Overlap-Area Registration Model for View Mosaics Using Matching Features

Xiaoyong Sun, *Student Member, IEEE* and Eric Dubois, *Fellow, IEEE*

Abstract

In this paper, we provide a new optimization model for registration of the overlap-area between two adjacent images taken by a camera mounted and rotated on a tripod. After overlap-area registration, two such views can be stitched into one picture with a larger field of view. The proposed model includes affine transformation adjustments and nonlinear focal-length adjustment. In order to minimize the stitching errors in the final view, we select a vertical stripe area for registration and stitching instead of using the whole overlap area. An implementation based on feature matching has been proposed to significantly reduce the computations. All transformations involved are represented as sampling structure changes, and the actual interpolation is carried out only when necessary in order to maintain the resolution as high as possible. In addition, a novel algorithm is proposed that can significantly reduce the accumulated errors when stitching multiple images to generate 360° panoramas. The solution of the global optimization for registration of the overlap areas between all the pre-captured images in generating such panoramas has been largely simplified with an iterative approach. The experimental results demonstrate that the proposed methods can yield good stitched views with minimal artifacts.

Index Terms

X. Sun and E. Dubois are both with the School of Information Technology and Engineering (SITE), University of Ottawa, Canada.

View Mosaic, Image-Based Rendering, Feature Point Matching, Image Registration, Panoramas.

I. INTRODUCTION

Recently, pictures with large field of view (FOV), such as panoramic views, have found a number of applications in image-based rendering (IBR) [1]. The FOV of an image is limited by the camera's optical system which is difficult and expensive to increase through strictly optical means. However, stitching several small-FOV images into one with a large FOV is more straightforward and easier to achieve through software implementation [2]. Stitching overlapped images along the horizontal direction to get wide-FOV pictures and even 360° panoramas [3] has become very common recently. In this situation, views from different directions are captured by a camera that is mounted on a tripod and rotated around its optical center. The adjacent images, with some overlap area, are then stitched together. More general scenarios for view mosaics can be found in [4].

Overlap-area registration is the most important step that must be accomplished before stitching. When depth variations are small compared with the average depth, such as in most outdoor-scene applications, the registration of adjacent images is not a big problem and most of the available commercial software, (e.g., [5]), can perform well. However, when the depth variation is large compared to the average depth, as in most indoor-environment applications, this depth variation will cause significant problems for overlap area registration between the adjacent views.

Previous work on overlap-area registration emphasizes recovery of the camera motion based on different models. Under this framework, various methods to register the overlap area of two adjacent images have been proposed based on an 8-parameter perspective transformation [6], a polynomial transformation with more freedom [7], and other geometric corrections [8]. A comparison of some commonly-used parametric models for camera motion recovery can be found in [9]. The registration and the stitching are usually implemented based on the whole overlap area, which can be problematic because the overlap area may

be too large for one global transformation to yield a good result. Thus local corrections have to be used for ghosting cancellation [6], which makes the registration more complex and introduces the possibility of causing further discontinuities among local blocks and even distortions, especially when dealing with high resolution pictures. Thus, the problem of effective overlap-area registration is still open.

The method proposed in this paper does not explicitly recover camera motion, but aims to reduce the registration error in the overlap area and to obtain a high-quality seamless mosaic view after stitching. Our model is simple but nonlinear, jointly adjusting the camera's pose and the camera's focal length parameter. The focal length is used in the warping procedure before stitching. Warping the overlapped images onto a common imaging surface is a popular technique in view mosaicking. This simulates the condition that the mosaic view was captured by a virtual camera with its imaging sensor in the shape of this common surface. In this way, the difference between the two images in the overlap area can be reduced; we refer to this procedure as warping for a virtual common imaging surface. The camera pose adjustment is carried out in this paper using affine transformations, but other transformations such as perspective transformations can equally well be used.

We observe that the overlap area registration and stitching of two adjacent views do not necessarily need to be carried out in the whole overlap area, but can be limited to a stripe-block area within it. This observation makes the registration easier than previous methods and it significantly reduces the discontinuities [10].

Like most other registration methods, the proposed registration method can use the image intensities at each pixel within the overlap area for registration. However, the amount of computation in the optimization is very large in this case, thus affecting the efficiency of the algorithms. The time-consuming optimizations usually end up locating a local minimum and may not give reliable results when the initial guess is not good enough [2]. An alternate implementation using registration based on feature matching is given in this paper, where feature points have been used as control points, as in [11]. After obtaining a set of reliable

matching features between two adjacent images, the overlap-area registration problem is converted to the position registration of the feature points. The optimization problem is consequently changed to using the positions of a limited number of feature points instead of using the image intensities or textures over the whole overlap area. Thus algebraic solutions for the optimization problem can be applied, significantly reducing the amount of computation. A great deal of research has been carried out in computer vision for fast and robust feature matching algorithms [12] to deal with various matching difficulties under different scenarios. Similar work on image mosaicking by stratified matching has recently appeared [13] but in a scenario different from this paper.

One application of view mosaics is to generate cylindrical panoramas, where a set of overlapped images that cover a 360° field of view are captured by a camera mounted and rotated on a tripod. Starting from one given image (the first image), the adjacent images are registered in the overlap area and stitched one by one until the last image has been stitched. The left side and right side of this mosaicked image finally have to be registered and stitched in the overlap area. However, the accumulated registration errors are usually very large. Thus, a global optimization on the overlap areas between all pre-captured images is required, although the solution might be too complex to obtain directly. In this paper, an iterative algorithm is proposed to obtain the global optimization solution.

Usually, the sophisticated algorithms for overlap registration that are required to achieve good results involve many transformations. Each transformation re-samples the original images and thus reduces their resolution. In this paper, all the transformations involved are regarded as sampling structure changes, and the actual interpolation is carried out only when necessary, at the final stage. In this way, the maximum resolution can be maintained.

In section II, we briefly describe the camera rotation model in obtaining the pre-captured images, the basic techniques that will be used in this paper and the methods to match features in the overlap area. The optimization models and our solutions to reduce the registration errors between two adjacent

overlapped images are proposed in section III. To minimize the discontinuities, the strategy of registration and stitching on a stripe-block area within the overlap area is introduced in section IV. Section V shows how multiple transformations, represented as changes of the image's sampling structures, can be cascaded. The amount of computation for the matching feature-based approach and the texture-based approach are also roughly compared in this section. The procedure to generate 360° panoramic views based on the proposed method are described in section VI with a novel iterative algorithm. The experimental results are illustrated in section VII, with our conclusions given in section VIII.

II. Camera movement model and basic techniques

In order to stitch two adjacent images, the difference between the two views in the overlap area must be minimized to obtain stitched images with good quality. From a general camera rotation model [8], we find that camera pose and movement deviating from the ideal case is the main cause of registration error [10]. Affine adjustment and focal-length adjustment are used in this paper to reduce the registration error in the overlap area caused by such non-ideal camera pose and movement. The affine transformation on a discrete image is defined first. Then, cylindrical warping is discussed in order to introduce the nonlinear part of the proposed model. In addition, the precise definition of the overlap area is given, followed by a brief discussion of feature detection and matching in the overlap area.

A. A general camera rotation model for analysis of registration errors

When a camera is mounted on a tripod and captures images at different rotation angles, its motion usually deviates from a pure rotation about the camera projection center. We refer to the camera's projection center as the camera position in the following sections. A general camera rotation model [8] is illustrated in Fig. 1. It is not easy to physically locate the projection center of a real camera. Thus the camera center usually moves along a circle (the circle may not even be on the level plane), as shown by the dashed arc in Fig. 1, rather than remaining fixed at a given point, such as O in Fig. 1.

Assume that two images I_1 and I_2 are captured from adjacent viewpoints (O_1 and O_2 in Fig. 1) with

