Robotics and Autonomous Systems 62 (2014) 1755–1767

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

A new approach to real-time mosaicing of aerial images

Taygun Kekec, Alper Yildirim, Mustafa Unel *

Sabanci University, Faculty of Engineering and Natural Sciences, Orhanli-Tuzla 34956, Istanbul, Turkey

HIGHLIGHTS

- A new real-time method to create mosaics of aerial images.
- Use of Separating Axis Theorem to detect image intersections.
- An affine refinement procedure for obtaining better global consistency.
- Linear optimization instead of computationally heavy non-linear optimization techniques.
- Experimental comparison with the state-of-the-art methods.

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form
14 July 2014
Accepted 25 July 2014
Available online 4 August 2014

Keywords:
Image mosaicing
Bundle adjustment
MLESAC
Separating axis theorem
Affine refinement
Multi-band blending
Gain compensation

ABSTRACT

We present a new image mosaicing technique that uses sequential aerial images captured from a camera and is capable of creating consistent large scale mosaics in real-time. To find the alignment of every new image, we use all the available images in the mosaic that have intersection with the new image instead of using only the previous one. To detect image intersections in an efficient manner, we utilize 'Separating Axis Theorem', a geometric tool from computer graphics which is used for collision detection. Moreover, after a certain number of images are added to the mosaic, a novel affine refinement procedure is carried out to increase global consistency. Finally, gain compensation and multi-band blending are optionally used as offline steps to compensate for photometric defects and seams caused by misregistrations. Proposed approach is tested on some public datasets and it is compared with two state-of-the-art algorithms. Results are promising and show the potential of our algorithm in various practical scenarios.

1. Introduction

Image mosaicing aims to increase visual perception by composing visual data obtained from separate images since a composite image provides richer description than individual images. Gaining and maintaining situational awareness from image mosaics is important for both civil and military applications. Inspection of the urban areas suffering from natural disasters and examination of the large plantations are possible civil areas of utilization. For military applications, image mosaicing can provide critical information about enemy activities in a broad perspective. Although there are many studies in the literature that focus on creating real-time image mosaics for different applications, there is still room for improvement due to the need for faster and more accurate mosaicing for a variety of practical scenarios.

Image mosaicing is the process of stitching many images together in order to create a larger, consistent and seamless composite image. The composite image can provide more information than spatially and temporally distinct separate images. Image mosaicing algorithms are frequently used in personal, medical and remote sensing applications. By using these algorithms, charming panoramas of natural scenes [1] can be obtained with inexpensive off-the-shelf cameras. In the context of medical imagery, mosaicing of retinal images [2] and tissues [3] produce impressive results. These algorithms are also used for creating large microscopic [4] and fingerprint imagery [5]. For remote sensing purposes, maps of an environment can be created using aerial [6] and underwater [7] images. They are also embedded as image stabilization and video compression routines [8] in video cameras and mobile platforms.

Finding the alignments of the images is the central part of all mosaicing algorithms. In literature, image alignment methods can be classified into two categories: dense and sparse methods. These are also known as direct and feature based approaches [9]. In direct approaches, whole image data is used instead of a set of sparse features extracted from the image. Within these approaches,

* Corresponding author. Tel.: +90 2164839549; fax: +90 2164839550.
E-mail addresses: ikekec@sabanciuniv.edu (T. Kekec), alperyildirim@sabanciuniv.edu (A. Yildirim), munel@sabanciuniv.edu, unel.mustafa@gmail.com (M. Unel).

http://dx.doi.org/10.1016/j.robot.2014.07.010
0921-8890/

transformation parameters and pixel correspondences are estimated simultaneously. These approaches also provide a higher accuracy when compared to the feature based approaches since all the available image information is considered in the estimation process. Although this provides more accuracy, initial estimation parameters must be in close proximity of the true solution and a high degree of overlap between the images is needed for convergence. Pioneering work in this field is done by Lucas and Kanade [10]. A nice overview on historical progress and extensions to this framework can be found in Baker's work [11].

In feature based methods, distinctive image features such as SIFT [12], SURF [13] and affine invariant regions [14] are used for the estimation of the alignment parameters. Sparse nature of the features accelerates the estimation process and eases the real-time operation.

