Globally Consistent Alignment for Planar Mosaicking via Topology Analysis

Menghan Xia, Jian Yao*, Renping Xie, Li Li

Computer Vision and Remote Sensing (CVRS) Lab, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, P.R. China

Wei Zhang

School of Control Science and Engineering, Shandong University, Jinan, Shandong, P.R. China

Abstract

Over the past decade, many image mosaicking methods have been proposed in robotic mapping and remote sensing applications. However, most of these methods mainly focus on the optimizing problem of minimizing the image alignment error, which can't necessarily guarantee the global consistency of the mosaicking result, especially for the case of wide-range pseudo-planar scenes prone to suffering from severe perspective distortion. In this paper, we propose a generic framework for globally consistent alignment of images captured from approximately planar scenes via topology analysis, capable of resisting perspective distortion meanwhile preserving local alignment accuracy. Firstly, to estimate the topological relations of images efficiently, we search for a main chain connecting all images over a fast built similarity table of image pairs (mainly for unordered image sequence), along which potential overlapping pairs are incrementally detected according to the gradually recovered geometrical positions and orientations. Secondly, all the sequential images are organized as a spinning tree through applying a graphic algorithm on the topological graph, so as to find the optimal reference image which minimizes the total number of error propagation. Thirdly, we perform the global consistent alignment with the topology analysis in an ingenious strategy that images are initially aligned by groups via the robust affine model, followed by the model

*Corresponding author.
Email address: jian.yao@whu.edu.cn (Jian Yao)
URL: http://cvrs.whu.edu.cn/ (Jian Yao)

refinement under the anti-perspective constraint, through which the optimal balance between aligning precision and global consistency can be achieved. Finally, experimental results on several challenging aerial image sets sufficiently illustrate the validity of the proposed approach.

Keywords: Topology Estimation, Reference Image, Graph Analysis, Global Consistency, Image Mosaicking

---

1. Introduction

Owing to the rapid developments in obtaining optical image data from areas beyond human reach, there is a high demand from different research and engineering fields for creating large range mosaicked images. In fact, image mosaicking is a procedure that merges two or more images with overlapping areas into a single composite image as seamless as possible in both geometry and color tone. The critical first step in the mosaicking process is accurately aligning images into a common coordinate system, which directly influences the mosaicking quality [1, 2, 3]. As a strict aligning model, homography is often used to describe the relationship between two images of a 3D plane or two images captured from the same camera center [4]. Because of the limitation of motionless position, two or multiple images captured from the same camera center and toward different orientations are mainly used to make a ground panorama with an omniscient point of view [5]. On the contrary, mosaicking images of a 3D planar scene permits the camera moving freely, which is popular in robotic mapping and remote sensing applications [6, 7]. Recently, some mosaicking methods not limited to this two geometric conditions have been proposed to extend the range of applications [8, 9, 10]. Specially, in this paper, we focus on mosaicking images from an approximately planar scene known as planar image mosaicking. Under the challenge of both pseudo-plane and accumulation error, a lot of related studies have been presented in the literature of the last decade. However, the performance considering both accurate alignment and global consistency still remains to be further improved.

Generally, the image alignment approaches can be divided into two categories: area-based approaches [11, 12] and feature-based ones [13]. Because of the high computational cost, the area-based approaches are seldom used in the mosaicking missions of large scale [14]. As for the planar mosaicking problem, such as aerial image mosaicking, the feature-based approaches are usually applied to recover the homography model between images [15, 16, 17] due to the fact that the ground scene can be regarded as an approximate plane observed from the aerial photographic camera. To improve the mosaicking result, many optimization algorithms have been proposed to achieve a global alignment. A typical global optimization method is "Bundle Adjustment" [18, 19], which aims at finding an optimal solution minimizing the total reprojection error [20]. To provide a good initial solution for global optimization, Xing et al. [21] proposed to first apply the Extended Kalman Filter [22] onto the local area, and then refine all the parameters globally. To avoid the non-linear optimization, Kekec et al. [23] employed the affine model to optimize the initial alignment made by the homography model in the global optimization. Some methods [24, 25] utilized the topological structure information of images to achieve a global registration. To prevent image suffering down-scaling effect, Elibol et al. [26] proposed to optimize point positions in the mosaicked frame and the alignment model in an alternate iteration scheme.

All those methods concentrating on the optimizing strategy seeking for an alignment with the least registration error can usually composite a satisfied mosaic image from several or dozens of images. However, sequential images taken from a wide-range area can always make the global consistency inaccessible for them, because in the case of pseudo-plane violating the strict geometric model, the least-registration-error principle is prone to causing a severe accumulation of perspective distortions. To release this problem, Caballero et al. [22] proposed to use the hierarchical models according to the alignment quality of images, where the model with less degree of freedom (DoF) is used for images with bigger parallax. The essence of this method is to make a trade-off between improving aligning precision and resisting perspective distortion. In fact, a more reasonable solution is to allow continuous transition between aligning models according the predefined constraint, instead of regarding the model selection as a binary problem. This idea has been detailedly investigated in our previous work [27].

As to the large-scale mosaicking problem, utilizing the topology among images is another effective way to improve the mosaicking result. On the one hand, the potential overlapping relations in topology contribute on the global alignment greatly by providing lots of joint constraints, on the other hand, based on the topological graph, some graphic algorithms can be applied to optimize aligning strategy and reduce error propagation. To estimate the topology efficiently, Elibol et al. [28] used the low-cost tentative matching combined with the Minimum Spanning Tree (MST) solution to detect overlapping relations in an iterative scheme and decide when to update the topological estimation via information-theory principles. The algorithm is efficient as a whole, but the strategy of detecting potential overlapping pairs is not efficient enough, because the detection and the alignment are divided into two independent steps in each iteration, which induce many invalid matching attempts. As for the selection of the reference image, Richard et al. [29] stated that a reasonable choice is the most central image geometrically. This idea is obviously reasonable due to the fact that the central image usually has the shortest distance to all other images on average. However, they didn't give any solution about how to find such an image. To solve this problem, Choe et al. [30] applied a graphic algorithm to select the reference image with the lowest cumulative registration error, but the registration error between each image pair have to be calculated in advance.

In this paper, for mosaicking images taken from a wide-range approximately planar scene, we propose to achieve a visually satisfactory mosaic image with both accurate alignment and global consistency through two technical means: (1) utilizing topology analysis to strengthen registration constraints and reduce error propagation; (2) adopting the alignment strategy of allowing continuous transition between different aligning models, to adaptively keep the optimal balance between alignment accuracy and global consistency (i.e., no obviously perspective distortion). Firstly, we initialize an approximate similarity matrix for image pairs in a fast way, which is combined with the Minimum Spanning Tree (MST) to find the main chain for an unordered image sequence. Then, other potential overlaps are detected incrementally with the gradually recovered geometric positions along the main chain. Because of the synchronism of overlap detection and image location, our proposed topology estimation strategy is more efficient than the method used in [28]. Secondly, all the sequential images are organized as a spinning tree through the classical Floyd-Warshall algorithm, so as to find the optimal reference image with the least cascading times when projecting other images to the reference plane. Obviously, such a reference image benefits in reducing error accumulation. Finally, a globally consistent alignment strategy is proposed to align images into a common coordinate system, which combines the affine model with the homography model effectively. The initial alignment is made by the robust affine model by groups and the globally homography refinement is followed under the anti-perspective constraint, to improve the alignment accuracy on the premise of global consistency not affected. Our proposed approach was sufficiently examined through several groups of experiments on two challenging aerial image datasets and the performances were comprehensively evaluated by comparing with the state-of-the-art algorithm and a famous commercial software.

The remainder of this paper is organized as follows. The proposed framework is detailed in Section 2, which is comprised of topology estimation, selection of reference image, and global alignment. Experimental results are provided in Section 3 followed by the conclusions are drawn and future works are provided in Section 4.

2. Our Approach

Aiming at achieving the mosaicking result with both accurate alignment and global consistency, we propose a generic framework for globally consistent alignment of images captured from an approximately planar scene as shown in Figure 1, which is composed of three modules: topology estimation, selection of reference image, and global alignment. First, the sequential images are inputed for topology estimation, through which the obtained topological graph and matching results are used to search the optimal reference image via graph algorithm and to provide feature correspondences for the global alignment, respectively. Finally, according to the reorganized aligning hierarchy, all the images are aligned by a specially designed double-model under the global optimization framework. Due to the versatile topology estimation, the proposed framework is suitable for both time-consecutive image sets and unordered image sets.

For the description convenience in the following, the frequently used notations in this

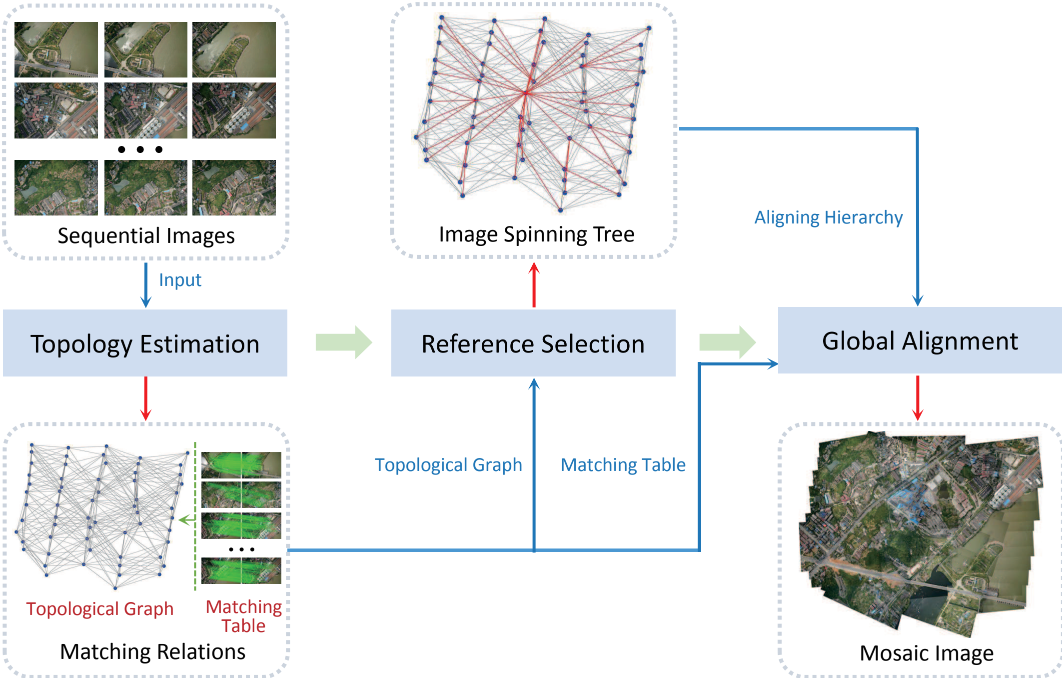


Figure 1: The flowchart of our proposed framework for globally consistent alignment of images. The blue and red thin arrows denote the input and output of each processing module, respectively, and the wide green arrows indicate the execution sequence.

paper are summarized below :

- $\mathbf{I}_i$ - the i-th image in the sequential images.
- $\mathbf{A}_i$ - the 3×3 affine transformation matrix (6DoF) relating $\mathbf{I}_i$ to the reference frame.
- $\mathbf{H}_i$ - the 3×3 homography transformation matrix (8DoF) relating $\mathbf{I}_i$ to the reference frame.
- $\mathbf{x} = [x, y, w]^\top$ - the homogeneous coordinate of a feature point.
- $\mathbf{x}_{i,j}^k$ - the 2D coordinate of the k-th matched feature in $\mathbf{I}_i$ corresponding to the k-th matched feature $\mathbf{x}_{j,i}^k$ in $\mathbf{I}_j$.
- $M_{i,j}$ - the total number of matches between $\mathbf{I}_i$ and $\mathbf{I}_j$.
- $\varpi(\mathbf{x}) = [x/w, y/w]^\top$ - the function transforming the homogeneous coordinate of a 2D point into the non-homogeneous coordinate.