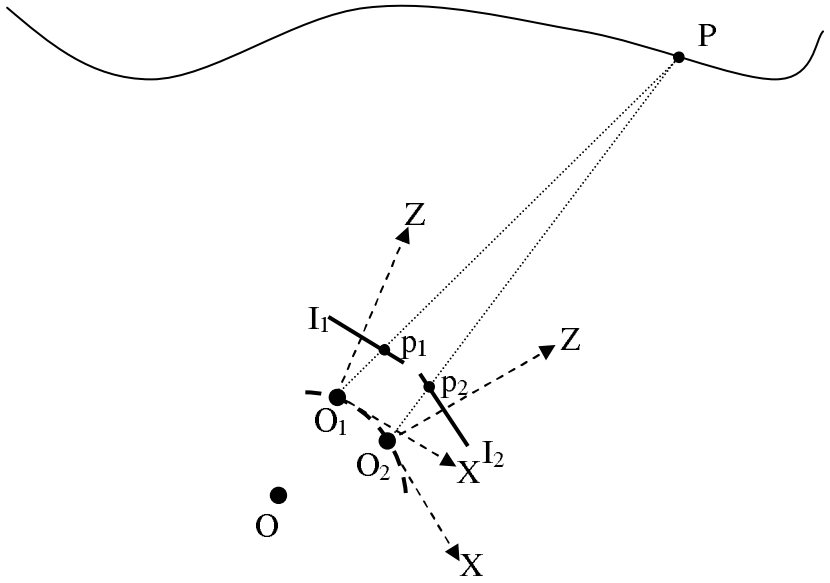

Fig. 1. The general camera rotation model

some overlap area. Using similar notation to [14], we assume that image I_1 is captured with a camera having a certain frame F_1 with origin at the camera center O_1 and I_2 is captured with a camera having frame F_2 with origin O_2. The coordinate vector of a scene point P in frame F_i $(i = 1, 2)$ is denoted ${}^{F_i}P = [{}^{F_i}P_x \, {}^{F_i}P_y \, {}^{F_i}P_z]^T$. Then ${}^{F_2}P = {}^{F_2}_{F_1}R \cdot {}^{F_1}P + {}^{F_2}O_1$, where ${}^{F_2}_{F_1}R$ is a rotation matrix and ${}^{F_2}O_1$ is the coordinate vector of O_1 in frame F_2 [14]. Using the pinhole camera model, the point P is projected to points p_1 and p_2 in F_1 and F_2 respectively, given by $p_i = f_i \cdot {}^{F_i}P / {}^{F_i}P_z$, where f_1 and f_2 are the focal lengths associated with the two camera positions. Combining these equations, we obtain

$$w_{12} \begin{bmatrix} p_{2x} \\ p_{2y} \\ f_2 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} + {}^{F_2}O_{1x} / {}^{F_1}P_z \\ r_{21} & r_{22} & r_{23} + {}^{F_2}O_{1y} / {}^{F_1}P_z \\ r_{31} & r_{32} & r_{33} + {}^{F_2}O_{1z} / {}^{F_1}P_z \end{bmatrix} \begin{bmatrix} p_{1x} \\ p_{1y} \\ f_1 \end{bmatrix} \tag{1}$$

where, $w_{12} = f_1 \cdot {}^{F_2}P_z / (f_2 \cdot {}^{F_1}P_z)$, ${}^{F_2}O_1 = [{}^{F_2}O_{1x} \quad {}^{F_2}O_{1y} \quad {}^{F_2}O_{1z}]^T$ and $p_i = [p_{ix} \quad p_{iy} \quad p_{iz}]^T$ $(i = 1, 2)$. r_{ij} $(i, j = 1, 2, 3)$ are the elements of 3 by 3 matrix ${}^{F_2}_{F_1}R$. The key observation here is that the transformation

to describe the positions of corresponding points in the overlap area of the two adjacent images is depth-dependent. Perfect registration can only be achieved with perfect 3D reconstruction of the depth distribution of the environment, which is difficult to implement in practice.

B. Discrete affine transformation

Affine transformations are normally defined on continuous-space images. For a discrete-space image I, the affine transformation is defined as follows. Assume that a continuous-space image corresponding to I can be obtained using a linear interpolation operator $\mathcal{H}$. Then the discrete affine transformation operator with parameter vector $\boldsymbol{t}$, denoted $\mathcal{A}_{\boldsymbol{t}}$, is defined by $(\mathcal{A}_{\boldsymbol{t}}I)(\boldsymbol{x}) = (\mathcal{H}I)(\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{x}_0)+\boldsymbol{d}+\boldsymbol{x}_0), \boldsymbol{x}\in\Lambda$, where $\boldsymbol{T} = \left[\begin{smallmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{smallmatrix}\right]$ and $\boldsymbol{d} = [d_1\ d_2]^T$. We define the parameter vector $\boldsymbol{t} = [t_{11}\ t_{12}\ t_{21}\ t_{22}\ d_1\ d_2]^T$ and $\boldsymbol{x}_0$ is an arbitrary reference point. In this paper, the image center $\boldsymbol{x}_c = (x_c, y_c)$ will be selected as the reference point, so we will not explicitly specify $\boldsymbol{x}_0$ in the following sections unless necessary. Λ is the sampling lattice where the image $(\mathcal{A}_{\boldsymbol{t}}I)(\boldsymbol{x})$ is defined. Of course, $(\mathcal{H}I)(\boldsymbol{x}')$ is only computed at the points $\boldsymbol{x}' = \boldsymbol{T}(\boldsymbol{x}-\boldsymbol{x}_c)+\boldsymbol{d}+\boldsymbol{x}_c, \boldsymbol{x}\in\Lambda$, using any suitable interpolation method such as bilinear, bicubic, or spline. The affine transformation of coordinates on $\mathbb{R}^2$ can also be represented as an operator $\mathcal{S}_{\boldsymbol{t}}$ where $\mathcal{S}_{\boldsymbol{t}}\boldsymbol{x} = \boldsymbol{T}(\boldsymbol{x}-\boldsymbol{x}_c)+\boldsymbol{d}+\boldsymbol{x}_c$. Thus, we have $(\mathcal{A}_{\boldsymbol{t}}I)(\boldsymbol{x}) = (\mathcal{H}I)(\mathcal{S}_{\boldsymbol{t}}\boldsymbol{x})$. In practical situations, $\boldsymbol{T}$ is non-singular, so that $\mathcal{S}_{\boldsymbol{t}}$ is invertible. The inverse is also an affine transformation, with $\mathcal{S}_{\boldsymbol{t}}^{-1}\boldsymbol{x}' = \boldsymbol{T}^{-1}(\boldsymbol{x}'-\boldsymbol{x}_c)-\boldsymbol{T}^{-1}\boldsymbol{d}+\boldsymbol{x}_c$.

C. Warping the image onto a cylindrical surface

In order to reduce the difference between two adjacent images in the overlap area, the images should be mapped onto a common surface such as a cylindrical or spherical surface, as if they were captured by a camera with its imaging sensor on this common surface. We choose a cylindrical surface as the common imaging surface in this paper since it is a popular choice to generate 360° panoramas. The warped images are referred to as *cylindrically-warped* images. The cylindrical radius is usually chosen to be the focal

length of the camera. This choice makes the geometric relationships between the source images and the warped images straightforward and thus the warping algorithm is easy to implement. Using similar notation as above, the cylindrical warping is denoted as $\mathcal{W}_f$ with parameter f, which is the focal length of the camera (assuming that the camera is rotated exactly around its optical center and the focal length is unchanged for adjacent images). Then, the warped image is $(\mathcal{W}_f I)(\boldsymbol{x}) = (\mathcal{H}I)(\boldsymbol{x}^{'}), \boldsymbol{x} \in \Gamma$, where Γ is a lattice in the cylindrical coordinate. The relationship between $\boldsymbol{x} = [x\ y]^T$ and $\boldsymbol{x}^{'} = [x^{'}\ y^{'}]^T$ can be represented as (obtained from [2]) $x^{'} = f \cdot \tan((x - x_c)/f) + x_c$ and $y^{'} = (y - y_c)/\cos((x - x_c)/f) + y_c$.

This one-to-one nonlinear mapping between $\boldsymbol{x}$ and $\boldsymbol{x}^{'}$ can also be represented by an operator $\mathcal{U}_f$ on $\mathbb{R}^2$, so that $(\mathcal{W}_f I)(\boldsymbol{x}) = (\mathcal{H}I)(\mathcal{U}_f \boldsymbol{x})$ with $\boldsymbol{x}^{'} = \mathcal{U}_f \boldsymbol{x}$. The corresponding inverse mapping defines the de-warping procedure that will be used later. The inverse of the above relationship is given by $x = f \cdot \tan^{-1}((x^{'} - x_c)/f) + x_c$ and $y = f \cdot (y^{'} - y_c)/(\sqrt{(x^{'} - x_c)^2 + f^2}) + y_c$, or $\boldsymbol{x} = \mathcal{U}_f^{-1} \boldsymbol{x}^{'}$. Operator $\mathcal{U}_f^{-1}$ denotes the sampling structure change from the warped image to the de-warped image in the de-warping transformation, $I(\boldsymbol{x}) = (\mathcal{H}(\mathcal{W}_f I))(\mathcal{U}_f^{-1} \boldsymbol{x})$.

One fixed focal length can only be used to warp all images if the camera was rotated exactly around its projection center and the camera focal length was unchanged during the capture procedure, because this focal length is the radius of the cylindrical surface. However, the camera usually does not rotate around its projection center, as illustrated in Fig. 1. The optimal virtual common imaging surface for view mosaicking is no longer a cylindrical surface with one fixed radius, or focal length. Thus, the focal length will be assumed variable and can be adjusted for each image. In practice, the real shape of the virtual common imaging surface may be too difficult to recover. Adjustment of focal length alone may be insufficient to recover the practical virtual common imaging surface. In section III-B, we will define the optimal focal length with respect to minimization of the registration errors in the overlap area.