Selecting an appropriate transformation model to compute the image alignments is an important step for image mosaicing. A hierarchy of transformations [15] exists under projectivity. Projective homography is the most general motion model for image mosaicing applications where the scene is planar and the camera undergoes a rigid motion [9]. For pure rotational camera motion, homography is a rotation matrix that has less independent parameters than a full homography and as a result, estimation becomes more stable [16,1]. However, this assumption is violated in airborne applications where parallax effects are present due to the camera translation.

Several different frameworks have been proposed to create image mosaics for various scenarios. One approach is to consider the mosaicing problem under a recursive estimation framework [17] where homography parameters are treated as the system states. Authors consider homography parameters as the system states. Whenever a loop is detected in the image sequence, an Extended Kalman Filter (EKF) is launched to tune transformation parameters through the loop. This way error is propagated through images and thus global consistency is improved. The analogy of mosaicing to Simultaneous Localization and Mapping (SLAM) problem has been noted by Civera et al. [18]. They utilize a SLAM framework for creating image mosaics in real-time. In the cited work, system states are composed of feature coordinates and the most recent pose parameters of the camera. While filter based approaches provide satisfactory results, their scalability can be questioned. For instance, in the case of large scale image mosaics which contain thousands of images, it becomes computationally infeasible to use the filtering approach since the size of the state vector grows very large.

An alternative formulation is to employ graph theory in mosaicing. Kang et al. formulate global consistency as finding optimal paths in the graph [19]. Elibol et al. utilize Minimum Spanning Tree (MST) algorithm to infer tentative topology of the mosaic with a reduced number of matching trials [20]. Choe et al. [2] focus on selecting optimal reference frame which is formulated as a shortest path problem on the graph using Floyd–Warshall algorithm. Kim and Hong [21] use sequential block matching in regularly spaced grid features. They reduce search space on the graph by using a sequential shortest-path algorithm.

In order to create globally consistent image mosaics, a nonlinear optimization algorithm, i.e. 'Bundle Adjustment' [22], can be run on the feature reprojection errors. Given a number of overlapping images, bundle adjustment aims to find parameters that minimize the total feature reprojection error. The minimization can be performed over motion parameters or structure parameters or both. Despite the fact that results can be impressive, this minimization is hard to perform in real-time. Although several variants of bundle adjustment exist and either sparsity of the structure is exploited [23,24] or multiple cores are being utilized [25], speed issues are still being investigated. This severely limits usage of bundle adjustment in robotics applications, especially for large scale data.

Image mosaicing can be easier if some prior data are used. For example, in the context of mosaicing where images are captured from a UAV, data from non-visual airborne sensors such as Inertial Measurement Unit (IMU) and GPS can be incorporated. Such sensors will allow orthorectification of the acquired imagery and limit the parameter space [26]. By narrowing the region of interest, computation time is also decreased during the matching procedure [27]. Initial works on aerial image mosaicing adopted robust model estimation techniques for feature matching such as RANSAC [28] and LMeds [29]. Various improvements have been introduced on classical RANSAC in terms of speed, accuracy and robustness. For example, RANSAC framework has been extended with various ideas such as MLE estimation [30], guided sampling procedure [31], exploitation of match similarities [32] and local optimizations [33].

In this paper, we propose a new mosaicing method for creating seamless mosaics from a set of sequential images acquired from a UAV. We are interested in the problem of creating mosaics of quasi-planar scenes. Our aim is to reach a reasonable accuracy in the mosaic without using a computationally expensive framework such as bundle adjustment. There are two main features of the proposed approach that distinguishes it from the existing methods. First, we use a tool from computer graphics literature, Separating Axis Theorem (SAT), to detect intersections between new and previous images in the mosaic where this information is used to obtain better estimates of the homographies. Theorem provides an efficient operation since it is composed of a small number of steps with basic geometric operations. To the best of our knowledge, this is the first use of SAT in mosaicing context. Second, we introduce an affine refinement procedure which enhances global consistency of the mosaic by performing a linear optimization on the transformation parameters of the recent images. This refinement is performed on a constant number of images which provides an approximately constant time operation. Although there are similar refinements in the literature, the way the cost function is formulated in the optimization problem is quite different from the existing refinement techniques. Proposed algorithm can be run on long image sequences where other frameworks are inconvenient due to their limited scalability. While there are many studies [34,35] that boost estimation process with auxiliary data, we avoid using such data and do not perform any correction with non-visual onboard sensors. This increases functionality of the algorithm since sensorial data may be inaccurate or unavailable. Our method is validated through several experiments conducted on some publicly available datasets and its performance is assessed with some of the state-of-the-art algorithms.