2.1. Fast Topology Estimation

The image topology of the surveyed area is usually represented by a graph where an image stands for a node and the overlapping relationship between image pair is denoted by an edge or a link. Topology estimation means to find the existing overlapping relationships among all images. In this section, we try to find all the potential overlapping image pairs by utilizing the gradually recovered geometric positions of images in the time-consecutive order on the mosaicking plane, instead of simply doing matching attempts. As for an unordered image sequence, finding a main chain connecting all images can make the problem the same as that of the time-consecutive image sequence. Therefore, an efficient strategy can be proposed to find the complete topology with the minimum number of image matching attempts.

2.1.1. Finding Main Chain with Most Reliability

For a sequence of n images, the main chain consisting of $(n-1)$ edges connects all the nodes/images in the graph. More strictly, it is defined as a spanning tree of an undirected graph in graphic theory [31, 32]. Obviously, there is no need to find a main chain for the time-consecutive image sequence due to that their time-consecutive links have implied a main chain already. That's to say, this step is mainly set for the case of finding the image topology of an unordered image set.

Given an unordered image set, we have to measure the similarities between image pairs in advance of finding a main chain. Here, the similarity measurement is intended to be computed in an approximate but efficient way. To achieve this goal, for each image, we select a subset of SURF features extracted from it, and the similarity between image pair is defined as the number of candidate point matches whose descriptor vector distances are less than some given distance. Specially, to increase the corresponding probability, the subsets are generated by selecting features extracted from the same scale layer in the SURF detector, instead of sampling randomly. In our experiments, the features from the second scale layer of the total four octaves were selected as the subset representing each image, which hold a stable ratio of $22 \pm 3\%$ almost for all kinds of images. The computational cost of this similarity measure is comparatively low, since it mainly involves computing the distances between a small set of descriptor vectors. Over the exhaustive comparison, all the similarity values between image pairs implying the initial overlapping information are organized in the form of a matrix $\mathbf{S}$, where $\mathbf{S}(i,j)$ represents the similarity between images $\mathbf{I}_i$ and $\mathbf{I}_j$. The value of $\mathbf{S}(i,j)$ from small to large means an increasing similarity between images $\mathbf{I}_i$ and $\mathbf{I}_j$, which can be regarded as the probability of images $\mathbf{I}_i$ and $\mathbf{I}_j$ sharing an overlap.

Although the similarity table built by this way is not reliable, it is qualified to provide the initial similarity information just for finding a main chain under an iterative scheme. Based on the similarity matrix, the reciprocals of those non-zero similarity values are set as the weights for the edges of the graph, i.e., $W(i,j) = \frac{1}{\mathbf{S}(i,j)}$. Given such a graph, we try to select a linkage path that connects all the nodes with the highest total reliability, i.e., the lowest sum of weights. This idea is effectively implemented in the following two-step iterative scheme.

Maximum Reliability: This is realized by finding the Minimum Spanning Tree (MST) of the current weighted graph. The MST is a spanning tree whose edges have the minimum total weight in all the spanning trees of the graph. So, the MST represents the connected tree composed of the most similar image pairs.

Check Connectivity: The algorithm tries to match all the image pairs in the MST. If all the matching attempts succeed completely, the MST is the targeted main chain and the iteration is terminated. On the contrary, when there exists any image pair failed to be matched, we have to modify the weights of the graph where the weights of successfully matched pairs are set as zero while the weights of matching-failed image pairs are set as an infinite value, then it turns to the next iteration.

To examine the difference of the main chains from a time-consecutive image sequence and an unordered image sequence, a subset of the first dataset described in Section 3 was selected to demonstrate the results of topology estimation, which had been performed in both the time-consecutive mode and the unordered one, respectively, as shown in Figure 2. Apart from the difference of the main chains, the topologies estimated in these two different modes are almost the same, which are compared quantitatively in Table 1.

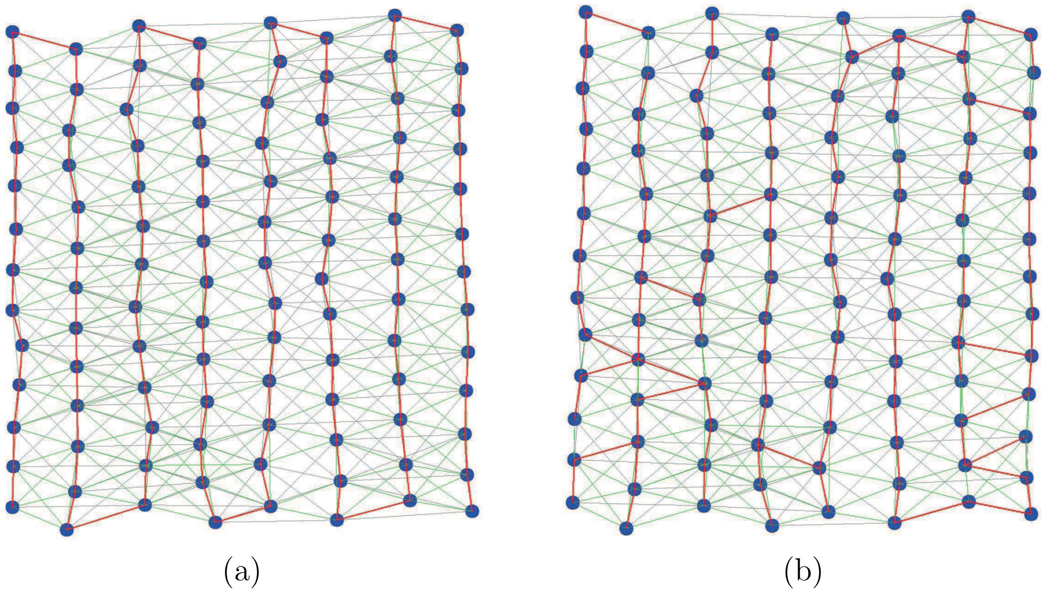


Figure 2: The estimated topologies of an image sequence (104 images) in the time-consecutive mode and the unordered one, respectively: (a) The topology estimated in the time-consecutive mode, where the red edges represent the prior main chain in the time-consecutive order and green edges indicate the numbers of matched features between image pairs are over 100 while gray ones indicate the numbers are less than 100; (b) The topology estimated in the unordered mode, where the red edges represent the main chain linked by the proposed iterative scheme and other edges have the same meanings as in (a).

2.1.2. Detecting Potential Overlapping Pairs

After having obtained the overlapping relationships along the main chain, we move on to detect other potential overlapping pairs for a more complete topology. As mentioned in the beginning of Section 2.1, we can recover the comparative geometric positions of sequential images in a common coordinate system according to the main chain. Based on the geometric information, the potential overlapping pairs can be detected easily. Therefore, there are two problems to be solved: 1) how to recover the comparative geometric positions with the main chain information; 2) how to detect the potential overlapping pairs based on the geometric relationships. In the proposed method, these two problems are solved in a collaborative way, instead of an independent way.

Firstly, we temporarily select a reference image as the mosaicking plane through applying the algorithm detailed in Section 2.2 on the main chain. To recover the comparative geometric positions, we employ the affine model to align images into the mosaicking plane, which is robust in locating the centroids of images. Compared with

Algorithm 1 Detecting potential overlapping pairs

Input: The image set $\mathcal{I} = \{\mathbf{I}_i\}_{i=1}^{n}$ arranged in some order.

Output: The set of overlapping image pairs $\mathcal{P} = \{\mathbf{P}_{ij}\}_{i \neq j}$.

Initialize the located image set $\widehat{\mathcal{I}} = \{\mathbf{I}_1\}$
for each image $\mathbf{I}_i \in \mathcal{I} \backslash \{\mathbf{I}_1\}$ **do**
  Align $\mathbf{I}_i$ with its direct reference image $\mathbf{I}_{\rho(i)}$.
  Initialize the overlapping pairs set $\mathcal{P}_i = \{\mathbf{P}_{i\rho(i)}\}$.
  for each image $\mathbf{I}_j \in \widehat{\mathcal{I}} \backslash \{\mathbf{I}_{\rho(i)}\}$ **do**
    yes/no $\leftarrow$ Detect the overlap between $\mathbf{I}_i$ and $\mathbf{I}_j$.
    if yes **then**
      $\mathcal{P}_i = \mathcal{P}_i \cup \{\mathbf{P}_{ij}\}$.
    end if
  end for
  Realign $\mathbf{I}_i$ with its neighborhood image set $\mathcal{P}_i$.
  $\mathcal{P} = \mathcal{P} \cup \mathcal{P}_i$
  $\widehat{\mathcal{I}} = \widehat{\mathcal{I}} \cup \{\mathbf{I}_i\}$
end for
return $\mathcal{P}$

the affine model, the classic homography model is prone to suffering from the perspective distortion, and the 2D rigid model tends to make a bending trajectory because of error accumulation, which is validated in Section 3.2. Specially, to improve the reliability of the image locations, the images on the main chain are aligned starting from the reference image one by one. As the images being located gradually, the potential overlapping relationships around the newly located image are detected and would be used for optimizing the position of this newly aligned image in the following. This strategy makes a significant contribution to improving the accuracy of the recovered geometric positions, because the simultaneously detected overlapping pairs can provide extra favorable constraints for aligning images. Given a newly aligned image $\mathbf{I}_i$, it checks whether there is any overlap with all the previously aligned image set $\widehat{\mathcal{I}} = \{\mathbf{I}_j\}_{j=1}^{m}$. For an image pair $\mathbf{I}_i$ and $\mathbf{I}_j$, the overlap detection is performed by calculating the distance

Table 1: Comparisons of our topology estimation running in both the time-consecutive mode (a) and the unordered mode (b) (with All-against-all as the ground truth).

Strategy	Successful Attempts	Total Attempts	% of Recall	% of computation on feature matching
Proposed Approach (a)	606	896	94.71	99.42
Proposed Approach (b)	595	905	92.10	84.14
All-against-all	646	5356	100.00	100.00

between their centroids as follows:

$$\delta_{ij} = \frac{\max(0, |c_i - c_j| - |d_i - d_j|/2)}{\min(d_i, d_j)}, \tag{1}$$

where c_i, c_j, d_i and d_j are the centroids and the diameters of the minimum boundary circles of the projection onto the mosaicking plane of $\mathbf{I}_i$ and $\mathbf{I}_j$, respectively. If $\delta_{ij} > 1$, there is no overlap. Otherwise, there may exist an overlap between $\mathbf{I}_i$ and $\mathbf{I}_j$, and we try to match them for verification. Of course, if the matching between $\mathbf{I}_i$ and $\mathbf{I}_j$ has been attempted during finding the main chain, there is no need to repeat the matching attempt again. The whole sketch procedure of our proposed topology estimation approach is described in Algorithm 1. When all the overlapping pairs are obtained, we redefine the similarity matrix as the final topological representation. The original similarity matrix is reset as a zero matrix firstly, and the value of $\mathbf{S}(i, j)$ is replaced with the number of matched points only if $\mathbf{I}_i$ and $\mathbf{I}_j$ have been matched successfully.

It should be noted that the major computation cost of the topology estimation is feature matching between images, as listed in the fourth column of Table 1. Because there is no global optimization or iterative detection, the image alignment and potential overlapping detection have a relatively low computation cost. Besides, as for an unordered image set, the initialization of the similarity matrix occupies the majority of the rest computation actually. That's to say, the topology estimation before image mosaicking is well worthy, which is fundamental to the following process while adds nearly no extra computation load except for the necessary feature matching.