D. Definition of the overlap area between adjacent images

Assume that I_1 and I_2 are two original adjacent images where I_1 is to the left of I_2. I_1 and I_2 are defined on sampling lattice Λ. Let W_1 define the area where image I_1 is defined and set $I_1(\boldsymbol{x}) = 0, \boldsymbol{x} \in \mathbb{R}^2 \setminus W_1$. Similarly, let W_2 define the area where image I_2 is defined with $I_2(\boldsymbol{x}) = 0, \boldsymbol{x} \in \mathbb{R}^2 \setminus W_2$. In both cases, $\boldsymbol{x}$ is in the image coordinate system, with origin at the top left of the image. The approximate registration of image I_1 and I_2 through a simple shifting technique is given by

$$\hat{\boldsymbol{d}}_0 = \arg\min_{\boldsymbol{d}_0 \in \Lambda} \frac{1}{|W_{12}(\boldsymbol{d}_0) \bigcap \Lambda|} \sum_{\boldsymbol{x} \in W_{12}(\boldsymbol{d}_0) \bigcap \Lambda} |I_1(\boldsymbol{x}) - I_2(\boldsymbol{x} - \boldsymbol{d}_0)|, \tag{2}$$

where $W_{12}(\boldsymbol{d}_0) = W_1 \bigcap (W_2 + \boldsymbol{d}_0)$ is the overlap between two regions W_1 and $(W_2 + \boldsymbol{d}_0)$. $|W_{12}(\boldsymbol{d}_0) \bigcap \Lambda|$ represents the number of pixels within area $W_{12}(\boldsymbol{d}_0)$. The overlap area between images I_1 and I_2 is defined as $W_{12}(\hat{\boldsymbol{d}}_0) = W_1 \bigcap (W_2 + \hat{\boldsymbol{d}}_0)$. The sub-image of I_1 in the overlap area is $I_{1\text{com}}$, where $I_{1\text{com}}(\boldsymbol{x}) = I_1(\boldsymbol{x}), x \in W_{12,\Lambda}(\hat{\boldsymbol{d}}_0)$ and $W_{12,\Lambda}(\hat{\boldsymbol{d}}_0) = W_{12}(\hat{\boldsymbol{d}}_0) \bigcap \Lambda$. Similarly, $I_{2\text{com}}(\boldsymbol{x}) = I_2(\boldsymbol{x} - \hat{\boldsymbol{d}}_0), x \in W_{12,\Lambda}(\hat{\boldsymbol{d}}_0)$ for the sub-image of I_2 in the overlap area.

The above registration processing is defined on the original images, so that W_1, W_2 and $W_{12}(\hat{\boldsymbol{d}}_0)$ are rectangular in shape. If the affine or/and warping transformations are applied on I_2, then W_2 and $W_{12}(\hat{\boldsymbol{d}}_0)$ may no longer be rectangular. We use $W^{'}$ to denote the general non-zero area of a transformed image I, so that we denote the overlap area of transformed images by $W_{12}^{'}(\hat{\boldsymbol{d}}_{\boldsymbol{0}}^{\boldsymbol{'}}) = W_1^{'} \bigcap (W_2^{'} + \hat{\boldsymbol{d}}_{\boldsymbol{0}}^{\boldsymbol{'}})$.

E. Matching-feature detection in the overlap area

Matching features are detected in images $I_{1\text{com}}$ and $I_{2\text{com}}$ in order to implement the feature-based approach for registration. In this paper, we simply match Harris corners [15] based on correlation. The matching relationships between the Harris corners in two sub-images are found through neighbor area searching based on a normalized correlation criterion. The epipolar constraints can be used to guide the matching process [16]. A publicly available software package that includes similar functions can be found on Roth's Projection Vision Toolkit (PVT) website [17].

For each feature point, or Harris corner P, located at $\boldsymbol{x}$ in $I_{1\text{com}}$, its matching point in $I_{2\text{com}}$ is searched from among the Harris corners located in a window centered at $\boldsymbol{x}$ in image $I_{2\text{com}}$. This window is the neighborhood searching window. The block-based normalized correlations between every Harris corner within the neighborhood searching window in $I_{2\text{com}}$ and the feature point P in $I_{1\text{com}}$ are calculated. The feature point with maximum normalized correlation is selected and its normalized correlation value is compared with a threshold to determine if it is a good matching point for P. Since the images $I_{1\text{com}}$ and $I_{2\text{com}}$ are very similar to each other, the neighborhood searching window can be set to a relatively small size, which makes matching relatively easy and reliable. More details on the above feature matching algorithm can be found in [18].

The relationship between $I_{1\text{com}}$ and $I_{2\text{com}}$ is represented by two sets of matching feature points $\text{MP}_1 = \{\boldsymbol{x}_{1,n} | n = 1, 2, ..., N\}$ and $\text{MP}_2 = \{\boldsymbol{x}_2(\boldsymbol{x}_{1,n}) | n = 1, 2, ..., N\}$ in $I_{1\text{com}}$ and $I_{2\text{com}}$, respectively. For any feature point $\boldsymbol{x}_{1,n} \in \text{MP}_1$, its matching point in $I_{2\text{com}}$ is denoted as $\boldsymbol{x}_2(\boldsymbol{x}_{1,n}) \in \text{MP}_2$. N is the total number of matching features in the common region. In the following section, we will use the position of these feature points to determine the transformations of the images I_1 and I_2 for overlap area registration.

Some recently developed more reliable matching features, such as SIFT corners [19], PCA-SIFT [20], etc., can also be used. The algorithms for Harris corner detection are relatively simple and thus save computations. For those overlap areas where there are few or no features that can be detected, the matching features can be selected from dense disparity maps. Usually, stitching of overlap areas having low texture is easy to achieve with few artifacts.

III. Overlap area registration using matching features as control points

In this section, we introduce our methods to reduce the registration errors. The affine transformation is used as one technique to adjust the camera pose changes, together with a nonlinear focal-length adjustment procedure. The objective function in the proposed optimization model aims at minimizing the residual

registration errors in order to obtain a seamless mosaic view after stitching. Thus the 'optimal' focal lengths obtained do not necessarily have any significant physical meaning.

A. Camera pose adjustment through affine transformation

The affine adjustment is applied on the *cylindrically-warped* images. For example, assume two *cylindrically warped* images are $\mathcal{W}_{f_0}I_1$ and $\mathcal{W}_{f_0}I_2$. Here, f_0 is the *a priori* estimated focal length of the camera.

In the texture-based approach [10], the affine adjustment on $\mathcal{W}_{f_0}I_2$ is defined by searching for the parameter vector $\hat{\boldsymbol{t}}_2$ that minimizes the average of absolute differences over the overlap area,

$$\hat{\boldsymbol{t}}_2 = \arg\min_{\boldsymbol{t}_2} \frac{1}{|W_{12}^{'}(\hat{\boldsymbol{d}}_{\boldsymbol{0}}^{'}) \bigcap \Gamma|} \sum_{\boldsymbol{x} \in W_{12}^{'}(\hat{\boldsymbol{d}}_{\boldsymbol{0}}^{'}) \cap \Gamma} |\mathcal{W}_{f_0}I_1(\boldsymbol{x}) - \mathcal{A}_{\boldsymbol{t}_2}(\mathcal{W}_{f_0}I_2)(\boldsymbol{x})| \tag{3}$$

where $W_{12}^{'}(\hat{\boldsymbol{d}}_{\boldsymbol{0}}^{'})$ is the overlap between the images $\mathcal{W}_{f_0}I_1$ and $\mathcal{W}_{f_0}I_2$ as defined in Section II-D.

For the feature-matching approach, the corresponding optimal affine transformation is given by

$$\hat{\boldsymbol{t}}_2 = \arg\min_{\boldsymbol{t}_2} \frac{1}{N} \sum_{n=1}^{N} |\mathcal{U}_{f_0}^{-1}\boldsymbol{x}_{1,n} - \mathcal{S}_{\boldsymbol{t}_2}^{-1}(\mathcal{U}_{f_0}^{-1}\boldsymbol{x}_2(\boldsymbol{x}_{1,n}))|^2, \tag{4}$$

where $\mathcal{U}_{f_0}^{-1}$ and $\mathcal{S}_{\boldsymbol{t}_2}^{-1}$ define the forward mapping for cylindrical warping and affine transformations, respectively. This is a standard least-squares problem. More details on the problem formulation can be found in [21]. It can easily be solved using the singular value decomposition (SVD) method or other related algebraic methods (usually $N \gg 3$).

The optimal parameter vector $\hat{\boldsymbol{t}}_2$ is selected to update the current $\mathcal{W}_{f_0}I_2$ by $\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{f_0}I_2)$. The positions of matching features are also updated following the optimal transformation. An affine adjustment can equivalently be applied on image $\mathcal{W}_{f_0}I_1$ to match $\mathcal{W}_{f_0}I_2$.

B. Focal-length adjustment method

The optimal focal length associated with each image for cylindrical warping is selected as a refinement so as to further minimize the registration error. In the proposed algorithm, the focal-length adjustment is

applied on original or de-warped images after some initial geometric transformation (such as the above affine adjustment) has been applied on two adjacent images. For example, assume that we want to apply focal-length adjustment on image I_2, i.e., to find the optimal focal length associated with I_2 that can minimize the overlap registration error with I_1.