The organization of the paper is as follows: in Section 2, image mosaicing is described and alignment methods are presented. In Section 3, proposed mosaicing approach is detailed. In Section 4, gain compensation and multi-band blending techniques that are used to obtain visually attractive image mosaics are suggested as offline steps. In Section 5, experimental results are presented where visual and quantitative results are provided. Finally the paper is concluded with some remarks in Section 6 and future directions are indicated.

2. Image mosaicing

Image mosaicing involves transforming images captured from different camera poses as if they are taken from a single camera and registering them on an image plane (mosaic plane or reference frame). The simplest way to register sequential images acquired from a UAV is to perform homography estimations between successive images (pairwise alignment). To create the mosaic, all the images must be aligned to the reference image. Let I_r be our reference image. Given that n images $I_0, I_1, I_2, \ldots, I_{n-1}$ from a planar

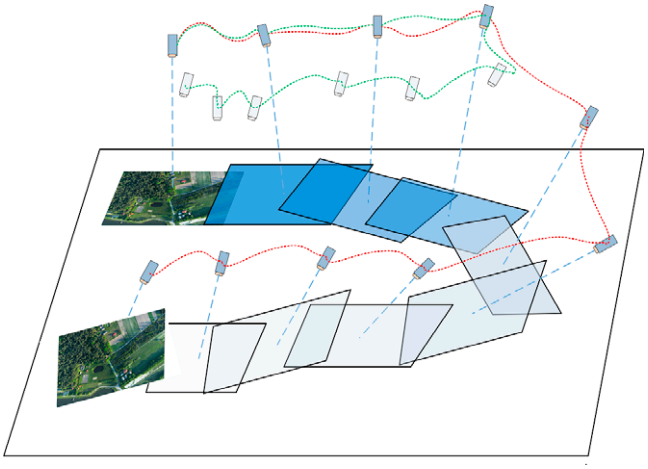


Fig. 1. Drift caused by estimation errors. UAV returns to the same area and snaps the same image from the initial position. True and estimated trajectories are shown with green and red dashed curves respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

scene and pairwise homographies $H_{01}, H_{12}, H_{23} \ldots, H_{(n-2)(n-1)}$ related to this set of images are known where H_{ij} is the transformation which aligns I_j to I_i, we can calculate the homography between a new image I_n and the reference image I_r as:

$$H_{rn} = \prod_{i=r}^{n-1} H_{i(i+1)}. \tag{1}$$

Although this approach seems to be straightforward, because of its multiplicative nature, errors tend to accumulate very fast which causes the images to drift in the mosaic. Drift of the images in the mosaic are depicted in Fig. 1.

Since a Normalized Direct Linear Transformation (NDLT) algorithm is used during the estimations of the pairwise homographies, geometric error is minimized approximately during the estimations [15]. In this case, the cost function can be given as:

$$J(H_{i(i+1)}) = \|x_i - H_{i(i+1)}x_{i+1}\|^2. \tag{2}$$

Note that error is defined on the image I_i. However, when we align I_i and I_j to the mosaic, homography between I_i and I_j will not have the minimum error property anymore since residual vectors between these images for the estimated pairwise homography are also transformed during the alignment.

An alternative approach is to estimate the homography directly between new image and the mosaic (reference image). In other words, the features of the new image I_i are extracted and matched with those of I_{i-1}. Then, matching features of image I_{i-1} are aligned to the mosaic using $H_{r,i-1}$ and estimation of H_{ri} is carried out using the aligned version of I_{i-1}. Consequently, the cost function for the estimation is modified as

$$J(H_{ri}) = \|H_{r(i-1)}x_{i-1} - H_{ri}x_i\|^2 \tag{3}$$

where x_i and x_{i-1} are matching features between I_i and I_{i-1}, respectively. Utilizing this approach is more advantageous since the estimation is realized directly on the reference image. We use this approach in our estimations. Results for pairwise and warped alignment approaches can be seen in Fig. 2 where better results are obtained with the warped alignment approach.

