVPR ,* 2007.

[7] L. Zelnik-Manor, G. Peters, and P. Perona, "Squaring the circle in panoramas," *in Proc. ICCV,* vol. 2, pp. 17-21, Oct. 2005.

[8] G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press 1998

[9] D. Zorin, A. H. Barr, "Correction of Geometric Perceptual Distortions in Pcitures," *in Proc. SIGGRAPH ,* 1995.

[10] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski, "Photographing Long Scenes with Multi-Viewpoint Panoramas," *in Proc. SIGGRAPH ,* 2006.

[11] Autostitch. http://www.autostitch.net/.

[12] Levenberg-Marquardt nonlinear least squares algorithms in C/C++. http://www.ics.forth.gr/ lourakis/levmar/

[13] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Communications of the ACM,* 24:381-395, 1981.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal on Computer Vision*, pp. 91-110, 2004.

[15] http://epaperpress.com/ptlens/.