First, we warp I_1 with some fixed focal length f_1. It can be the focal length initially estimated from camera calibration, or it can be the focal length obtained from a previous optimization of focal length applied on image I_1. The estimated focal length does not need to be very accurate, which makes the calibration inexpensive. Then, the search for an optimal focal length $\hat{f}_2$ for I_2 using image intensity matching yields [10]

$$\hat{f}_2 = \arg\min_{f_2} \frac{1}{|W'_{12}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}}) \bigcap \Gamma|} \sum_{\boldsymbol{x} \in W'_{12}(\hat{\boldsymbol{d}}'_0) \cap \Gamma} |\mathcal{W}_{f_1} I_1(\boldsymbol{x}) - \mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{f_2} I_2)(\boldsymbol{x})| \tag{5}$$

where $W'_{12}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}})$ is the overlap area between the images $\mathcal{W}_{f_1} I_1$ and $\mathcal{W}_{f_{20}} I_2$ and $\mathcal{A}_{\hat{\boldsymbol{t}}_2}$ is the optimal affine transformation defined in Section III.A. f_{20} is the initial guess of f_2, which also can be the focal length initially estimated from camera calibration, or it can be the focal length from a previous optimization of focal length adjustment applied on image I_2 in an iterative procedure.

The above optimization involves a large amount computation. The alternative approach proposed in this paper is to find the optimal focal length using the positions of matching features in the overlap area. The optimization model for the feature-based approach is then

$$\hat{f}_2 = \arg\min_{f_2} \frac{1}{N} \sum_{n=1}^{N} |\mathcal{U}_{f_1}^{-1} \boldsymbol{x}_{1,n} - \mathcal{S}_{\hat{\boldsymbol{t}}_2}^{-1}(\mathcal{U}_{f_2}^{-1} \boldsymbol{x}_2(\boldsymbol{x}_{1,n}))|^2, \tag{6}$$

which is a scalar optimization problem. After the optimal focal length $\hat{f}_2$ is found, warping with this focal length is applied on I_2 for updating the positions of the features in the feature-based approach, for further processing. As we mentioned, focal-length adjustment is only applied on de-warped images in the proposed algorithm.

In the matching-feature approach, areas with more matching features get better registration. Usually,

these areas are rich in texture and good stitching is required. On the other hand, we can give different weights to the registration errors for matching features at different locations in equation (4) and (6) to adjust residual registration errors at different locations in the overlap area. This observation also supports the following idea: a better registration result can be obtained if the matching features on a narrow stripe area are chosen (special weights assignment), as will be done in Section IV.

C. The overall overlap area registration algorithm

The overall optimization model to register the overlap area between two adjacent images using the texture-based approach is

$$[\hat{\boldsymbol{t}}_2, \hat{f}_1, \hat{f}_2] = \arg \min_{\boldsymbol{t}_2, f_1, f_2} \frac{1}{|W_{12}^{'}(\hat{\boldsymbol{d}_0^{'}}) \bigcap \Gamma|} \sum_{\boldsymbol{x} \in W_{12}^{'}(\hat{\boldsymbol{d}_0^{'}}) \cap \Gamma} |\mathcal{W}_{f_1} I_1(\boldsymbol{x}) - \mathcal{A}_{\boldsymbol{t}_2}(\mathcal{W}_{f_2} I_2)(\boldsymbol{x})| \tag{7}$$

where $W_{12}^{'}(\hat{\boldsymbol{d}_0^{'}})$ is the overlap area between the images $\mathcal{W}_{f_{10}} I_1$ and $\mathcal{W}_{f_{20}} I_2$. f_{10} and f_{20} are initial values of f_1 and f_2, respectively. In the feature-based approach, the corresponding optimization model is

$$[\hat{\boldsymbol{t}}_2, \hat{f}_1, \hat{f}_2] = \arg \min_{\boldsymbol{t}_2, f_1, f_2} \frac{1}{N} \sum_{n=1}^{N} |\mathcal{U}_{f_1}^{-1} \boldsymbol{x}_{1,n} - \mathcal{S}_{\boldsymbol{t}_2}^{-1}(\mathcal{U}_{f_2}^{-1} \boldsymbol{x}_2(\boldsymbol{x}_{1,n}))|^2. \tag{8}$$

Finding jointly optimal $\hat{\boldsymbol{t}}_2, \hat{f}_1, \hat{f}_2$ requires the solution of a complex nonlinear optimization problem. Instead, the following alternating optimization algorithm can be used in an iterative way:

1) set $f_{10} = f_{20} = f_0$ and $g = 0$.
2) The images I_1 and I_2 are warped using focal lengths f_{10} and f_{20}, respectively. The warped version of image I_1 is $\mathcal{W}_{f_{10}} I_1$ and that of I_2 is $\mathcal{W}_{f_{20}} I_2$.
3) Apply affine adjustment on image $\mathcal{W}_{f_{20}} I_2$ to obtain $\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{f_{20}} I_2)$ after registering the overlap area between $\mathcal{W}_{f_{10}} I_1$ and $\mathcal{W}_{f_{20}} I_2$.
4) Apply focal-length adjustment on I_1 to obtain $\mathcal{W}_{\hat{f}_1} I_1$, and based on which the focal-length adjustment can be applied on I_2 to obtain $\mathcal{W}_{\hat{f}_2} I_2$ after de-warping $\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{f_{20}} I_2)$.
5) $g = g + 1$; stop? if not, set $f_{10} = \hat{f}_1$ and $f_{20} = \hat{f}_2$, go to 3).

The conditions to terminate the iterations can be: (1) the registration errors are small enough compared to a predefined threshold; or (2) the change of the registration errors is small enough compared to a predefined threshold; or (3) the number of the iterations g has reached a predefined value. In practice, we used g to terminate the iterations. The actual implementations were carried out using the feature-based approaches. The globally optimal parameters for equation (8) may not necessarily be found in the above alternating optimization. However, the simulation results show that the solutions obtained are generally good enough for the current applications.

IV. Choosing a narrow stripe block for stitching to minimize the discontinuities

The observation that the final stitching of the two images can be implemented just on a narrow stripe area is important. We perform the optimization on all possible stripe blocks with fixed size in the overlap area, instead of over the whole overlap area, and thus a better registration result can be obtained. The updated overall optimization can be expressed as,

$$[\hat{\boldsymbol{t}}_2, \hat{f}_1, \hat{f}_2, \hat{i}] = \arg \min_{\boldsymbol{t}_2, f_1, f_2, i} \frac{1}{|W'_{12,i}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}}) \bigcap \Gamma|} \sum_{\boldsymbol{x} \in W'_{12,i}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}}) \cap \Gamma} |\mathcal{W}_{f_1} I_1(\boldsymbol{x}) - \mathcal{A}_{\boldsymbol{t}_2}(\mathcal{W}_{f_2} I_2)(\boldsymbol{x})| \tag{9}$$

where the stripe block $W'_{12,i}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}})$ is a sub-area of the overlap area with fixed size, i.e. $W'_{12,i}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}}) \subset W'_{12}(\hat{\boldsymbol{d}}'_{\boldsymbol{0}})$. The subscript i denotes the location where the stripe block starts. We are looking for a stripe block within the overlap area that gives the minimum registration error after the proposed affine adjustment and focal-length adjustment have been carried out on the image pair I_1 and I_2. Parameter vector $\hat{\boldsymbol{t}}_2$ and parameters $\hat{f}_1$ and $\hat{f}_2$ on this stripe block are the optimal ones. In practice, the alternating optimization method can also be used by finding the optimal $\hat{\boldsymbol{t}}_2$, $\hat{f}_1$, and $\hat{f}_2$ for the whole overlap area and then searching for the optimal stripe area within the whole overlap area.

In the feature-matching-based implementation, the following observations have to be considered:

- Due to the irregular distribution of the feature points in the overlap area, the stripe block for registration should be large enough to contain a sufficient number of feature points.

- The stripe block where the registration is carried out is not necessarily the same as the block where the stitching will be implemented; the former can be larger and contain the latter.
- There may be no matching features in the best stripe block for stitching.

As a consequence, we define two types of stripe blocks: stripe blocks for registration and stripe blocks for stitching. The optimal affine transformation parameters and optimal focal length are obtained based on the position registration of the matching features in the stripe block for registration and the location of the optimal stripe block for registration is recorded. The stripe block for registration should be large enough to contain sufficient matching features.