As all the images are aligned to a common reference frame, it can be questioned if the choice of the reference image changes the results. Since the homography maps the image coordinates of a scene point in one camera to another, coordinates in the reference frame are found by mapping the point via its global homography. As a result, it can be presumed that the image mosaic composed of the aligned images is taken by one camera which is located at the reference camera frame. For the case where the dominant plane defining the scene is not parallel to the plane of the reference image, perspective distortions may occur in the mosaic image depending on the severeness of the scenario. Distortion manifests itself as the growth or shrink of the separate images which is caused by the change of the scene depth with respect to the reference camera frame.

In our algorithm, homography estimations are also performed with respect to the reference image. This reveals a possibility that estimation quality of the homographies may depend on the reference image selection. For the case where image plane of the reference image is not parallel to the scene plane, similar to the case of separate images, feature reprojection errors also manifest growth and shrink behavior. This means feature reprojection errors of the scene points closer to the image plane will have a leverage effect on the minimization which can spoil the estimation quality. An ideal reference image should be taken perpendicular to the

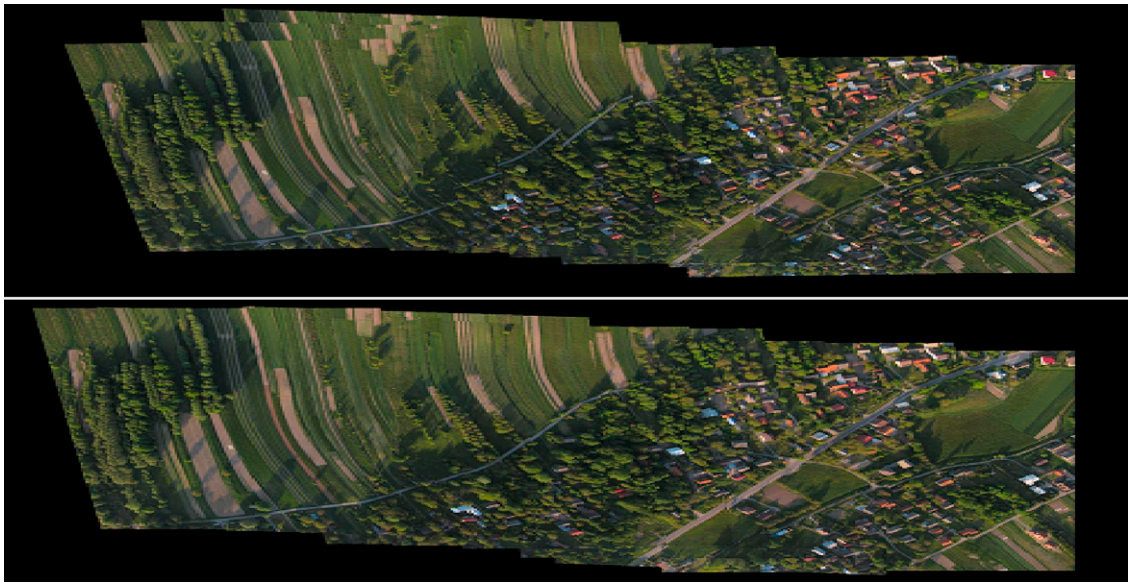


Fig. 2. Comparison of alignment using the cost function in Eq. (2) (top image) and alignment using the cost function in Eq. (3) (bottom image). Note that misregistration errors in the leftmost frames are prevented in the second case.

scene and should contain scene features which form a plane parallel to the dominant scene. Since the ground images captured from the UAVs approximately hold this condition, it does not pose a serious problem to our algorithm for generic cases. For other cases, selection of the reference image can be handled via a small external adjustment at the initialization of the algorithm if necessary.

3. Proposed mosaicing approach

The homography estimation process discussed in Section 2 includes the estimation between two images. However, estimating the homography by using only the previous image can be error-prone for mosaicing applications. For a more robust estimation, considering all of the previously aligned images which intersect the new image can be more beneficial. Since it is computationally expensive to check feature matches between the new and all of the previous images, number of these matching trials must be decreased. To this end, we propose to use a geometric tool called 'Separating Axis Theorem' to detect the previous images intersecting the new image since only aligned images intersecting each other are supposed to have common features.