2.2. Optimal Reference Image Selection

As we known, the images alignment is realized through warping each image into the mosaicking plane which is always set by selecting one of the sequential images (named as the reference image). An image without direct overlap to the reference image has to be projected to the mosaicking plane by cascading a series of relative transformation models between other intermediate images. Obviously, less intermediate images used for cascading makes less error accumulation. In fact, there may exist more than one path with the same cascading numbers from an image to another. Considering each cascading implies a different error, we would rather to select the path with the least accumulation error. In terms of this, the optimal reference image should give the lowest sum of accumulation errors from all the other images to the reference image plane. To address this problem, we construct an undirected and weighted graph based on the estimated topology in Section 2.1. According to the similarity matrix obtained by the topology estimation, those image pairs with non-zero values of similarities are linked with edges. As far as the weight (or cost) of an edge concerned, there are two kinds of settings in the existing literature: the reciprocal of the number of matched features [28] and the registration error between the image pair [30]. The former is intuitive and efficient while the latter perceives the error directly at the cost of calculating registration error between all available image pairs in advance. Considering the association between the number of matched features and the registration error, we creatively set the weight of an edge in the graph as follows:

$$w_{ij} = \begin{cases} \text{inf}, & \text{if} \quad M_{i,j} = 0, \\ \dfrac{1}{\log(M_{i,j} + \varepsilon)}, & \text{if} \quad M_{i,j} > 0, \end{cases} \tag{2}$$

where $M_{i,j}$ denotes the total number of matches between $\mathbf{I}_i$ and $\mathbf{I}_j$, and ε is a constant for regularization, which is set as 50 in our experiments. This weight setting equation, which describes the contribution of matched features to the registration accuracy, has the advantages of both efficiency and effect.

Based on the weighted graph, the optimal reference image selection problem is formulated as finding a node with the least total weight of the shortest paths to all the other nodes, which can be solved by the Floyd Warshalls all-pairs shortest path

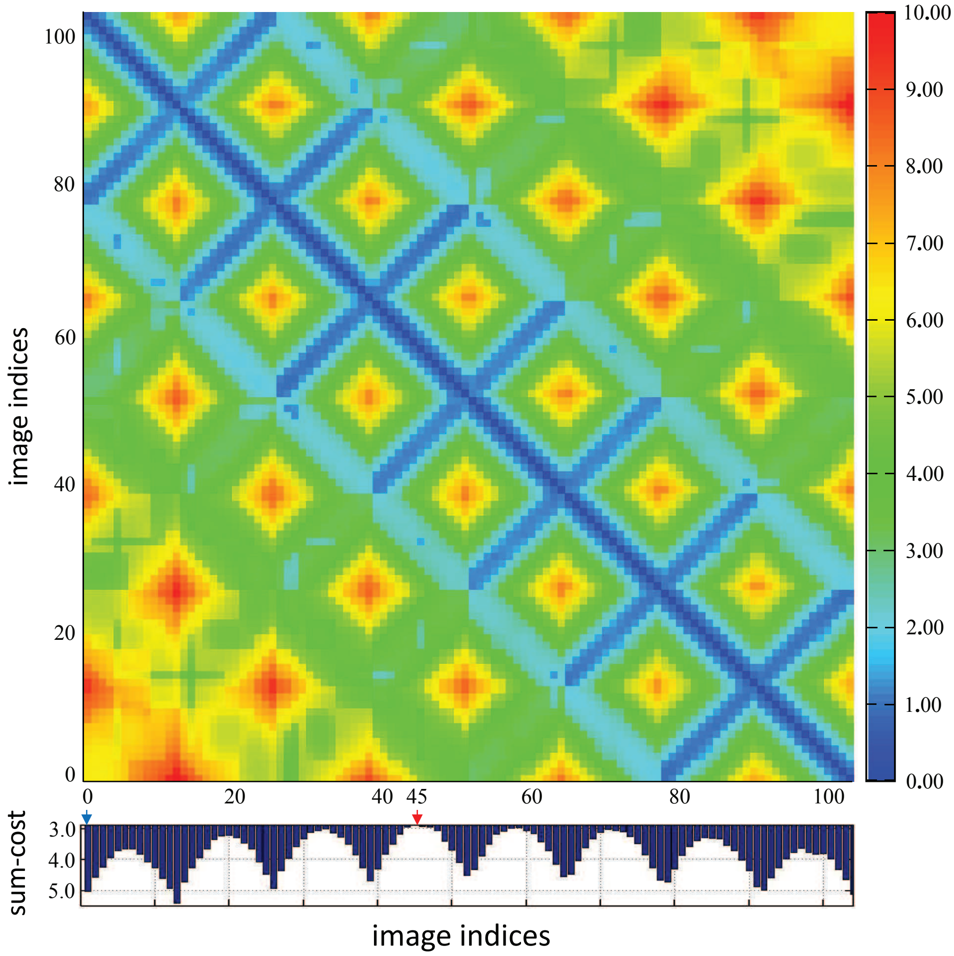


Figure 3: The cost matrix of all-pairs shortest path calculated from a sequence of 104 images. Below is attached the bar chart depicting the mean cumulative cost of each column of the cost matrix. As labeled with a red arrow in the bottom indices, the 45-th column of the cost matrix has the minimum total cost with the mean cumulative cost of 3.04. That's to say, the 45-th image is the optimal reference image. However, the conventional idea to naively select the first image as the reference image gives an much higher mean cumulative cost of 5.23, as labeled with a blue arrow in the bottom indices.

algorithm [33, 34]. The dynamic programming strategy is applied in this algorithm with the computation complexity of $O(3)$, so it is more efficient than running n times of a single source shortest path algorithm. With this algorithm, all shortest paths from a node to any other node can be obtained. When there are n images in a sequence, we build a $n \times n$ size symmetric cost matrix $\mathbf{W}$ where each element records the cost of the shortest path between two images. After running this algorithm, the cost of the shortest path from $\mathbf{I}_i$ to $\mathbf{I}_j$ is saved in $\mathbf{W}(i, j)$. Therefore, the i-th row or column of matrix $\mathbf{W}$

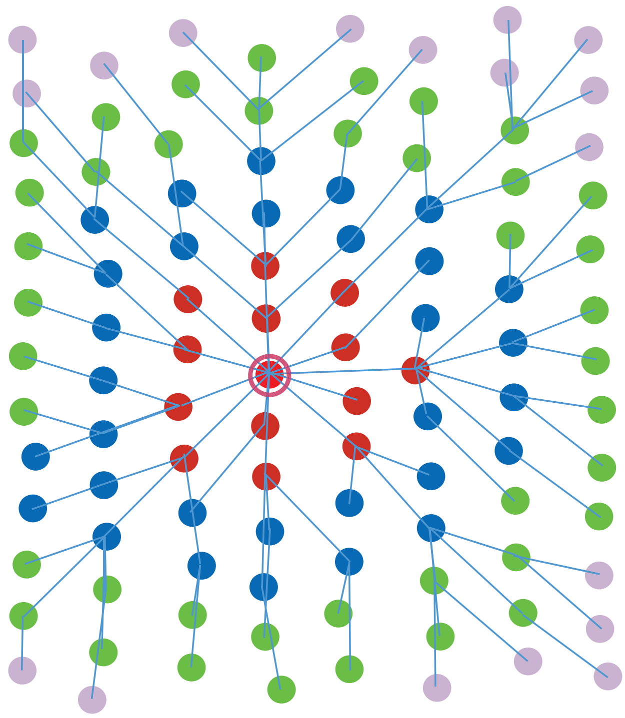


Figure 4: The spinning tree of the graph with the optimal reference image as the root node (marked with red ring). Nodes in different levels of the tree are marked with different colors, and the blue lines link each node and its parent node, which imply the shortest paths from all the other images to the reference image.

indicates the cost of every shortest path from other images to $\mathbf{I}_i$. On this occasion, the accumulated cost of each column in the cost matrix $\mathbf{W}$ can be calculated and the column with the minimum accumulated cost is selected as the reference image. To demonstrate the procedure, the cost matrix $\mathbf{W}$ of a sequence of 104 images is visualized in Figure 3, and the 45-th column with the minimum total cost is labeled with the red arrow in the bottom indices. Specially, the conventional strategy of selecting the first image as the reference image is also highlighted as a comparison. Considering the amount of images, the gap of the mean cumulative costs between the two strategies can make a big difference to the mosaicking result.

Actually, each row in $\mathbf{W}$ corresponds a spinning tree with the node of this row as the root node, which describes the hierarchical relationship of the image nodes. With the selected reference image, the spinning tree of the image sequence described in Figure 3, is displayed in Figure 4. The spinning tree indicates the direct reference image of each image (parent node in graphic terms), which determines the aligning order of images in the following global alignment.

2.3. Globally Consistent Alignment

In general, both the locally aligning accuracy and the global consistency are two basic factors determining the quality of mosaicking result. Under a strict transformation model, these two factors can be guaranteed in a coherent way, where the higher aligning precision contributes on the better global consistency. However, in most practical applications, the observing scenes of pesudo-planes make the frequently-used homography model just an approximate transformation between images. In this case, the aligning model of higher degrees of freedom (DoF) usually makes more accurate alignment but suffers more severe perspective distortion meanwhile, and vice versa. Therefore, we have to deal with these two factors in a trade-off way. To keep the optimal balance between them, the model with a relatively low DoF is employed to make the initial alignment of images robustly, which will be refined with a higher DoF to improve the aligning precision under the anti-perspective constraint.

2.3.1. Robust Alignment by Affine Model

For a robustly initial alignment, we would rather to use the affine model which compromises between the 2D rigid transformation and the homography transformation. On the one hand, the approximately coplanar constraint of images is partly implied in the six-parameter affine model which can suppress severe perspective distortion to some extent, on the other hand, the affine transformation is able to provide a qualified initial solution for the following homography refinement.

According to the spinning tree mentioned in Section 2.2, the sequential images are aligned group by group in the order of breadth-first search, which can decrease the accumulation error of alignment compared to the way of one by one. In this paper, when aligning a new group of images to the reference frame, the overlapping relations between all the previously aligned images and the newly added images, and the overlapping relations between intra-group will be jointly used in the optimization framework. Let $\mathcal{I} = \{\mathbf{I}_i\}_{i=1}^{s}$ be the set of previously aligned images. The affine transformation set

$\mathcal{A} = \{\mathbf{A}_i\}_{i=s+1}^{s+m}$ of the newly added image group $\mathcal{G} = \{\mathbf{I}_i\}_{i=s+1}^{s+m}$ for alignment will be optimized by minimizing the combination of two cost functions as below :

$$E(\mathcal{A}) = E_1(\mathcal{A}|\mathcal{I},\mathcal{G}) + E_2(\mathcal{A}|\mathcal{G}), \tag{3}$$

where the first energy term $E_1(\mathcal{A}|\mathcal{I},\mathcal{G})$ is related to the overlapping relations between $\mathcal{I}$ and $\mathcal{G}$ as follows:

$$E_1(\mathcal{A}|\mathcal{I},\mathcal{G}) = \sum\nolimits_{\mathbf{I}_i\in\mathcal{I},\mathbf{I}_j\in\mathcal{G}} \sum\nolimits_{k=1}^{M_{i,j}} \|\varpi(\mathbf{A}_i\mathbf{x}_{i,j}^k) - \varpi(\mathbf{A}_j)\mathbf{x}_{j,i}^k\|^2, \tag{4}$$

and the second energy term $E_2(\mathcal{A}|\mathcal{G})$ is related to the overlapping relations in $\mathcal{G}$ as follows:

$$E_2(\mathcal{A}|\mathcal{G}) = \sum\nolimits_{\mathbf{I}_i,\mathbf{I}_j\in\mathcal{G}} \sum\nolimits_{k=1}^{M_{i,j}} \|\varpi(\mathbf{A}_i\mathbf{x}_{i,j}^k) - \varpi(\mathbf{A}_j\mathbf{x}_{j,i}^k)\|^2, \tag{5}$$

where the meanings of the notations $\varpi(\cdot)$, $\mathbf{A}_i$, $M_{i,j}$ and $\mathbf{x}_{i,j}^k$ are given in the beginning of Section 2.