If the overlap area of two adjacent views is not very large, we can simplify the algorithm by choosing one reasonable stripe block for registration near the central part of the overlap area. In this case, we will not search for the optimal stripe block for registration but simply search for the optimal stripe block for stitching. In our implementation, the width of the stripe block for registration was usually selected to be five to six times larger than that of the stripe block for stitching. This is not necessarily optimal but it gives good results and it reduces the computations. In this way, the optimization on the stripe block for registration including both focal-length adjustment and affine transformation is

$$[\hat{\boldsymbol{t}}_2, \hat{f}_1, \hat{f}_2] = \arg \min_{\boldsymbol{t}_2, f_1, f_2} \frac{1}{N_j} \sum_{n \in \xi_j} |\mathcal{U}_{f_1}^{-1} \boldsymbol{x}_{1,n} - \mathcal{S}_{\boldsymbol{t}_2}^{-1} (\mathcal{U}_{f_2}^{-1} \boldsymbol{x}_2(\boldsymbol{x}_{1,n}))|^2 \tag{10}$$

where ξ_j is a set which contains an index list of all matching features that are located in the selected stripe block for registration. N_j is the number of matching features in ξ_j.

Within the stripe area for registration, we search for an optimal narrow stripe block for stitching, where the texture registration error is minimized. The stripe blocks are of fixed size and can overlap with each other. The location of the optimal stripe block for stitching can be obtained from

$$\hat{j} = \arg \min_j \frac{1}{|W_{r,j} \bigcap \Gamma|} \sum_{\boldsymbol{x} \in W_{r,j} \bigcap \Gamma} |\mathcal{W}_{\hat{f}_1} I_1(\boldsymbol{x}) - \mathcal{A}_{\hat{\boldsymbol{t}}_2} (\mathcal{W}_{\hat{f}_2} I_2)(\boldsymbol{x})| \tag{11}$$

where $W_{r,j} \subset W_r$ is one possible stripe block for stitching, with j standing for its start position. W_r is

the stripe block for registration. It can be the optimal stripe block for registration $W'_{12,\hat{i}}(\tilde{\boldsymbol{d}}'_{\mathbf{0}})$, or specified as above, depending the size of overlap area.

In the algorithm, the position of the finally selected optimal stripe block for stitching is determined by the texture difference on each possible stripe block instead of the difference between the positions of the matching feature points. The reason is that the visual stitching discontinuities will be visible through texture discontinuities.

It is possible that a sharp luminance change might appear when stitching on a narrow stripe area if the luminance change between two adjacent images is very large. For this kind of situation, a multi-resolution-based stitching method is suggested. Each image for stitching is first decomposed into two sub-band components through a complementary pair of 2D filters. The bandwidth of the lowpass filter is very narrow. Then, the lowpass components of two images are stitched on the whole overlap area and the highpass components of two images are stitched on the narrow stripe area. After obtaining the lowpass and highpass components of the stitched images, the final stitched image can be obtained by adding its lowpass component and highpass component. In this way, sharp luminance changes can be avoided in the stitched images and the stitching quality is good due to the proposed stitching method on a narrow stripe area. The method has been tested with successful results. The stitching implemented through blending. The correspondent weighted pixels in the stitching area but from different source images are added to generate the transition area in the stitched images. The weights change linearly along the horizontal direction in the stitching areas. The summation of correspondent weights at any pixel in the stitching area equals one. Further details on this blending method can be found in [2]. Other stitching algorithms [22] can also be used.

V. Resolution and computation considerations

Each transformation essentially changes the sampling structure where the discrete image is defined. Thus interpolation will be used, which usually reduces the picture's resolution due to the low-pass filtering effect in interpolation.

Fortunately, the above transformations in the registration procedure can be cascaded. Each transformation will only result in a new sampling structure defined on the original continuous image. The interpolation from the original discrete image will be applied only when it is necessary, for example when evaluating $\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2)(\boldsymbol{x}) = (\mathcal{H} I_2)(\mathcal{S}_{\hat{\boldsymbol{t}}_2}(\mathcal{U}_{\hat{f}_2}^{-1}\boldsymbol{x}))$. Because the registration is implemented based on the positions of matching features, no error due to re-sampling will affect the registration precision.

We use $\mathcal{W}_{\hat{f}_1} I_1 \oplus \mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2)$ to represent the image resulting from stitching $\mathcal{W}_{\hat{f}_1} I_1$ and $\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2)$. In the applications of obtaining large FOV pictures by stitching two adjacent views, the final de-warping process can also be cascaded as one of the sampling structure changes. The de-warping processing can be represented by $\mathcal{W}_{\hat{f}_{12}}^{-1}(\mathcal{W}_{\hat{f}_1} I_1 \oplus \mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2))$ and we can see

$$\mathcal{W}_{\hat{f}_{12}}^{-1}(\mathcal{W}_{\hat{f}_1} I_1 \oplus \mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2)) = \mathcal{W}_{\hat{f}_{12}}^{-1}(\mathcal{W}_{\hat{f}_1} I_1) \oplus \mathcal{W}_{\hat{f}_{12}}^{-1}(\mathcal{A}_{\hat{\boldsymbol{t}}_2}(\mathcal{W}_{\hat{f}_2} I_2)) \tag{12}$$

since the same focal length $\hat{f}_{12} = (\hat{f}_1 + \hat{f}_2)/2$ is used and the same reference point is chosen, which is the center of final stitched image which can be calculated before stitching is actually carried out.

The views from panoramas can be obtained in a similar way. The reference center for de-warping and a focal length have to be specified for a segmented view from the panorama in order to implement de-warping.

The proposed method for overlap area registration based on matching feature positions can save a large number of computations since the number of matching features is much lower than the number of pixels in the overlap area. Once the matching features are obtained, the amount of computation in the proposed optimization model to find optimal affine parameters and focal length is negligible when

using the matching-feature approach compared with using the texture-based approach. In addition, the cost functions of the optimization models in the matching-feature approach are directly related to the positions of the matching features. In the texture-based approach, the sampling structures of the images are first changed through the applied transformations and then required texture information is obtained from interpolation to evaluate the cost functions. As a consequence, the texture-based approach requires more additional computations.

In the feature-based approach, there are extra computations in the feature detection and matching, which serve as overhead computations in the overall computation amount. The amount of computation for feature detection and matching is different from method to method. Usually, the searching for matching features is the most expensive step. In the proposed method, the overlap areas are quite similar and thus the searching windows can be set to relatively small sizes, which greatly reduces the computation. In addition, a great deal of research, such as that in [23], is focusing on robust algorithms with low cost. As a consequence, the overhead computations can also be much lower compared with the optimizations in the texture-based approach.

VI. Generating panoramic views

One of the most important applications for view mosaics is to generate panoramic views for IBR applications. In order to generate a cylindrical panorama, we need to stitch several images into one. Assume that images I_1, I_2, I_3, ...I_M are taken in different directions by a camera mounted on a tripod and rotated clockwise roughly around the camera's center. The pre-captured images are overlapped with each other, and cover a 360° view. Thus the right part of I_M overlaps with the left part of I_1.

Usually, the overlap area in the images I_1 and I_2 is first registered through some kind of transformations, such as the proposed algorithm. The transformed images are $I_1^{'}$ and $I_2^{'}$. $I_{1,2}^{'}$ represents the resulting image after stitching images $I_1^{'}$ and $I_2^{'}$. Then the overlap area between the images $I_{1,2}^{'}$ and I_3 is registered, while

transforming I_3 to $I_3^{'}$. Image $I_{1,2,3}^{'}$ is obtained after stitching $I_{1,2}^{'}$ and $I_3^{'}$. Similar procedures are carried out until the transformed image $I_M^{'}$ has been stitched and the stitched image $I_{1,2,...,M}^{'}$ is obtained. Finally, the left side of image $I_{1,2,...,M}^{'}$ has to be registered and stitched with its right side in the overlap area to generate the full panorama. However, the differences between these two parts in the overlap area are usually very large due to the accumulated registration errors, making registration very difficult. The situation is illustrated in Fig.2, where we see that the transformation between I_M and $I_M^{'}$ is very large

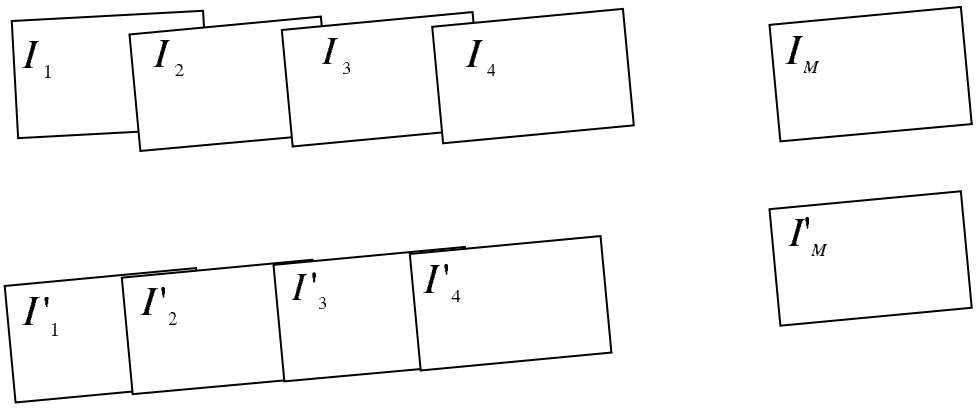

Fig. 2. Illustration of the effect of accumulated errors in 360° panoramic view generation

due to the effect of accumulated errors, which brings significant changes on image $I_M^{'}$. This makes left side of $I_1^{'}$ and right side of image $I_M^{'}$ in the overlap area very different, and the registration of these two sides in the overlap area can be very difficult. The registration errors usually remain very large. In [6], the accumulated global rotation angle is assigned to the transformations on I_1, I_2, .., I_M after finding this global rotation angle through an initial tentative registration. Then registrations and stitching are carried out again considering the assigned rotation angles in each transformation. This will affect the registration results and the way to assign the accumulated global rotation angle on each transformation is tentative. Several trials may be required to experimentally approach a good result. A global optimization is essentially required to minimize the overall overlap area registration errors. However, the algorithm to

obtain the solution of the global optimization may be very complex.