3.1. Detection of image intersections by using separating axis theorem

Separating Axis Theorem (SAT) is a popular tool in computer graphics which can be used to detect collisions between objects [36]. For 2D case, theorem simply states that if there exists a line for which the intervals of projection of the two objects onto that line do not intersect, then the objects do not intersect. Such a line is called a separating line or, more commonly, a separating axis. Since translated version of a separating line is also a separating line, it is sufficient to consider the lines passing through the origin. Given a line passing through the origin and with unit-length direction $\vec{d}$, projection of a convex set C onto this line is given by the following interval:

$$[\lambda_{min}(\vec{d}), \lambda_{max}(\vec{d})] = [min\{\vec{d} \cdot \vec{X} : \vec{X} \in C\}, max\{\vec{d} \cdot \vec{X} : \vec{X} \in C\}]. \quad (4)$$

To see if two convex sets C_i and C_j are separated, one can check the following simple conditions:

$$\lambda^i_{min}(\vec{d}) > \lambda^j_{max}(\vec{d}) \quad \text{or} \quad \lambda^i_{max}(\vec{d}) < \lambda^j_{min}(\vec{d}) \quad (5)$$

where the superscript denotes index of the object. For convex polygons, considering a finite set of unit-length directions is enough to conclude if two objects are separated. These unit-length directions are the unit edge normals of the objects. An illustration of the theorem is depicted in Fig. 3.

Since images aligned to the mosaic are 2D convex objects, SAT can be used to detect intersections between the new and the previous images. To employ SAT, we must know the layout of all images on the mosaic which we can be computed by using the homographies of those images. As we do not have the homography of the new image, we perform an initial estimation between the new and previous image and obtain an estimate for the homography of this image.

We represent each image by their four vertices and these vertices form a quadrilateral when aligned to the mosaic by its homography. As we look for the previous images intersecting the new image, SAT is employed between the new image and all of the previous images one by one. Since it is enough to choose the unit-length directions $(\vec{d})$ as the edge normals of the convex objects, we need to perform the operations in (4) and (5) at most eight times for each image pair which is a very efficient procedure. Suppose we need to check two aligned images if they are separated. SAT can be performed by the following steps:

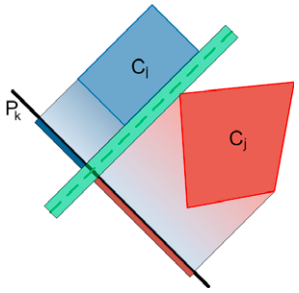


Fig. 3. An illustration of SAT. For a separating axis P_k, projected convex sets do not intersect.

1. Edge normals are obtained from the vertices of the aligned images (eight normals in total) and they are normalized to obtain the unit-length directions $\vec{d}$.
2. Operation in Eq. (4) is performed for both images by using directions $\vec{d}$ and vertices of the images (denoted with $\vec{X}$ in the equation)
3. Condition given in Eq. (5) is checked for all $\vec{d}$ directions.
4. If there exists a $\vec{d}$ for which the condition holds, it is concluded that these two images are separated which means it is unnecessary to perform matching trials for this image pair.

3.2. Homography estimation using intersecting images

As we determine all the previous images overlapping with the new image by using SAT, these images can be used to obtain a better estimate for the homography of the new image. Assume that there are n images in the mosaic overlapping with the new image. To estimate the alignment of a new image with respect to the reference image, we construct the following cost function where all the previous images and their homographies are incorporated:

$$J(H_{rn}) = \sum_{i=1}^{n-1} \|H_{ri}x_i - H_{rn}x_n\|^2 \quad (6)$$

where x_i and x_n denote the set of feature matches between the overlapping image I_i and the new image, I_n.

It should be noted that a different sampling scheme known as MLESAC [30] is employed during the homography estimations instead of classical RANSAC [28] as an MLE estimation can be beneficial for the mosaicing of quasi-planar scenes.