As a group of linear equations, Eq. (3) can be solved fast by the Singular Value Decomposition (SVD) method. Note that both the epipolar constraint and the appropriately homography constraint are employed to remove outliers in the SURF points matching algorithm. In fact, we also normalize the coordinates of matched points according to the method proposed in [35], in order to increase the numerical stability by improving the condition number of the coefficient matrix. What's more, the robust estimator MLESAC [36] is used to exclude outliers for affine estimation because it is beneficial for the image mosaicking of quasi-planar scenes.

2.3.2. Model Refinement under Anti-Perspective Constraint

The affine models recovered by groups are mainly used to achieve the robust initial alignment, which guarantee the mosaicking result against the perspective distortion well. However, the aligning precision needs to be further improved due to that the DoF of the aligning model is limited and no global optimization is performed. To improve the aligning accuracy to some extent but not to induce the perspective distortion, the energy function should allow to transit the affine model to the homography model with a higher DoF under some reasonable constraint. In fact, such constraint has been implied in the affine model which has the anti-perspective property relative to the homography model. So, the deviation between the optimal homography transformation and the initially estimated affine transformation is set as a regularization term in the proposed optimization framework.

As the images are aligned by groups, the affine models of all the images $\mathcal{I} = \{\mathbf{I}_i\}_{i=1}^n$ can be obtained, denoted as $\mathcal{A} = \{\mathbf{A}_i\}_{i=1}^n$, which are used as the initial parameters for the homography model in the final global optimization. The homography models $\mathcal{H} = \{\mathbf{H}_i\}_{i=1}^n$ with respective to the reference frame will be optimized in the energy function composed of two mutually contrary terms. The data term set for minimizing the sum of squares of the feature registration errors between images is denoted as:

$$E_d(\mathcal{H}) = \sum\nolimits_{\mathbf{I}_p,\mathbf{I}_q\in\mathcal{I}} \sum\nolimits_{k=1}^{M_{p,q}} \|\varpi(\mathbf{H}_p\mathbf{x}_{p,q}^k) - \varpi(\mathbf{H}_q\mathbf{x}_{q,p}^k)\|^2, \tag{6}$$

where all the aligning models have more free parameters to adjust the positions of points on the mosaicking plane, which is bound to increase the whole precision of alignment. Besides, the residual error is prone to distributing evenly under an uniform energy framework.

Another optimization objective is to keep the global consistency by suppressing the accumulation of the perspective distortions which may emerge in the transition from the affine model to the homography model. The regularization term from the idea that the optimal homography transformation should be close to the initially estimated affine transformation, is expressed as the displacements of the warped features from their initial positions as follows:

$$\begin{aligned} E_r(\mathcal{H}) \;=\; & \sum\nolimits_{\mathbf{I}_p,\mathbf{I}_q\in\mathcal{I}} \sum\nolimits_{k=1}^{M_{p,q}} \big(\|\varpi(\mathbf{H}_p\mathbf{x}_{p,q}^k) - \mathbf{A}_p\mathbf{x}_{p,q}^k\|^2 \\ & \qquad + \|\varpi(\mathbf{H}_q\mathbf{x}_{q,p}^k) - \mathbf{A}_q\mathbf{x}_{q,p}^k\|^2\big)\,. \end{aligned} \tag{7}$$

As depicted in Eq. (7), the regularization term is also denoted by the distances of image feature points as the data term does, which saves the troublesome normalized problem between different kinds of energy terms. So far, the energy terms defined in Eq. (6) and Eq. (7) can be linearly combined to define the final energy function as follows:

$$E(\mathcal{H}) = E_d(\mathcal{H}) + \lambda E_r(\mathcal{H}), \tag{8}$$

where λ is the weight coefficient used for balancing the two terms E_d and E_r, which should be set to an appropriate small value since the constraint isn't a strict one. Theoretically, a bigger value of λ strengthens the global consistency while decreases the accuracy of the local alignment. We set its value from 0.01 to 0.05 in all our experiments. As a typical non-linear least squares problem, Eq. (8) can be solved by the Levenberg-Marquardt (LM) algorithm. However, considering the specialty of this problem, we employ the sparse LM algorithm [37] to save memory and to speed up the computation, which is stated detailedly in Appendix A.

3. Experimental Results

To make a comprehensive study of our approach, three groups of experiments were conducted, including the evaluation on the topology estimation, the evaluation on the selection of initial model, and the comprehensive evaluation on the mosaicking results. Two sets of representative aerial images acquired by different flight platforms and over different landforms, respectively, were used as the experimental dataset. The first dataset, consisting of 744 images from 24 sequentially ordered strips, was captured at a flight height of about 780 meters over an urban area. The original images, with a forward overlapping rate of about 60% were down-sampled to the size of 1000×642 in our experiments. The second dataset, consisting of 130 images with the down-sampling size of 800×533, was captured by an unmanned aerial vehicle (UAV) with a forward overlapping rate of about 70%, which observes a suburb area containing mountains.

Due to the limit of pages, more experimental results and analysis are presented at `http://cvrs.whu.edu.cn/projects/PlanarMosaicking/`, where the dataset and the source code are publicly available for download.

3.1. Evaluation on Topology Estimation

In this section, the topology estimation module of the proposed approach was compared with the classic all-against-all strategy and the state-of-the-art algorithm implemented according to [28] (we name it as Fast-Topology hereafter). The comparisons were performed on the estimated topology of the aforementioned two datasets. To test our approach diversely, the aerial image sequence and the UAV image sequence were respectively processed in two different modes for topology estimation: the time-consecutive mode and the unordered mode. As a robust but exhaustive strategy, matching all-against-all was always used for comparison in topology estimation, the detected overlapping pairs by which can be regarded as the ground truth. Moreover, the successfully matched image pairs and the total matching attempts are combined to evaluate the topology estimation results as the quantitative metrics.

The topology estimation results of the two datasets are summarized in Table 2 and Table 3, respectively, where the first column lists the tested methods, the second column corresponds to the numbers of successfully matched image pairs, the third column contains the total numbers of matching attempts, the last two columns give the percentages of the second and third columns with respect to the all-against-all strategy. As the tables show, both our approach and Fast-Topology [28] almost recovered the complete topology as the all-against-all strategy did, but with a much less amount of image matching attempts. Although there are some omissions with respect to all-against-all, the major overlapping relations had been detected successfully in our approach, which can be observed in the topological graph depicted in Figure 5(a) and Figure 6(a), respectively. It implies that most of the undetected overlapping pairs probably share very small overlapping areas and make little difference to the mosaicking results. Compared to Fast-Topology [28], our approach has roughly the same recall rates but less total matching attempts, which benefits from two key strategies used in the potential overlapping pairs detection. The one is the selection the temporary reference image, which is determined by applying the strategy detailed in Section 2.2 on the main chain, instead of setting the first image simply like Fast-Topology. The other is that the position of the newly added image is simultaneously adapted along with the potential overlapping relations being detected, which improves the alignment accuracy and so does the efficiency. Differently with ours, the procedures of detecting the potential overlapping pairs and adapting alignment of images with the detecting results are divided into two independent steps in Fast-Topology. Therefore, it inevitably introduced many unnecessary matching attempts because of the inaccurate alignment in the first few iterations though it can find most of the existing overlapping relations after several iterations.

As mentioned in Section 2.2, the estimated topology is used to search for the optimal reference image, by the way of which the images are organized as a spinning tree implying

Table 2: Comparisons of the topology estimation obtained by different approaches on the first dataset (with All-against-all as the ground truth).

Strategy	Successful Attempts	Total Attempts	% of Recall	% of Attempts As to All-against-all
Our Approach	5197	7771	97.83	2.81
Fast-Topology [28]	5229	9601	98.43	3.47
All-against-all	5312	276396	100.00	100.00

Table 3: Comparisons of the topology estimation obtained by different approaches on the second dataset (with All-against-all as the ground truth).

Strategy	Successful Attempts	Total Attempts	% of Recall	% of Attempts As to All-against-all
Our Approach	781	934	95.36	11.14
Fast-Topology [28]	793	1336	96.83	15.93
All-against-all	819	8385	100.00	100.00

the aligning order for the global alignment. Here, the spinning trees with the reference image as the root node, are expressed by a group of red edges of the topological graph in Figure 5(b) and Figure 6(b), corresponding to the first and second datasets, respectively. It's easy to find that the selected reference images can always locate in the central part geometrically, no matter of the ruled aerial data or the strip shaped UAV data. Noticeably, the layouts of the image centroids recovered via the final global alignment, depicted in Figure 5(b) and Figure 6(b), are more neat (accurate) than those depicted in Figure 5(a) and Figure 6(a), respectively. This is reasonable, because the global alignment for compositing a good mosaic includes both the topology analysis and the global optimization while the topology estimation just aims at finding the topology in an efficient way.

3.2. Evaluation on Initial Model Selection

In the period of recovering initial alignment described in Section 2.3.1, the selection of the transformation model among *rigid*, *affine* and *homography* models can make differences to the final mosaicking result. To amplify the influence of error factors, we

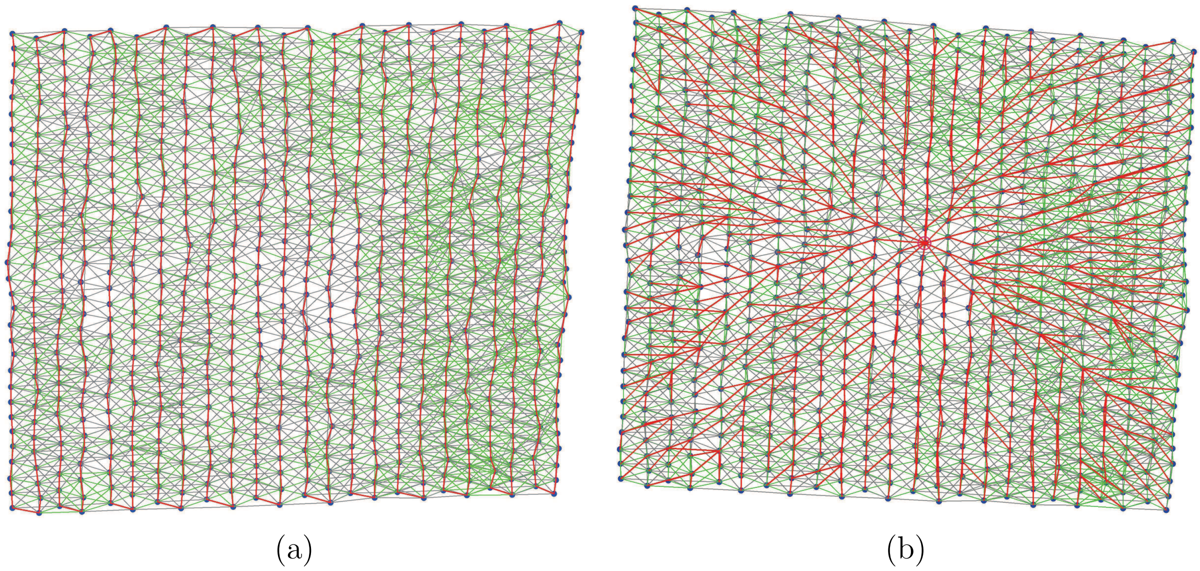


Figure 5: The estimated topology of the first dataset (744 images) highlighted for different aims: (a) The estimated topology with the prior main chain in the time-consecutive order marked with red edges; (b) The spinning tree generated by searching for the optimal reference image (the node with red ring), marked with red edges on the estimated topological graph. Different from (a), the geometric positions in (b) were recovered by the final global alignment. The edges in green and gray indicate the numbers of matched features between image pairs more and less than 100, respectively.