One observation associated with this kind of situation on panoramic view generation is that every adjacent image pair with overlap usually can be registered very well individually. Based on this observation, a novel algorithm to approach the optimal global registration results in an iterative fashion is proposed. In order to describe the proposed algorithm, the overlap area registration problem is first formulated as a coordinate system conversion.

We associate a 2D coordinate system G_i with each pre-captured image I_i $(i = 1, 2, ..., M)$. Without loss of generality, the origin of the 2D coordinate system is selected at the left-top corner of the image. On the top of Fig. 3, the first three pre-captured images and the coordinate systems associated with them are illustrated. In order to generate the panorama, the ideal transformations convert each G_i into $G_i^{'}$ so that there are only horizontal offsets between any two $G_i^{'}$. The bottom of Fig. 3 shows the ideal relationship between $G_i^{'}$ for the first three pre-captured images after registration. Assume $\mathcal{M}_i$ represents the transformation defined on image I_i which transfers the coordinates of points in G_i to their new coordinates in $G_i^{'}$. For a scene point P, assuming the coordinate of its imaging point ${}^{i}p$ in G_i is ${}^{i}\boldsymbol{x}$, the coordinate of its imaging point ${}^{i}p^{'}$ in the $G_i^{'}$ is $\mathcal{M}_i({}^{i}\boldsymbol{x})$. For the proposed registration algorithm, $\mathcal{M}_i({}^{i}\boldsymbol{x}) = \mathcal{S}_{\boldsymbol{t}_i}^{-1}(\mathcal{U}_{f_i}^{-1}({}^{i}\boldsymbol{x}))$ with affine parameter vector $\boldsymbol{t}_i$ and parameter f_i for focal length adjustment defined on I_i.

Consequently, the overlap area registration problem is equivalent to a coordinate system conversion. Suppose I_i and I_j are two adjacent images (where $j = i+1 \mod M$). Two sets of matching feature points ${}^{i,j}\text{MP}_i = \{{}^{i}\boldsymbol{x}_{(i,j),n} | n = 1, 2, ..., N_{i,j}\}$ and ${}^{i,j}\text{MP}_j = \{{}^{j}\boldsymbol{x}({}^{i}\boldsymbol{x}_{(i,j),n}) | n = 1, 2, ..., N_{i,j}\}$ are in the coordinate systems G_i (in image I_i) and G_j (in image I_j), respectively. ${}^{i,j}\text{MP}_i$ denotes the set of matching features between I_i and I_j and ${}^{i}\boldsymbol{x}_{(i,j),n}$ denotes the coordinate in I_i (G_i coordinate system) of one of these matching features. Its correspondent matching feature in I_j is at position ${}^{j}\boldsymbol{x}({}^{i}\boldsymbol{x}_{(i,j),n})$ in the coordinate system G_j. The total number of the matching features between I_i and I_j is $N_{i,j}$. The ideal transformations

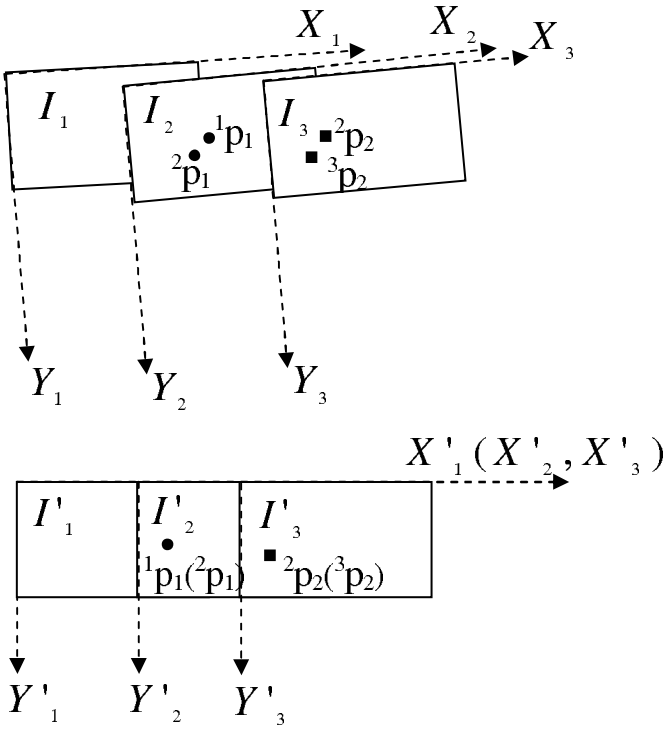

Fig. 3. Illustration of new local minimization model

$\mathcal{M}_i$ and $\mathcal{M}_j$ are defined such that,

$$\mathcal{M}_i({}^i\boldsymbol{x}_{(i,j),n}) = \mathcal{M}_j({}^j\boldsymbol{x}({}^i\boldsymbol{x}_{(i,j),n})) + {}^{i,j}\boldsymbol{x}_0 \tag{13}$$

for all matching features. ${}^{i,j}\boldsymbol{x}_0 = [{}^{i,j}x_0, 0]$ denotes the horizontal translation between $G_i^{'}$ and $G_j^{'}$. In practice, $\mathcal{M}_i$, $\mathcal{M}_j$ and ${}^{i,j}\boldsymbol{x}_0$ are found by minimizing $\sum_{n=1}^{N_{i,j}} |\mathcal{M}_i({}^i\boldsymbol{x}_{(i,j),n}) - \mathcal{M}_j({}^j\boldsymbol{x}({}^i\boldsymbol{x}_{(i,j),n})) - {}^{i,j}\boldsymbol{x}_0|^2$, such as in the algorithm proposed in Section III to register two adjacent overlapped images. In order to generate the panoramas, the above minimization has to be carried out on all pairs of adjacent overlapped images. The objective function for overall optimization is $\sum_{i=1}^{M}(\sum_{n=1}^{N_{i,j}} |\mathcal{M}_i({}^i\boldsymbol{x}_{(i,j),n}) - \mathcal{M}_j({}^j\boldsymbol{x}({}^i\boldsymbol{x}_{(i,j),n})) - {}^{i,j}\boldsymbol{x}_0|^2)$. A very complex global optimization algorithm might be required to solve this problem.

In this paper, an efficient algorithm is proposed to solve the above optimization problem iteratively. A set of local optimizations are carried out to approach the solution of the global problem. One local optimization can be illustrated using Fig. 3. In the following description, the goal is to obtain the optimal

transformation defined on I_j (where $j = 2$ in Fig. 3). Instead of searching for the optimal transformation applied on I_j to match the overlap area between image I_i ($i = j - 1 \mod M$, where $i = 1$ in Fig. 3) and I_j, the optimal transformations applied on I_j are defined to minimize the registration errors both between I_i and I_j and between I_j and I_k ($k = j + 1 \mod M$, where $k = 3$ in Fig. 3). In the proposed feature matching approach, the matching features on both left and right sides of I_j are matched with their correspondences in the adjacent images, I_i and I_k. For example, on the top of Fig. 3, ${}^{1}p_1$ and ${}^{2}p_1$ are a pair of matching features between I_1 and I_2 on the left side of I_2 and ${}^{2}p_2$ and ${}^{3}p_2$ are a pair of matching features between I_2 and I_3 on the right side of I_2. Ideally, for all matching features, both between I_i and I_j and between I_j and I_k, the relationships

$$\begin{aligned} \mathcal{M}_i({}^{i}\boldsymbol{x}_{(i,j),n}) &= \mathcal{M}_j({}^{j}\boldsymbol{x}({}^{i}\boldsymbol{x}_{(i,j),n})) + {}^{i,j}\boldsymbol{x}_0 \\ \mathcal{M}_j({}^{j}\boldsymbol{x}_{(j,k),n}) &= \mathcal{M}_k^{'}({}^{k}\boldsymbol{x}({}^{j}\boldsymbol{x}_{(j,k),n})) + {}^{j,k}\boldsymbol{x}_0 \end{aligned} \tag{14}$$

are satisfied, where ${}^{j}\boldsymbol{x}_{(j,k),n}$ denotes the coordinate of one matching feature between I_j and I_k in G_j with its correspondent matching feature at position ${}^{k}\boldsymbol{x}({}^{j}\boldsymbol{x}_{(j,k),n})$ in G_k. We are looking for the optimal transformation $\mathcal{M}_j$, and thus both $\mathcal{M}_i$ and $\mathcal{M}_k^{'}$ are temporarily fixed. $\mathcal{M}_k^{'}$ can be the initial transformation or the optimal transformation found in the last iteration when using an iterative method. The correspondent objective function for optimization is

$$\frac{1}{N_{i,j}}\sum_{n=1}^{N_{i,j}}|\mathcal{M}_i({}^{i}\boldsymbol{x}_{(i,j),n}) - \mathcal{M}_j({}^{j}\boldsymbol{x}({}^{i}\boldsymbol{x}_{(i,j),n})) - {}^{i,j}\boldsymbol{x}_0|^2 + \frac{1}{N_{j,k}}\sum_{n=1}^{N_{j,k}}|\mathcal{M}_j({}^{j}\boldsymbol{x}_{(j,k),n}) - \mathcal{M}_k^{'}({}^{k}\boldsymbol{x}({}^{j}\boldsymbol{x}_{(j,k),n})) - {}^{j,k}\boldsymbol{x}_0|^2. \tag{15}$$