3.3. Affine refinement

In the proposed estimation process, alignment of new images are estimated by using their feature matches with the previous images. During the estimation, homographies of the previous images $H_{r1}, H_{r2}, \ldots, H_{r(n-1)}$ are fixed and alignment of the new image (I_n) is estimated under this constraint. As a result, we obtain a locally optimal estimate of the homography for the given image. To obtain globally optimal results, all of the homographies must be estimated jointly. However, updating the alignment of all images in each step of the algorithm cannot be handled in real-time because of the increasing computational complexity of the estimation process with the number of images in the mosaic. As a result, we propose a partial global minimization process which aims to improve the global consistency of the mosaic by considering a fixed number of previous images. We enhance global error properties of the mosaic with affine refinement while retaining real-time capabilities.

In the literature, there are studies pursuing analogues goals with our local refinement procedure. Sawhney et al. [37] propose

to refine the registration parameters of the images after they are roughly aligned to the mosaic. Gauss–Newton iterations are used in the joint optimization of the motion parameters of all images from this rough alignment. Gracias et al. [38] use affine model for image motions and update all the parameters at each time step via recursive least-squares estimation. Pizarro and Singh [39] offer affine motion model for mosaicing of the underwater images for the initial alignments of the images. They propose to estimate affine transformations for the images as an initial operation which can be performed by using linear least squares. This estimation is used to determine the topology of the mosaic which is later used in the nonlinear optimization process where global alignment is obtained. Sibley [40] and Davis [41] propose partial global optimization procedures similar to ours in their estimations for robotics and mosaicing applications, respectively. Sibley [40] proposes a local bundle adjustment procedure for robotics applications where only a small portion of the state vector (composed of robot poses and landmarks) is optimized which results in a constant time algorithm. In the context of image mosaicing, Davis [41] offers a linear least-squares refinement in which global registration parameter estimates are updated by imposing pairwise relations of images. Global registration parameters are refined in such a way that pairwise homographies obtained from these parameters deviate minimally from the pairwise estimations obtained from the image pairs.

In our method, we assume that the relation between the current and the globally optimal version of the aligned images can be described by an affine transformation. Given n consecutive images which are aligned to the mosaic, the problem is to estimate affine transformations to be applied on these images which minimize the sum-of-squares of the feature reprojection errors between the image pairs. Cost function for this optimization problem can be expressed as

$$C_{int}(A_{1:n}) = \sum_{i,j \in \substack{\text{chosen}\\ \text{images}}}^{n} \|A_i \Phi_{ij}^i - A_j \Phi_{ij}^j\|^2 \tag{7}$$

where Φ denotes the set of feature match coordinates of the aligned images and A denotes the affine transformation to be applied on a given image. Subscript of Φ implies the image pair that feature set belongs to and superscript implies the image whose features are considered. For example, Φ_{ij}^i includes feature coordinates of the aligned image i obtained from the feature matching procedure between the images i and j. A_i denotes the 3×3 affine transformation to be applied on the warped image i. Our purpose is to find affine transformations that minimize C_{int}. Assume that, at time t, refinement will be performed on the recently added n images in the mosaic. Minimization of C_{int} implies an enhanced internal consistency between these n images. However, this cost function ignores the feature reprojection errors between the chosen images and the rest of the mosaic. For this reason, we propose a new term C_{ext}, which considers the consistency between chosen images and rest of the mosaic. This new term can be expressed as

$$C_{ext}(A_{1:n}) = \sum_{i \in \substack{\text{chosen}\\ \text{images}}} \sum_{j \in \substack{\text{rest}\\ \text{of the}\\ \text{mosaic}}} \|A_i \Phi_{ij}^i - \Phi_{ij}^j\|^2. \tag{8}$$

Consequently, by considering both internal consistency of n images and external consistency of these n images with the mosaic, we first propose to update our cost function by a linear combination of C_{int} and C_{ext}. However, system of equations constructed from these terms become ill-conditioned. For this reason, we add a regularization term to our cost function to regularize the system of equations. Since we assume that features of the warped images are close to their optimal position in the mosaic, all of the estimated affine transformations must be close to the identity. Accordingly, we choose to penalize the differences of the affine transformations from the identity, which in turn implies penalizing the displacements of the warped features from their initial positions. Regularization term can be written as