Table 4: Root-Mean-Square (RMS) errors through selecting different transformation models for initial alignment in the proposed approach (GR: Global Refinement; Unit: pixel).

Models	Strip Aerial Images			Block UAV Images		
	#Matches	RMS	RMS (GR)	#Matches	RMS	RMS (GR)
Rigid	131279	3.142	1.247	48783	5.112	1.985
Affine	131279	2.825	1.117	48783	4.421	1.743
Homography	131279	2.459	0.808	48783	3.605	1.485

specially selected a strip-shaped aerial image subset and a block UAV image subset from the first dataset and the second one, respectively, and the image on the end was set as the reference image. The comparative analyses were made on both alignment precision and global consistency, where the numerical results are shown in Table 4 while the global consistency can be judged via the visual results shown in Figure 7.

As for the strip aerial images, the homography model employed as the initial model has the best alignment precision, but suffers severe an accumulation of the perspective

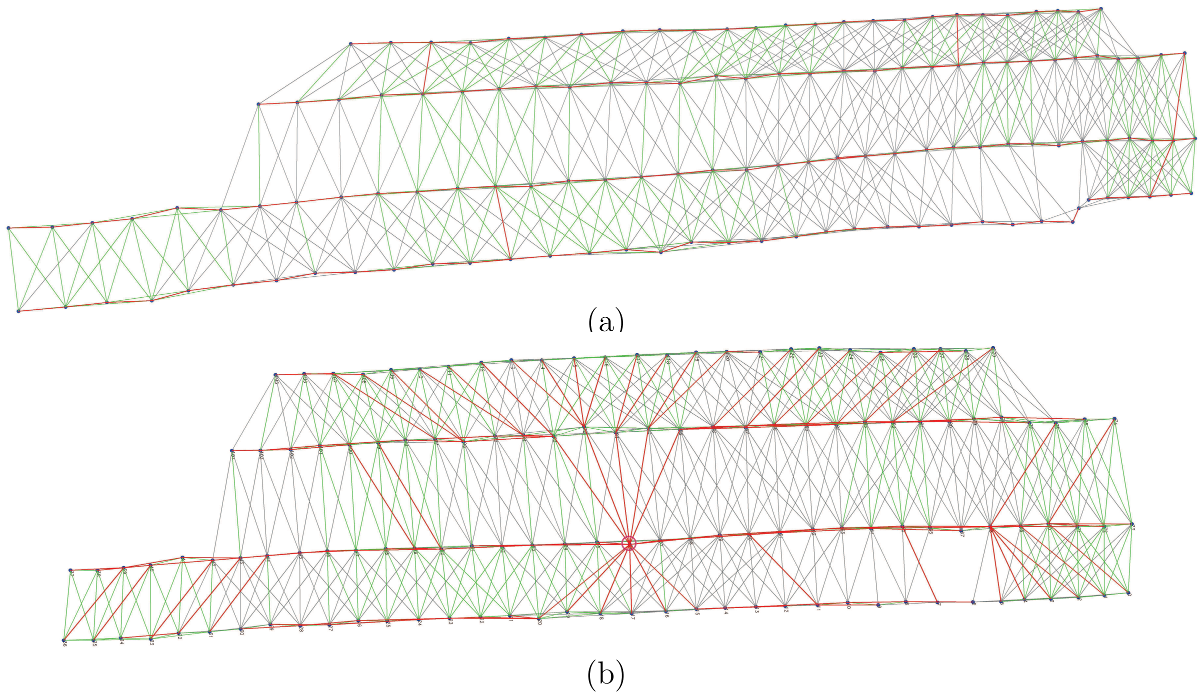


Figure 6: The estimated topology of the second dataset (130 images) highlighted for different aims: (a) The estimated topology with the main chain, linked by the proposed iterative scheme, labeled in red; (b) The spinning tree generated by searching for the optimal reference image (the node with red ring), marked with red edges on the estimated topological graph. Different from (a), the geometric positions in (b) were recovered by the final global alignment. The edges in green and gray indicate the numbers of matched features between image pairs more and less than 100, respectively.

distortions meanwhile due to that it has the highest DoF for alignment. However, the mosaicking result based on the rigid transformation as the initial model shows a bending tendency with the lowest accuracy although it doesn't induce a severe perspective distortion. This is because the rigid model of 3 free parameters just allows the image translation and rotation, which is not enough to describe the truly geometric relations between images and prone to resulting in the accumulation of rotation or translation. Compromising between them, the affine model with a moderate DoF, has made a good balance between the aligning accuracy and the global consistency, which gives the most visually satisfactory mosaicking result.

Because of the low flight altitude, the comparatively large-depth-difference ground greatly decreases the aligning precision for the UAV image sequence. In this case, the perspective distortion is still noticeable for the mosaicking result of the homography

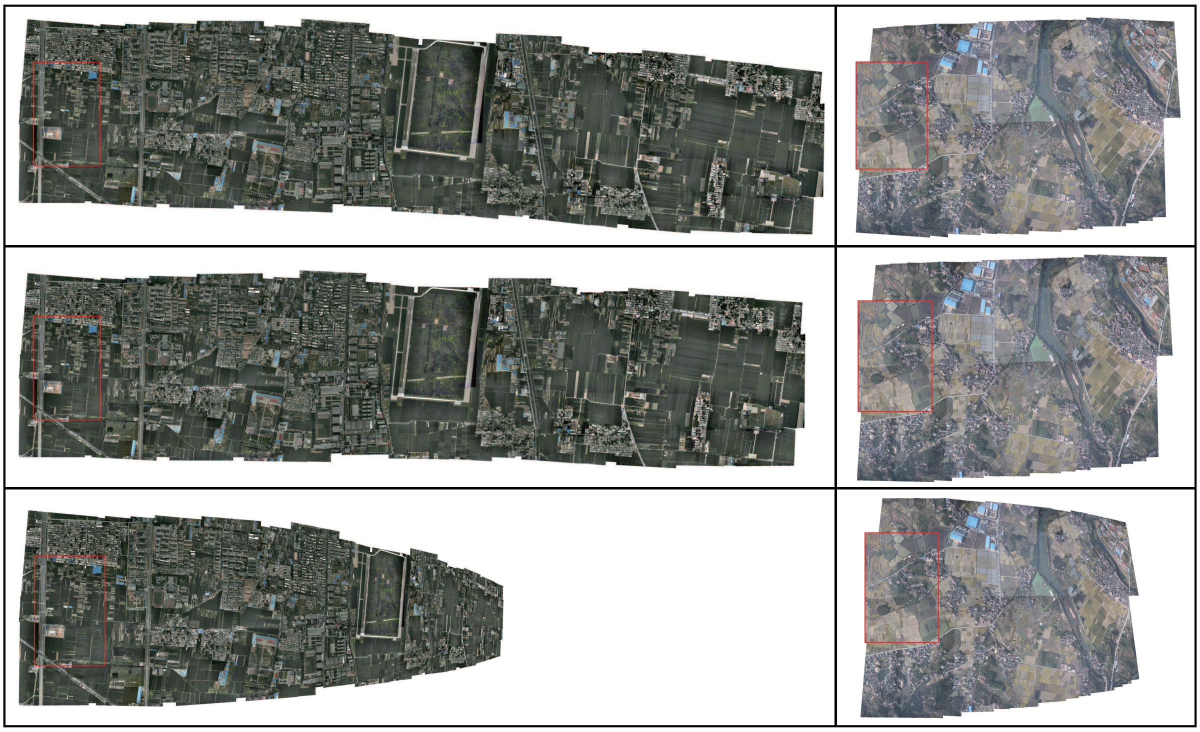


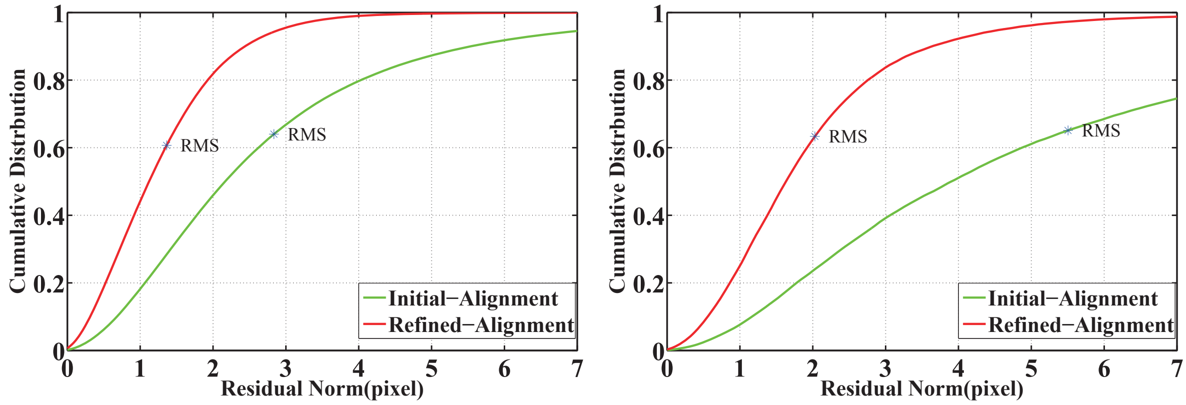
Figure 7: The thumbnails of the mosaicking results on the aerial images (Left) and the UAV images (Right) where the rigid model in the first row, the affine model in the second row, and the homography model in the last row were chosen for initial alignment, respectively. Notice that the reference image of each mosaic is marked with a red rectangular box.

model, even though the images were taken from a small block area. Differently, the rigid model achieves an as good visual result as the affine model does for this UAV image sequence, though its aligning precision is a little inferior to that of the affine model. Conclusively, the affine model has the best comprehensive property to provide a robust initial alignment, so it's the most reasonable choice of the initial aligning model in our approach.

3.3. Comprehensive Evaluation on Mosaicking Results

The final mosaicking results of our approach were evaluated in both qualitative and quantitative forms. Firstly, we compared the mosaicked images generated by our approach with those created by a commercial software named PTGui [1] on visual effects. Since aiming at comparing the alignment results only, the following seamline detection and

[1] http://www.ptgui.com/



(a) (b)

Figure 8: Cumulative probability distributions of the residual error norms with and without the global refinement performed in our approach: (a) the error analysis for the first dataset; (b) the error analysis for the second dataset. The green curves depict the aligning error of our approach with only the initial alignment, while the red curves represent the aligning error of our approach with the fullset of alignment. The blue marks on curves indicate the RMS errors.

tonal correction were skipped in PTGui and our image stacking order was made consistent with that of PTGui. The comparative results of the first and the second datasets are illustrated in Figure 9 and Figure 10, respectively.

From the mosaics shown in Figure 9, the two mosaics have similar visual effects as a whole, both of which take on a pretty good global consistency. However, when it comes to the local aligning accuracy, our approach has an obvious superiority over PTGui, which can be observed from some enlarged regions listed in the right column of Figure 9. As for the UAV data, the large-depth-difference ground makes the assumption of planarity of the scene weaker, which increases the difficulty to keep the global consistency. A slightly down-scale tendency in the left part can be found in the mosaicking result of our approach in Figure 10(a). Since some strong constraints were employed for keeping the scale of each image consistent, the mosaicking result of PTGui nearly suffered no perspective distortions, but in the mean while, the alignment precision was destroyed greatly. For a detail comparison, a serial of enlarged typical regions are listed in the middle line of Figure 10, which illustrate the excellent performance of our approach in the aspect of aligning accuracy.