The overall procedure to generate panoramas using the matching-feature approach is as follows:

1) Apply warping with initial focal length f_0 on all pre-captured images I_1, I_2, ... I_M. These are the initial transformations $\mathcal{M}_1^{'}$, $\mathcal{M}_2^{'}$, ... $\mathcal{M}_M^{'}$; set $g = 0$.
2) set $j = 2$ and $\mathcal{M}_1 = \mathcal{M}_1^{'}$.
3) Search for the best transformation $\mathcal{M}_j$ (including both affine transformation and focal length adjustment) for I_j by minimizing equation (15).

4) $j = j + 1 \mod M$.

5) $j = 2$? if not, go to step 3); otherwise, go to step 6).

6) $g = g + 1$; stop? if not, set $\mathcal{M}_1^{'} = \mathcal{M}_1$, $\mathcal{M}_2^{'} = \mathcal{M}_2$, ..., $\mathcal{M}_M^{'} = \mathcal{M}_M$ go to step 2).

The condition to terminate the iteration can be: (1) the registration errors are small enough compared to a predefined threshold; or (2) the change of the registration errors is small enough compared to a predefined threshold; or (3) the number of the overall iterations g has reached a predefined value. In practice, we used g to terminate the iteration. Good results can be obtained by three to five overall iterations. It should be noted that the offsets between two adjacent images, or ${}^{i,j}\boldsymbol{x}_0$, are variable in each local optimization and the optimal offsets are obtained after all transformations have been found. The techniques of registration and stitching on a stripe area can be incorporated in the above algorithm in a straightforward manner.

VII. Experimental results

In this section, we first test the proposed algorithm for the stitching of two adjacent images with some overlap area. In this test, a pair of adjacent images is stitched to synthesize a view with a larger FOV. Then, the results are shown when stitching two registered adjacent images on stitching areas with different widths. The registration is implemented using both the traditional approach and the proposed approach for comparison. The mean square errors (MSE) between the positions of matching features in the overlap areas (equivalent to the obtained minimum in equations (4), (6), etc.) will be used to evaluate the registration results. Using the changes of these MSE values, the contributions from different adjustments that have been presented can be illustrated. Finally, the proposed algorithm for generation of 360° panoramas is implemented as one application for view mosaicking. In all experiments, the original images were captured by a digital camera mounted and rotated on a tripod, at different rotation angles. The camera poses were just roughly adjusted to be levelled by hand. The image size for all pre-captured

images is 1024 pixels in width and 768 pixels in height.

Fig. 4 shows the results obtained in experiment 1. The overlap images $I_{1\text{com}}$ and $I_{2\text{com}}$ with detected Harris corners are shown in Fig. 4 (a), (b). The geometric relationships between the matching features after registration using equation (2) are shown in Fig. 4 (c). The geometric relationships between these matching features after overlap area registration using the proposed algorithm are shown in Fig. 4 (d). If we simply warp the images with the initial focal length and adjust the two images through shifting along horizontal and vertical directions (traditional registration approach), the stitched results are shown in Fig. 4 (e). The stitching was carried out in the indicated stripe area. Fig. 4 (f) shows the stitching result after applying the proposed registration algorithm and blending on the indicated stripe block. The blending is carried out in the area between the two vertical lines in each stitched image. It is clear that the match is much better and that the stitched view is significantly improved with the proposed registration algorithm (for example as the edges of the shelves near the bottom). The authors have also tested some commercial software packages, and none of them give better or equivalent results compared to the results from the proposed method. In this experiment, we can notice that there is a possible faulty matching feature at the tops of the images, which are low texture areas. However, it does not significantly affect the proposed registration algorithm and the stitching result because the number of matching features in the low texture area is very small.

Experiment 2 was carried out on a pair of images intentionally chosen to have large depth variations. Due to the large depth variations for these two adjacent images, it is very difficult to get a good registration result in the overlap area. The overlap images are shown in Fig. 5 (a), (b). Fig. 5 (c) shows the stitching result using traditional registration approach and blending in a wide area, with the stitching result using traditional registration approach but blending in a stripe area shown in Fig. 5 (d). Fig. 5 (e) shows the stitching result after registration using proposed method but blending in a wide area, with the stitching result after registration using proposed method and blending on a stripe area shown in Fig. 5(f). We can

TABLE I

MSE BETWEEN THE POSITIONS OF MATCHING FEATURES IN EXPERIMENT 1 AND 2. (1) AFTER RELATIVE SHIFTING. (2) AFTER INITIAL WARPING AND RELATIVE SHIFTING. (3) AFTER APPLYING AFFINE ADJUSTMENT ON I_1. (4) AFTER APPLYING FOCAL-LENGTH ADJUSTMENT ON I_1. (5) AFTER APPLYING FOCAL-LENGTH ADJUSTMENT ON I_2.

Experiment	(1)	(2)	(3)	(4)	(5)
1	0.3364	0.3063	0.2182	0.2176	0.2158
2	1.1798	0.6034	0.5872	0.5743	0.5675

see that the proposed method can give better registration results and the technique of stitching on a stripe area can significantly improve the visual quality of the stitched images.

Table I shows the MSE between the positions of matching features during the registrations in Experiment 1 and 2. The values were calculated based on the same set of matching features in the overlap area in each experiment, respectively. The data in column (1) are the minimal MSE values between the positions of the matching features after applying the traditional registration approach. The data in column (2) show minimal MSE values between the positions of the matching features after initial warping and relative shifting of two adjacent images along horizontal and vertical directions. The data in column (3) show the MSE between the positions of the matching features after affine adjustment. The data in column (4) and (5) show the MSE between the positions of the matching features after focal length adjustments on two adjacent images, respectively. Because the widths of overlap areas, distributions and numbers of detected matching features are different in Experiment 1 and 2, the MSE values between Experiment 1 and 2 are not comparable.

In the last experiment, we tested the proposed registration algorithm in generating 360° panoramas. A tentative panoramic view is generated using the traditional approach in order to check the imaging conditions of the pre-captured images. The procedure was described in Section IV with $M = 12$. We

found that there is a huge vertical offset (around 200 pixels) between the left side and right side of the stitched images $I'_{1,2,...,12}$ due to the accumulated errors. A simple global rotation introduced severe distortion, bringing significant difference between the left side and right side of $I'_{1,2,...,12}$ in the overlap area.

The proposed algorithm for global registration in generating panoramas was implemented and the final stitching results on the overlap area between the first image and the last image are shown in Fig. 6. Fig. 6 (a) to (c) show the stitching results after one iteration, two iterations and three iterations, respectively. The results show that the proposed algorithm converges very fast in obtaining an optimized solution.

VIII. CONCLUSIONS

This paper has proposed an overlap-area registration technique for view mosaicking using an affine transformation adjustment together with a nonlinear refinement through focal length adjustment. A feature-based overlap-area registration algorithm has been described for implementation of the proposed method. The principle of using matching feature points as control points for view transformations can be applied more broadly. One major advantage of using matching feature points as control points is that the amount of computation can be greatly reduced. Moreover, the strategy of registration and stitching on a stripe block within the overlap area can significantly improve the visual quality of the mosaicked images.

In our approach, the registration of the adjacent images was formulated as a general coordinate-system transformation module which is independent of the particular registration methods. The framework for generating panoramas through adjacent image registration can then be clearly illustrated. Due to the complexity in obtaining a global optimization solution for the registration errors in the overlap areas of all pre-captured images, a novel simple iterative algorithm was proposed with the experimental results showing that the results converge to a good solution very quickly.

In addition, we observe that these view transformations can be considered as sampling structure changes

which can be combined into one step in order to avoid unnecessary interpolation, which will generally reduce the texture resolution of the images. The principle can be applied to other applications involving multiple transformations on the images, when some or all of these transformations can be combined into a single one.

The proposed algorithm was applied to view stitching using images taken by a conventional digital camera mounted on a tripod and pointed in different directions with some overlap area between the adjacent views. Methods to extend the proposed algorithm for view stitching in more general situations and other types of cameras are currently being studied.