$$C_{reg}(A_{1:n}) = \sum_{i,j \in \substack{\text{chosen}\\ \text{images}}} \|(A_i - \mathrm{I})\Phi_{ij}^i\|^2 + \|(A_j - \mathrm{I})\Phi_{ij}^j\|^2 \tag{9}$$

where I is the 3×3 identity matrix. Eqs. (7)–(9) can be linearly combined to obtain the final cost function

$$f(A_{1:n}) = C_{int} + \lambda_1 C_{ext} + \lambda_2 C_{reg} \tag{10}$$

where λ_1 and λ_2 are the weights for C_{ext} and C_{reg} terms. Since every affine transformation has 6 independent parameters, for n images the solution vector will have $6n$ parameters. This optimization problem can be solved in an efficient manner since it can be expressed as a linear-least-squares problem defined on a limited number of images.

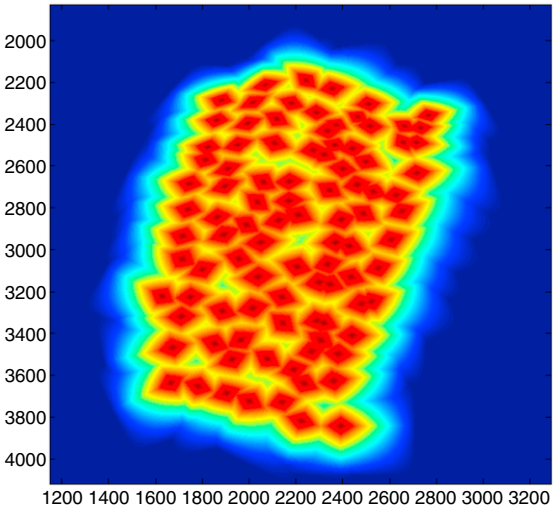


Fig. 4. Resampled weight functions of each image in the mosaic.

4. Offline enhancements

When the complete mosaic is obtained by aligning the images, results are post-processed with gain compensation [1] and multi-band blending operations [42]. By using these operations, seams caused by the illumination differences and misregistrations are reduced and visually appealing results are obtained. Finally, a contrast stretching procedure is applied on the mosaic images to compensate for a possible loose of contrast in the composite images.

4.1. Gain compensation

One of the main constituent of the seams in the mosaic images is the illumination differences in the images. These differences can be corrected by using gain compensation. Gain compensation is based on an optimization problem by which we obtain gain values for all the images that minimize sum-of-squares of the illumination differences across the overlapping regions of the images. Given n images, cost function of the optimization problem is given by the following equation [1]:

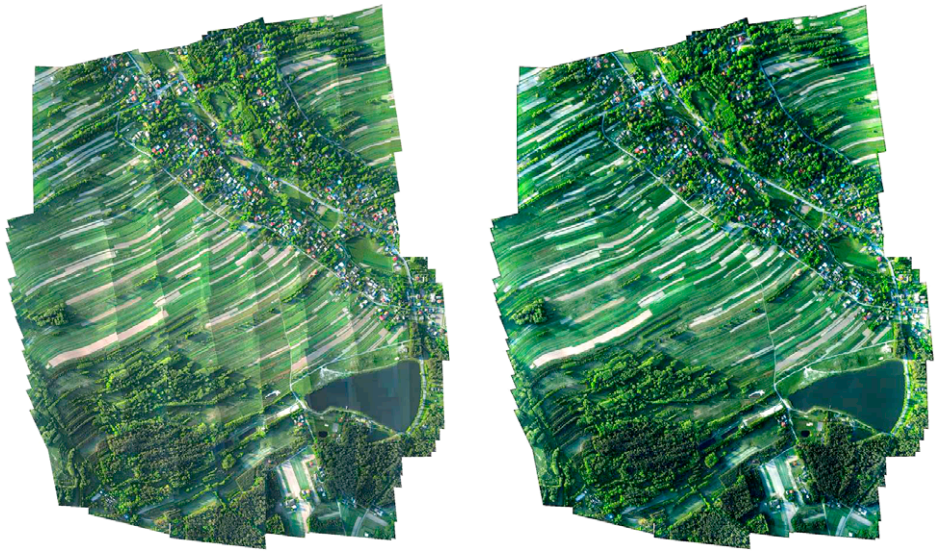
$$e = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} N_{ij} \left(\frac{(g_i I_{ij} - g_j I_{ji})^2}{\sigma_N^2} + \frac{(1 - g_i)^2}{\sigma_g} \right). \tag{11}$$

(a) Czyste.