Without precision analysis in PTGui, the quantitative evaluation of our approach

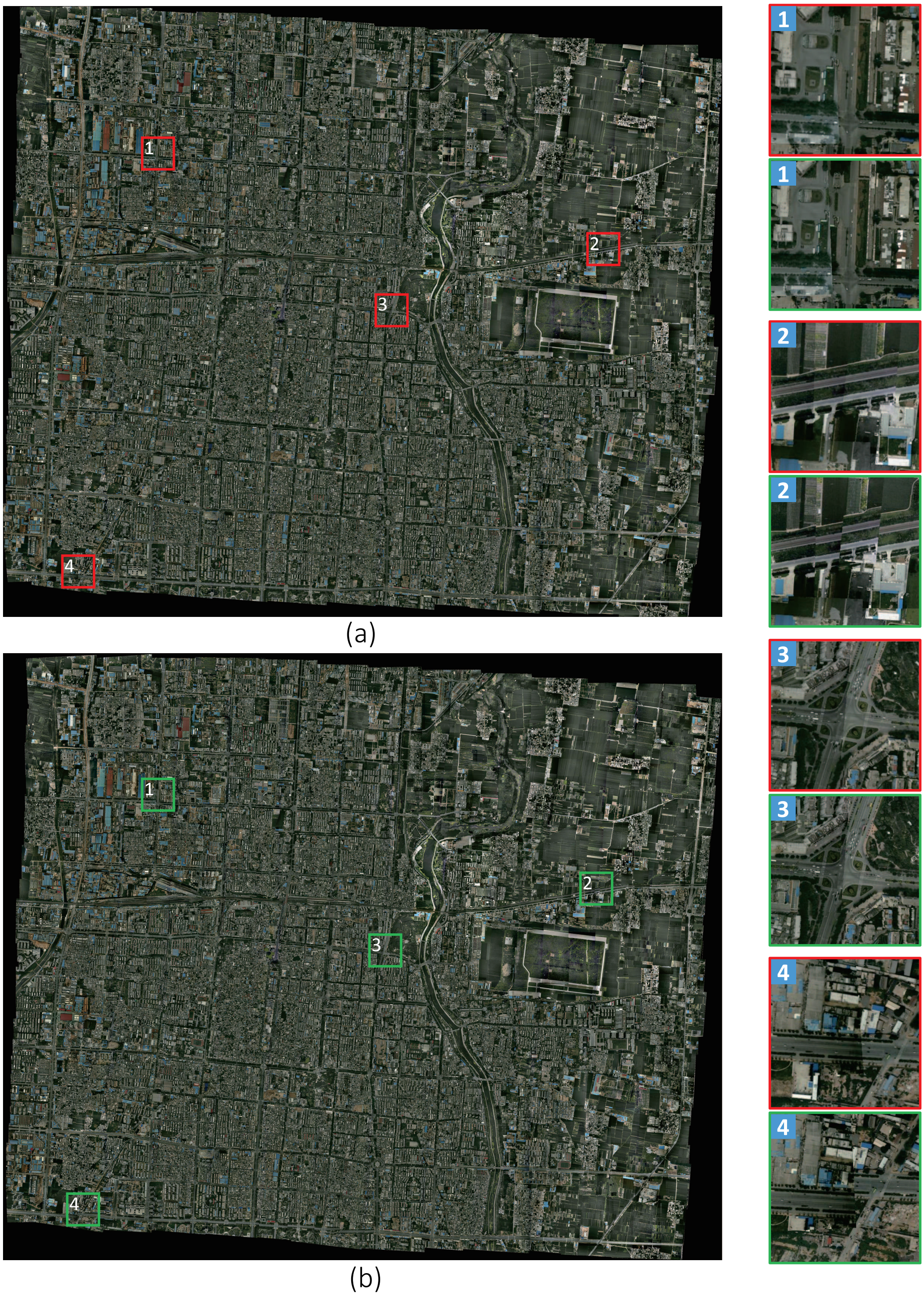


Figure 9: The mosaics composited from the first dataset (744 images) by: (a) our approach and (b) PTGui, respectively. Several typical regions grabbed from the mosaics are enlarged in pairs in the right column.

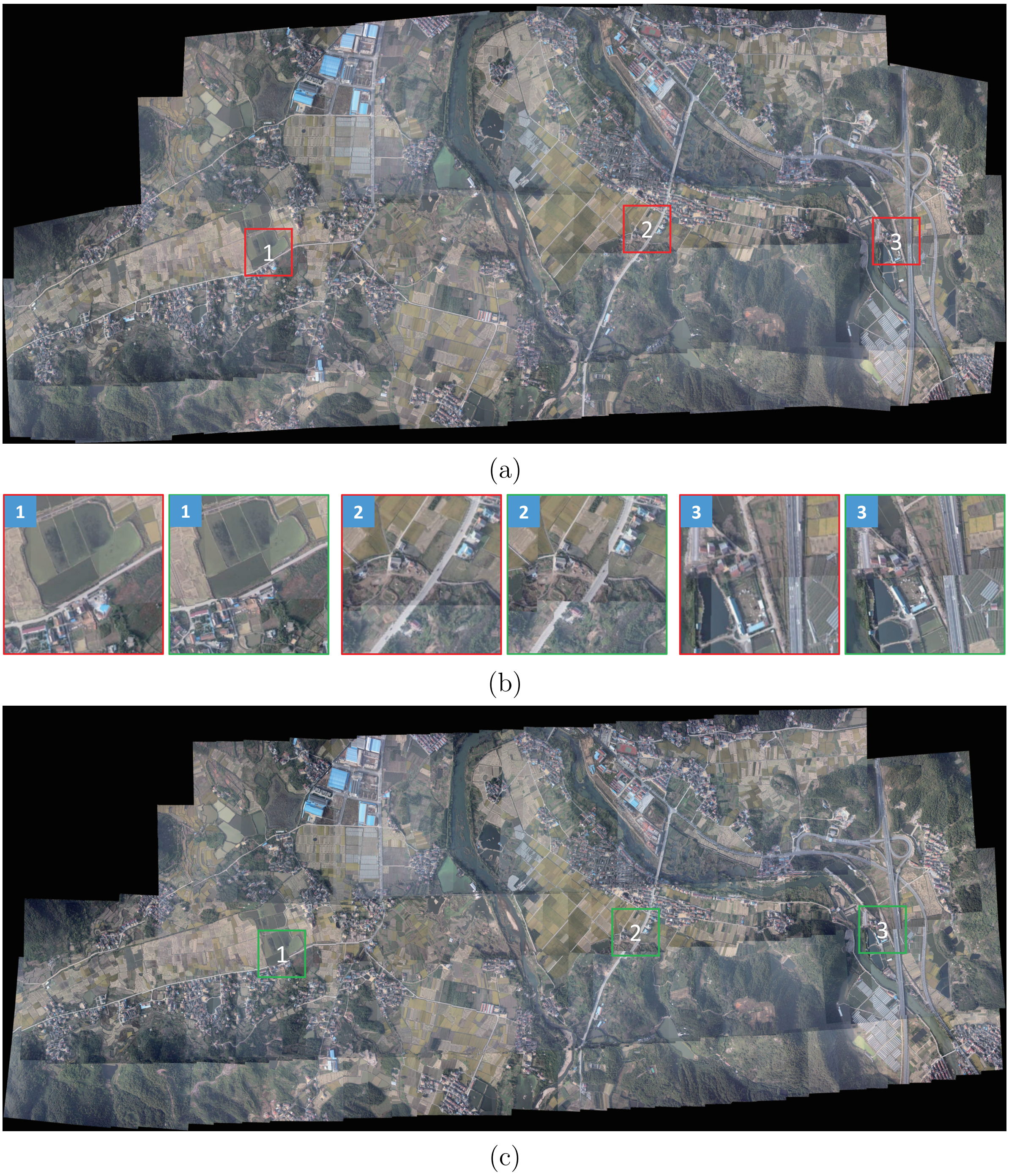


Figure 10: The mosaics composited from the second dataset (130 images) by: (a) our approach and (c) PTGui, respectively. Several typical regions grabbed from the mosaics are enlarged in pairs in (b).

474 was performed in two aspects. As an alignment precision, the registration error of our approach, running with the initial alignment only and the full set of global alignment, respectively, were compared in the form of cumulative probability distribution, as

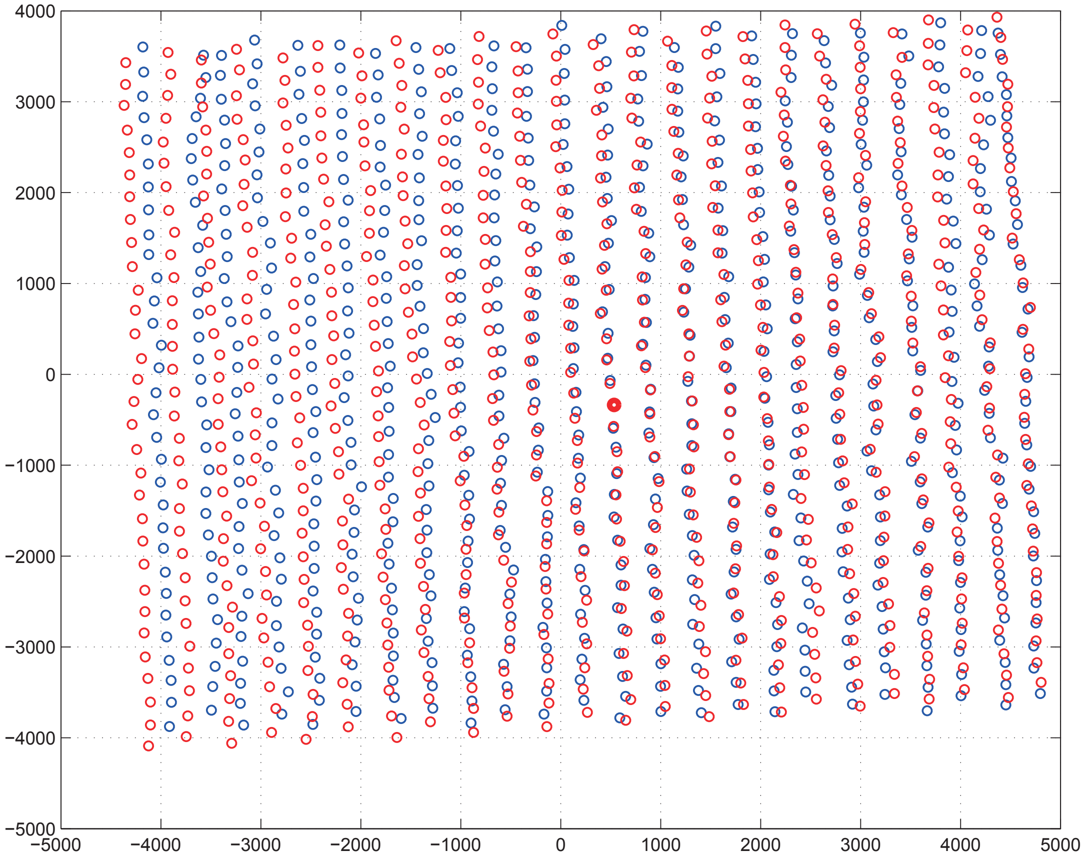


Figure 11: Distributions of image centroids on the mosaic computed by two different approaches. The red circles are the centroids recovered by the proposed approach, and the blue ones represent the result of the pose-based approach. The solid red circle stands for the centroid of the reference image, from which different groups are strictly superimposed as a base point.

displayed in Figure 8. From the comparisons, it's easy to find that the aligning precision increases a lot with the help of the homography refinement, while the global consistency is not affected during the transition from the affine model to the homography one, as can be observed in Figure 9(a) and Figure 10(a). This is what we try to achieve, namely to keep an optimal balance between the alignment accuracy and the global consistency.