ACKNOWLEDGEMENT

This research was supported by Science and Engineering Research Canada (NSERC). The authors would like to thank Dr. Gerhard Roth for discussions on algorithms for feature matching.

REFERENCES

[1] H.-Y. Shum, S. Kang, and S.-C. Chan, "Survey of image-based representations and compression techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 1020–1037, November 2003.

[2] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Application*, vol. 16, pp. 22–30, March 1996.

[3] S. E. Chen, "Quicktime VR — an image-based approach to virtual environment navigation," *Computer Graphics (SIGGRAPH'95)*, pp. 29–38, August 1995.

[4] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1144–1154, October 2000.

[5] PanoGuide. Available: http://www.panoguide.com/.

[6] H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," *Intern. J. Comput. Vis.*, vol. 36, no. 2, pp. 101–130, 2000.

[7] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Mosaic representation of video sequences and their applications," *Signal Process., Image Commun.*, vol. 8, no. 4, pp. 327–351, 1996.

[8] Z. Zhu, G. Xu, E. M. Riseman, and R. Hanson, "Fast generation of dynamic and multi-resolution 360-degree panorama from video sequence," *Proc. IEEE Int. Conf. on Multimedia Computing and Systems*, vol. 1, pp. 400–406, 1999.

[9] M. Traka and G. Tziritas, "Panoramic view construction," *Signal Process., Image Commun.*, vol. 18, pp. 465–481, July 2003.

[10] X. Sun and E. Dubois, "A novel algorithm to stitch multiple views in image mosaics," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. III–481 – III–484, May 2004.

[11] T. Beier and S. Neely, "Feature-based image metamorphosis," *Computer Graphics (SIGGRAPH'92)*, pp. 35–42, 1992.

[12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, vol. 2, pp. 257–263, 2003.

[13] Y. Kanazawa and K. Kanatani, "Image mosaicing by stratified matching," *Image Vis. Comput.*, vol. 22, pp. 93–103, 2004.

[14] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[15] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. The Fourth Alvey Vision Conference*, pp. 147–151, 1988.

[16] E. Vincent and R. Laganière, "Matching with epipolar gradient features and edge transfer," *Proc. IEEE Int. Conf. Image Processing*, pp. 277–280, September 2003.

[17] G. Roth (2003, May 23). Projective Vision Toolkit. Avaible: http://www.cv.iit.nrc.ca/ $\sim$ gerhard/PVT/.

[18] A. Whitehead and G.Roth, "The projective vision toolkit," *Proc. Modelling and Simulation*, pp. 204–209, May 2000.

[19] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Intern. J. Comput. Vis.*, vol. 2, no. 60.

[20] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 511–517, 2004.

[21] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Prentice Hall, 2001.

[22] J. Davis, "Mosaics of scenes with moving objects," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 354–360, 1998.

[23] D. M. Mount, N. S. Netanyahu, and J. LeMoigne, "Efficient algorithms for robust feature matching," *Pattern Recognit.*, vol. 32, pp. 17–38, 1999.

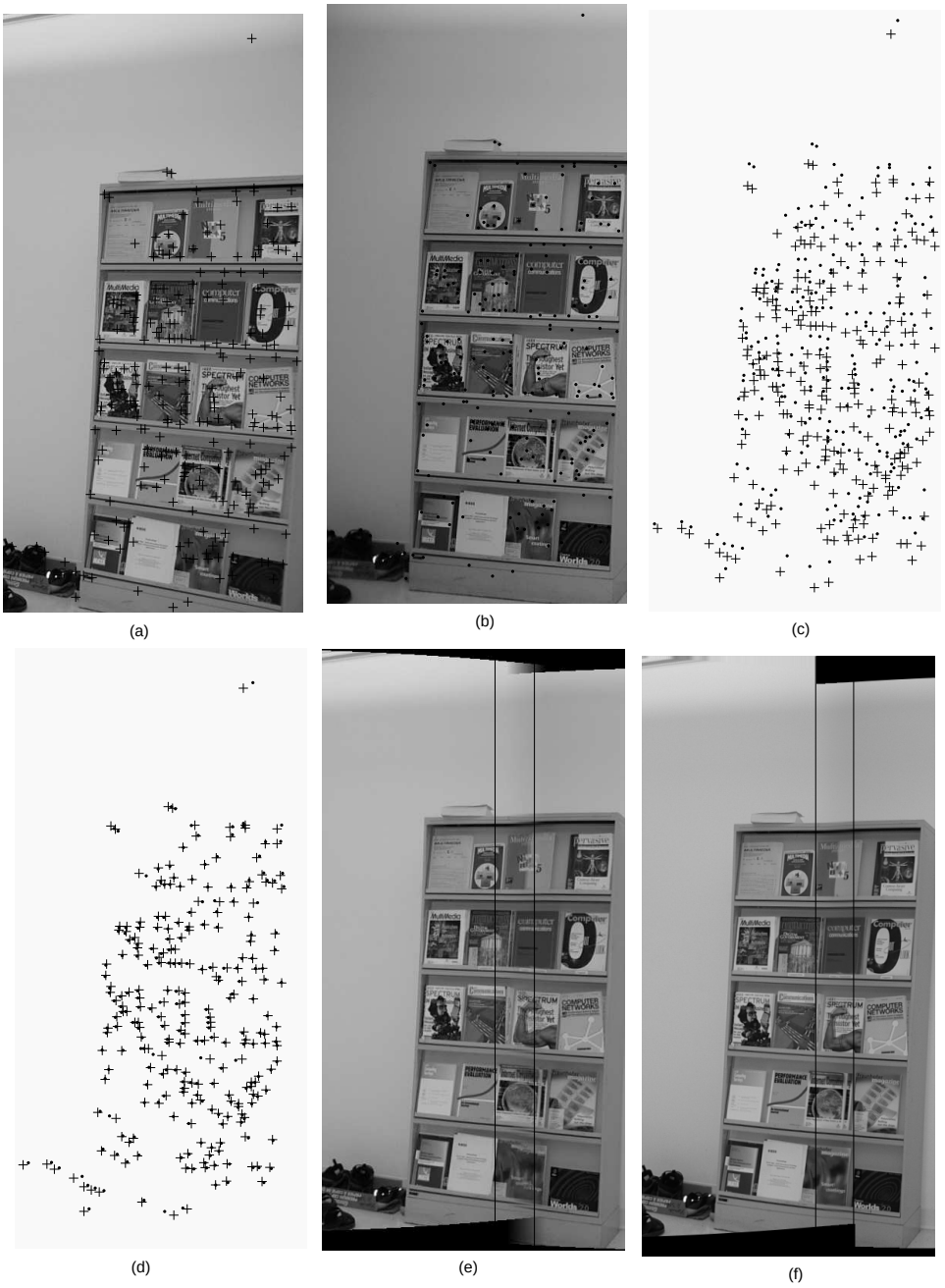

Fig. 4. Results for experiment 1. (a) The matching features in $I_{1\text{com}}$. (b) The matching features in $I_{2\text{com}}$. (c)The relationships between matching features after registration using equation (2). (d) The relationships between matching features after registration using the proposed algorithm. (e) Stitching result with traditional registration approach but blending on a stripe block. (f) Stitching result after registration using proposed method and blending on a stripe block.

Fig. 5. Results for experiment 2. (a)$I_{1\text{com}}$. (b) $I_{2\text{com}}$. (c)Stitching result with traditional registration approach and blending in a wide area. (d) Stitching result using traditional registration approach but blending in a stripe area. (e)Stitching result after registration using proposed method but blending in a wide area. (f) Stitching result after registration using proposed method and blending on a stripe area.

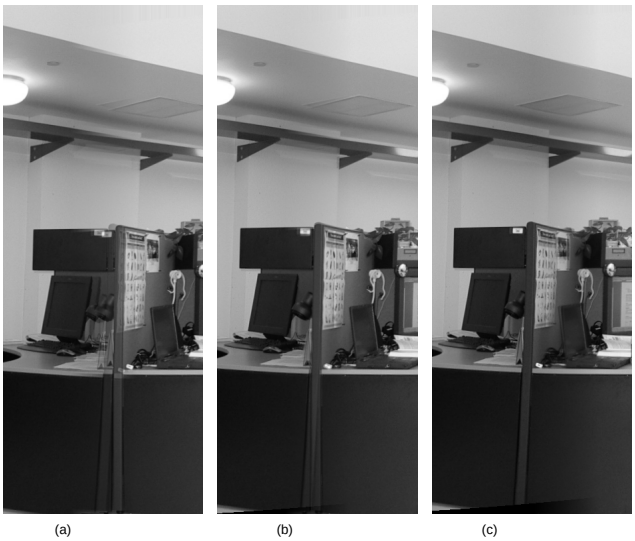

Fig. 6. The stitching results in the overlap area between the first image and the last image in generating panoramas with different iterations using proposed algorithm: (a) after 1 iteration (b) after 2 iterations (c) after 3 iterations