Moreover, the available poses of the first dataset, which were recovered by the rigid block adjustment of photogrammetry field, were used to calculate the homography models according to the formula in [4] under the assumption of the ground being a plane. Considering the pesudo-planarity of the ground scene, they are not accurate enough to be used as the ground truth, but they are qualified to evaluate the global consistency as a reference, since the pose parameters can be regarded as no accumulation error. The recovered image centroids of the first dataset obtained by our approach and the reference models, are illustrated in Figure 11. It shows that the two groups of centroids have a similar distribution form but there are also some displacements between corresponding centroids which average at 5.16 pixels. Obviously, the displacements in the right part are much smaller than those in the left part, because the right part has more dense and strong topological relationships suppressing accumulation errors. In fact, as an image mosaicking approach based on the 2D feature registration, the recovered geometric positions are accurate enough to keep the global consistency of a mosaic, which emphasizes more on the visual effects than the geometric measurements. What's more, because of no image registration based optimization performed, the pose-based approach obtains a terrible image aligning accuracy as the RMS error of 103.9 pixels, which is much inferior to that of 1.36 pixels in our approach. Therefore, our approach has a good property of alignment accuracy and global consistency in the terms of image mosaicking.

4. Conclusion and Future Works

In this paper, a topology analysis based generic framework was proposed for mosaicking sequential images of an approximately planar scene, which is composed of three steps : topology estimation, reference image selection, and global alignment. Specifically, it's adapted to both ordered and unordered image sequences. To estimate the topology robustly, we perform the image location and the potential overlapping pairs detection in a collaborative way, which results in that our approach for topology estimation significantly outperformed the state-of-the-art method in the aspect of efficiency. Based on the topological graph, the optimal reference image is found by graph analysis and all the images are organized as a spinning tree which gives the reference relationships for each image. With the result of topological analysis, we propose a global alignment strategy of allowing the continuous transition between the affine model and the homography one according to the energy definition, which can keep the optimal balance between the global consistency and the aligning accuracy adaptively. The proposed framework was tested with several datasets and the experimental results illustrate the superiority of our approaches. However, as stated in this paper, the global consistency and the alignment accuracy need to be treated in a trade-off way in the case of pseudo-plane. Therefore, the ideal following process of this would be optimal seamline selection, which removes the residual parallax by crossing areas with less misalignment. This is meaningful for compositing a mosaicked image of high quality, and it will be studied in the future work.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Project No. 41571436), the Hubei Province Science and Technology Support Program, China (Project No. 2015BAA027), the National Natural Science Foundation of China (Project No. 41271431), and the Jiangsu Province Science and Technology Support Program, China (Project No. BE2014866).

Appendix A. Optimization Derivation for Model Refinement Under Anti-Perspective Constraint

All the terms in the energy definition in Eq. (8) for model refinement under anti-perspective constraint are quadratic, which need to be linearized by the Taylor expansion for the iterative optimization. Generally, the first-order Taylor series expansion is accurate enough for the optimization problem of quadratic functions.

Here, we define the parameter vector of the homography matrix $\mathbf{H}_i$ as $\theta_i = [h_1^i, h_2^i, h_3^i, h_4^i, h_5^i, h_6^i, h_7^i, h_8^i]^\top, i \in [1, n]$, and the initial value of θ_i is defined as $\bar{\theta}_i = [\bar{h}_1^i, \bar{h}_2^i, \bar{h}_3^i, \bar{h}_4^i, \bar{h}_5^i, \bar{h}_6^i, \bar{h}_7^i, \bar{h}_8^i]^\top$. Taking a pair of matching points $\{\varpi(\mathbf{x}_{ij}^k) = (x, y), \varpi(\mathbf{x}_{ji}^k) = (x', y')\}$ from $\mathbf{I}_i$ and $\mathbf{I}_j$ for example, Eq. (8) can be written as :

$$\begin{aligned} f_k = &\left(\frac{h_1^i x + h_2^i y + h_3^i}{h_7^i x + h_8^i y + 1} - \frac{h_1^j x' + h_2^j y' + h_3^j}{h_7^j x' + h_8^j y' + 1}\right)^2 + \left(\frac{h_4^i x + h_5^i y + h_6^i}{h_7^i x + h_8^i y + 1} - \frac{h_4^j x' + h_5^j y' + h_6^j}{h_7^j x' + h_8^j y' + 1}\right)^2 \\ &+ \lambda \left[\left(\frac{h_1^i x + h_2^i y + h_3^i}{h_7^i x + h_8^i y + 1} - x_0\right)^2 + \left(\frac{h_1^j x' + h_2^j y' + h_3^j}{h_7^j x' + h_8^j y' + 1} - x'_0\right)^2 \right. \\ &\left. + \left(\frac{h_4^i x + h_5^i y + h_6^i}{h_7^i x + h_8^i y + 1} - y_0\right)^2 + \left(\frac{h_4^j x' + h_5^j y' + h_6^j}{h_7^j x' + h_8^j y' + 1} - y'_0\right)^2 \right], \end{aligned} \quad \text{(A.1)}$$

where $[x_0, y_0]^\top = \varpi(\mathbf{A_i}\mathbf{x}_{ij}^k)$ and $[x'_0, y'_0]^\top = \varpi(\mathbf{A_j}\mathbf{x}_{ji}^k)$, are the constant terms which can be calculated in advance. Eq. (A.1) is expanded in the form of the first-order Taylor series as:

$$
\begin{aligned}
f_k \approx \bar{f}_k &+ \frac{\partial f_k}{\partial h_1^k} dh_1^i + \frac{\partial f_k}{\partial h_2^i} dh_2^i + \frac{\partial f_k}{\partial h_3^i} dh_3^i + \frac{\partial f_k}{\partial h_4^i} dh_4^i + \frac{\partial f_k}{\partial h_5^i} dh_5^i + \frac{\partial f_k}{\partial h_6^i} dh_6^i + \frac{\partial f_k}{\partial h_7^i} dh_7^i + \frac{\partial f_k}{\partial h_8^i} dh_8^i \\
&+ \frac{\partial f_k}{\partial h_1^j} dh_1^j + \frac{\partial f_k}{\partial h_2^j} dh_2^j + \frac{\partial f_k}{\partial h_3^j} dh_3^j + \frac{\partial f_k}{\partial h_4^j} dh_4^j + \frac{\partial f_k}{\partial h_5^j} dh_5^j + \frac{\partial f_k}{\partial h_6^j} dh_6^j + \frac{\partial f_k}{\partial h_7^j} dh_7^j + \frac{\partial f_k}{\partial h_8^j} dh_8^j,
\end{aligned}
\tag{A.2}
$$

where $\bar{f}_k$ is the values of f_k when substituting $\bar{\theta}_i$ and $\bar{\theta}_j$ into Eq. (A.1). $d\theta_i = [dh_1^i, dh_2^i, dh_3^i, dh_4^i, dh_5^i, dh_6^i, dh_7^i, dh_8^i]^\top$ represents the delta value of $\theta_i, i \in [1, n]$. The partial derivatives of functions f_k with respect to θ_i and θ_j are listed as below:

$$
\begin{cases}
\dfrac{\partial f_k}{\partial h_1^i} = \dfrac{K_1 x}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \quad \dfrac{\partial f_k}{\partial h_2^i} = \dfrac{K_1 y}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \quad \dfrac{\partial f_k}{\partial h_3^i} = \dfrac{K_1}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \\
\dfrac{\partial f_k}{\partial h_4^i} = \dfrac{K_2 x}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \quad \dfrac{\partial f_k}{\partial h_5^i} = \dfrac{K_2 y}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \quad \dfrac{\partial f_k}{\partial h_6^i} = \dfrac{K_2}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \\
\dfrac{\partial f_k}{\partial h_7^i} = \dfrac{-K_1(\bar{h}_1^i x + \bar{h}_2^i y + \bar{h}_3^i)x}{(\bar{h}_7^i x + \bar{h}_8^i y + 1)^2} + \dfrac{-K_2(\bar{h}_3^i x + \bar{h}_4^i y + \bar{h}_5^i)x}{(\bar{h}_7^i x + \bar{h}_8^i y + 1)^2}, \\
\dfrac{\partial f_k}{\partial h_8^i} = \dfrac{-K_1(\bar{h}_1^i x + \bar{h}_2^i y + \bar{h}_3^i)y}{(\bar{h}_7^i x + \bar{h}_8^i y + 1)^2} + \dfrac{-K_2(\bar{h}_3^i x + \bar{h}_4^i y + \bar{h}_5^i)y}{(\bar{h}_7^i x + \bar{h}_8^i y + 1)^2}, \\
\dfrac{\partial f_k}{\partial h_1^j} = \dfrac{K_3 x}{\bar{h}_7^j x + \bar{h}_8^j y + 1}, \quad \dfrac{\partial f_k}{\partial h_2^j} = \dfrac{K_3 y}{\bar{h}_7^j x + \bar{h}_8^j y + 1}, \quad \dfrac{\partial f_k}{\partial h_3^j} = \dfrac{K_3}{\bar{h}_7^j x + \bar{h}_8^j y + 1}, \\
\dfrac{\partial f_k}{\partial h_4^j} = \dfrac{K_4 x}{\bar{h}_7^j x + \bar{h}_8^j y + 1}, \quad \dfrac{\partial f_k}{\partial h_5^j} = \dfrac{K_4 y}{\bar{h}_7^j x + \bar{h}_8^j y + 1}, \quad \dfrac{\partial f_k}{\partial h_6^j} = \dfrac{K_4}{\bar{h}_7^i x + \bar{h}_8^i y + 1}, \\
\dfrac{\partial f_k}{\partial h_7^j} = \dfrac{-K_3(\bar{h}_1^j x + \bar{h}_2^j y + \bar{h}_3^j)x}{(\bar{h}_7^j x + \bar{h}_8^j y + 1)^2} + \dfrac{-K_4(\bar{h}_3^j x + \bar{h}_4^j y + \bar{h}_5^j)x}{(\bar{h}_7^j x + \bar{h}_8^j y + 1)^2}, \\
\dfrac{\partial f_k}{\partial h_8^j} = \dfrac{-K_3(\bar{h}_1^j x + \bar{h}_2^j y + \bar{h}_3^j)y}{(\bar{h}_7^j x + \bar{h}_8^j y + 1)^2} + \dfrac{-K_4(\bar{h}_3^j x + \bar{h}_4^j y + \bar{h}_5^j)y}{(\bar{h}_7^j x + \bar{h}_8^j y + 1)^2},
\end{cases}
$$

where K_1, K_2, K_3, and K_4 are computed as:

$$
\begin{cases}
K_1 = \dfrac{2(\bar{h}_1^i x + \bar{h}_2^i y + h_3^i)}{\bar{h}_7^i x + h_8^i y + 1} - \dfrac{2(\bar{h}_1^j x' + \bar{h}_2^j y' + \bar{h}_3^j)}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} + 2\lambda(\dfrac{\bar{h}_1^i x + \bar{h}_2^i y + \bar{h}_3^i}{\bar{h}_7^i x + \bar{h}_8^i y + 1} - x_0), \\
K_2 = \dfrac{2(\bar{h}_4^i x + \bar{h}_5^i y + \bar{h}_6^i)}{\bar{h}_7^i x + \bar{h}_8^i y + 1} - \dfrac{2(\bar{h}_4^j x' + \bar{h}_5^j y' + \bar{h}_6^j)}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} + 2\lambda(\dfrac{\bar{h}_4^i x + \bar{h}_5^i y + \bar{h}_6^i}{\bar{h}_7^i x + \bar{h}_8^i y + 1} - y_0), \\
K_3 = -\dfrac{2(\bar{h}_1^i x + \bar{h}_2^i y + \bar{h}_3^i)}{\bar{h}_7^i x + \bar{h}_8^i y + 1} + \dfrac{2(\bar{h}_1^j x' + \bar{h}_2^j y' + \bar{h}_3^j)}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} + 2\lambda(\dfrac{\bar{h}_1^j x' + \bar{h}_2^j y + \bar{h}_3^j}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} - x'_0), \\
K_4 = -\dfrac{2(\bar{h}_4^i x + \bar{h}_5^i y + \bar{h}_6^i)}{\bar{h}_7^i x + \bar{h}_8^i y + 1} + \dfrac{2(\bar{h}_4^j x' + \bar{h}_5^j y' + \bar{h}_6^j)}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} + 2\lambda(\dfrac{\bar{h}_4^j x' + \bar{h}_5^j y' + \bar{h}_6^i}{\bar{h}_7^j x' + \bar{h}_8^j y' + 1} - y'_0).
\end{cases}
$$

For the convenience of descriptions in the following, the matrix form of Eq. (A.2) are written as the standard equation of the Least Square optimization:

$$[v_k] = \begin{bmatrix} \cdots & \frac{\partial f_k}{\partial h_1^i} & \frac{\partial f_k}{\partial h_2^i} & \frac{\partial f_k}{\partial h_3^i} & \frac{\partial f_k}{\partial h_4^i} & \frac{\partial f_k}{\partial h_5^i} & \frac{\partial f_k}{\partial h_6^i} & \frac{\partial f_k}{\partial h_7^i} & \frac{\partial f_k}{\partial h_8^i} & \cdots \\ \cdots & \frac{\partial f_i}{\partial h_1^j} & \frac{\partial f_i}{\partial h_2^j} & \frac{\partial f_k}{\partial h_3^j} & \frac{\partial f_k}{\partial h_4^j} & \frac{\partial f_k}{\partial h_5^j} & \frac{\partial f_k}{\partial h_6^j} & \frac{\partial f_k}{\partial h_7^j} & \frac{\partial f_k}{\partial h_8^j} & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ d\theta_i \\ \vdots \\ d\theta_j \\ \vdots \end{bmatrix} - \left[-\bar{f}_k\right].$$

The above equation is expressed with the corresponding matrix labels as:

$$\mathbf{V}^k = \mathbf{J}^k\mathbf{X} - \mathbf{L}^k, \tag{A.3}$$

where the dots in the Jacobi matrix $\mathbf{J}^k$ represent a series of zeros, and the dots in $\mathbf{X}$ indicate the other unknown parameters in $\{d\theta_i\}_{i=1}^n$. $\mathbf{V^k}$ is the residual error of a pair of matching points. Hereafter, we name $\mathbf{J}^k$ and $\mathbf{L}^k$ as the coefficient matrix and the constant matrix, respectively.

As can be seen, a pair of matching points from two images provides an equation with 16 unknown parameters. Supposing that n images have m pairs of overlapping relations and there are s matching points of each image pair in average, then we obtain a Jacobi matrix with the size of $m \times s$ rows and $8 \times m$ columns and a constant matrix with the size of $m \times s$ rows and 1 column. In each iteration, $\mathbf{J}_{ms\times 8n}$ and $\mathbf{L}_{ms\times 1}$ have to be recalculated and the corresponding solution vector $\mathbf{X}_{8n\times 1} = [d\theta_1^\top, ..., d\theta_n^\top]^\top$ can be solved with the following equation:

$$\mathbf{X}_{8n\times 1} = (\mathbf{J}_{ms\times 8n}^\top \mathbf{J}_{ms\times 8n})^{-1}(\mathbf{J}_{ms\times 8n}^\top \mathbf{L}_{ms\times 1}). \tag{A.4}$$

The initial solution of $\{\theta_i\}_{i=1}^n$ for next iteration is updated by adding up $\mathbf{X}_{8n\times 1}$ and the initial solution used in this iteration. As the iteration goes, the updated solution will converge to the optimal solution gradually unless the initial solution provided at the very beginning is not accurate enough. However, when the amount of images is large, the size of the Jacobi matrix will be very huge and makes a challenge to the memory of computer.

In fact, we can calculate $\{\theta_i\}_{i=1}^n$ directly if $\mathbf{TJ}_{8n\times 8n} = \mathbf{J}_{ms\times 8n}^\top \mathbf{J}_{ms\times 8n}$ and $\mathbf{TL}_{8n\times 1} = \mathbf{J}_{ms\times 8n}^\top \mathbf{L}_{ms\times 1}$ have been obtained. So, to reduce the required memory space and the computation time, we manage to compute $\mathbf{TJ}_{8n\times 8n}$ and $\mathbf{TL}_{8n\times 1}$ by adding up the matrix ${\mathbf{J}^i}^\top \mathbf{J}^i$ and the matrix ${\mathbf{J}^i}^\top \mathbf{L}^i$ calculated from each pair of matching points, instead of building the large $\mathbf{J}$ and $\mathbf{L}$ beforehand. The improved computation formula is defined as:

$$\begin{cases} \mathbf{TJ}_{8n\times 8n} = \sum_{i=1}^{ms} {\mathbf{J}^i}^\top_{1\times 8n} \mathbf{J}^i_{1\times 8n}, \\ \mathbf{TL}_{8n\times 1} = \sum_{i=1}^{ms} {\mathbf{J}^i}^\top_{1\times 8n} \mathbf{L}^i_{2\times 1}. \end{cases} \tag{A.5}$$

Then, the solution can be obtained in this way as:

$$\mathbf{X}_{8n\times 1} = \mathbf{TJ}^{-1}_{8n\times 8n} \mathbf{TL}_{8n\times 1}. \tag{A.6}$$

What's more, considering the sparsity of $\mathbf{J}^i$, the computation of the matrix multiplication in Eq. (A.5) can be improved further in the complexity of both time and space.

References

[1] E. Zagrouba, W. Barhoumi, S. Amri, An efficient image-mosaicing method based on multifeature matching, Machine Vision and Applications 20 (3) (2009) 139–162.

[2] J. Chen, H. Feng, K. Pan, Z. Xu, Q. Li, An optimization method for registration and mosaicking of remote sensing images, Optik - International Journal for Light and Electron Optics 125 (2) (2014) 697–703.

[3] B. Zitova, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (11) (2003) 977–1000.

[4] L. Kang, L. Wu, Y. Wei, B. Yang, H. Song, A highly accurate dense approach for homography estimation using modified differential evolution, Engineering Applications of Artificial Intelligence 31 (4) (2014) 68–77.

[5] Z. Wang, Y. Chen, Z. Zhu, W. Zhao, An automatic panoramic image mosaic method based on graph model, Multimedia Tools and Applications 75 (5) (2015) 2725–2740.

[6] N. R. Gracias, S. Van Der Zwaan, A. Bernardino, S.-V. Jos, Mosaic-based navigation for autonomous underwater vehicles, IEEE Journal of Oceanic Engineering 28 (4) (2003) 609–624.

[7] Y. Xu, J. Ou, H. He, X. Zhang, J. Mills, Mosaicking of unmanned aerial vehicle imagery in the absence of camera poses, 2016 8 (3) (2016) 204.

[8] Y. He, R. Chung, Image mosaicking for polyhedral scene and in particular singly visible surfaces, Pattern Recognition 41 (3) (2008) 1200–1213.

[9] J. Zaragoza, T.-J. Chin, M. Brown, D. Suter, As-projective-as-possible image stitching with moving DLT, IEEE Trasactions on Pattern Analysis and Machine Intelligence 36 (7) (2014) 1285–1298.

[10] C.-H. Chang, Y. Sato, Y.-Y. Chuang, Shape-preserving half-projective warps for image stitching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3254–3261.

[11] D. Patidar, A. Jain, Automatic image mosaicing: an approach based on FFT, International Journal of Scientific Engineering and Technology 1 (1) (2011) 01–04.

[12] S. Ghannam, A. L. Abbott, Cross correlation versus mutual information for image mosaicing, International Journal of Advanced Computer Science and Applications 4 (11) (2013) 94–102.

[13] A. Elibol, R. Garcia, O. Delaunoy, N. Gracias, Efficient Topology Estimation for Large Scale Optical Mapping, Springer, 2013, Ch. A New Global Alignment Method for Feature Based Image Mosaicing, pp. 25–39.

[14] S. Ali, C. Daul, E. Galbrun, Guillemin, Anisotropic motion estimation on edge preserving Riesz wavelets for robust video mosaicing, Pattern Recognition 51 (2016) 425–442.

[15] Y. Chen, J. Sun, G. Wang, Minimizing geometric distance by iterative linear optimization, in: IEEE International Conference on Pattern Recognition, Vol. 29, 2010, pp. 1–4.

[16] W. Mou, H. Wang, G. Seet, L. Zhou, Robust homography estimation based on nonlinear least squares optimization, Mathematical Problems in Engineering 2014 (1) (2013) 372–377.

[17] L. Zhang, Y. Li, J. Zhang, Y. Hu, Homography estimation in omnidirectional vision under the L_∞-norm, in: IEEE International Conference on Robotics and Biomimetics, 2010, pp. 1468–1473.

[18] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon, Vision Algorithms: Theory and Practice, Lecture Notes in Computer Science, Springer, 2000, Ch. Bundle adjustmenta modern synthesis, pp. 298–372.

[19] K. Konolige, Sparse sparse bundle adjustment, in: British Machine Vision Conference, 2010, pp. 1–10.

[20] M. Li, D. Li, D. Fan, A study on automatic UAV image mosaic method for paroxysmal disaster, in: Proceedings of the International Society of Photogrammetry and Remote Sensing Congress, 2012.

[21] C. Xing, J. Wang, Y. Xu, A robust method for mosaicking sequence images obtained from UAV, in: International Conference on Information Engineering and Computer Science (ICIECS), 2010.

[22] F. Caballero, L. Merino, J. Ferruz, A. Ollero, Homography based Kalman filter for mosaic building. applications to UAV position estimation, in: IEEE International Conference on Robotics and Automation, 2007, pp. 2004–2009.

[23] A. Y. Taygun Kekec, M. Unel, A new approach to real-time mosaicing of aerial images, Robotics and Autonomous Systems 62 (12) (2014) 1755–1767.

[24] E.-Y. Kang, I. Cohen, G. Medioni, A graph-based global registration for 2D mosaics, in: International Conference on Pattern Recognition, Vol. 1, 2000, pp. 257–260.

[25] R. Marzotto, A. Fusiello, V. Murino, High resolution video mosaicing with global alignment, in: IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, 2004, pp. 692–698.

[26] A. Elibol, R. Garcia R Elibol A, Gracias Na, O. Delaunoy, N. Gracias, A new global alignment method for feature based image mosaicing, Advances in Visual Computing, Lecture Notes in Computer Science 5359 (7) (2008) 257–266.

[27] M. Xia, J. Yao, L. Li, X. Lu, Globally consistent alignment for mosaicking aerial images, in: IEEE International Conference on Image Processing, 2015.

[28] A. Elibol, N. Gracias, R. Garcia, Fast topology estimation for image mosaicing using adaptive information thresholding, Robotics and Autonomous systems 61 (2) (2013) 125–136.

[29] R. Szeliski, Image alignment and stitching: A tutorial, Foundations and Trends in Computer Graphics and Vision 2 (1) (2006) 1–104.

[30] T. E. Choe, I. Cohen, M. Lee, G. Medioni, Optimal global mosaic generation from retinal images, in: IEEE International Conference on Pattern Recognition, Vol. 3, 2006, pp. 681–684.

[31] B. Bollobs, Modern graph theory, Springer Science & Business Media, 2013.

[32] R. L. Graham, P. Hell, On the history of the minimum spanning tree problem, Annals of the History of Computing 7 (1) (1985) 43–57.

[33] R. W. Floyd, Algorithm 97: shortest path, Communications of the ACM 5 (6) (1962) 345–345.

[34] T. H. Cormen, Introduction to algorithms, MIT press, 2009.

[35] R. I. Hartley, In defense of the eight-point algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (6) (1997) 580–593.

[36] P. Torr, A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, Computer Vision and Image Understanding 78 (1) (2000) 138–156.

[37] M. I. A. Lourakis, Sparse non-linear least squares optimization for geometric vision, in: Computer Vision–ECCV 2010, Vol. 6312, 2010, pp. 43–56.