(b) Munich quarry.

(c) Savona highway.

Fig. 5. Sample images from the aerial image datasets.



(a) Before postprocessing. (b) After postprocessing.

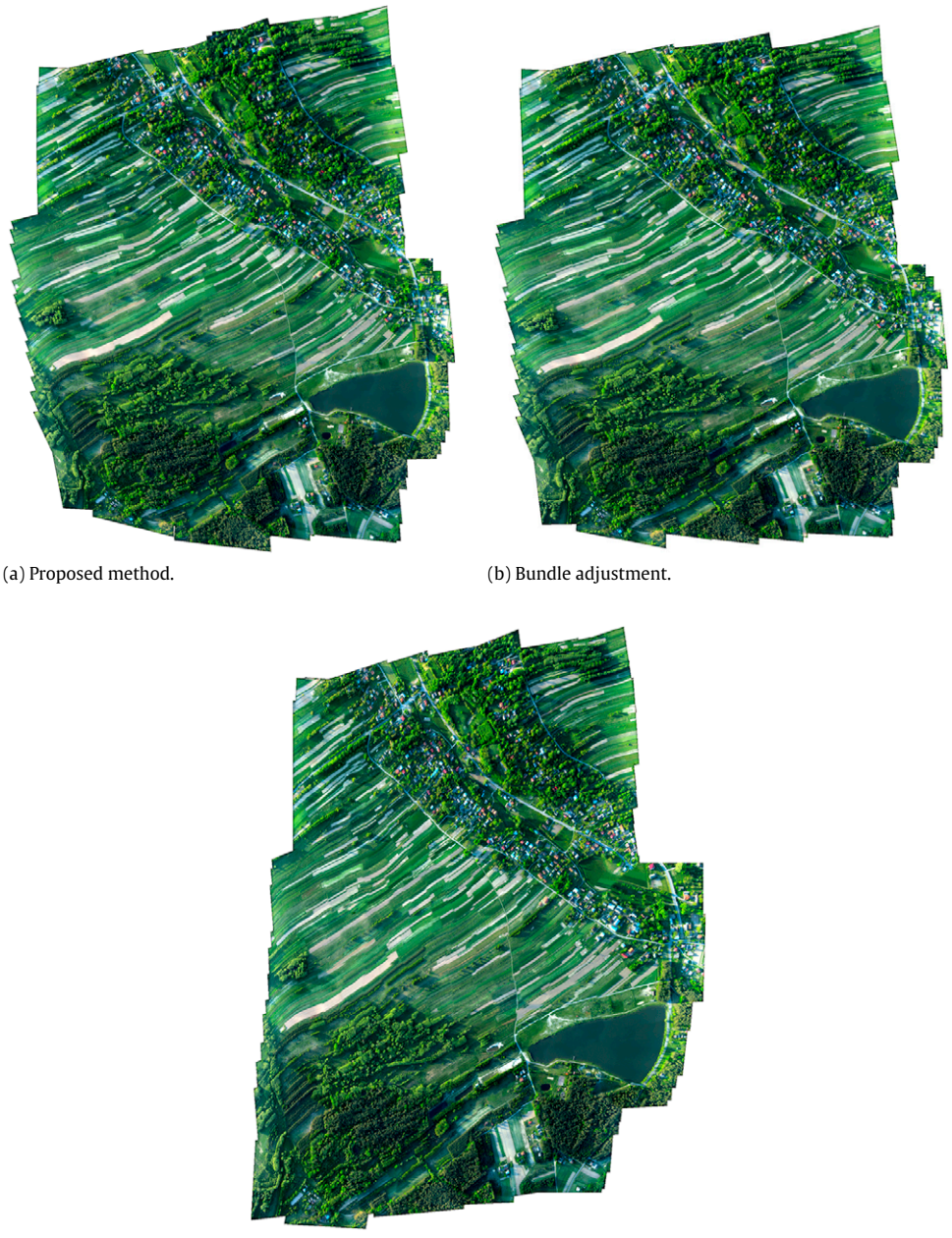
Fig. 6. Mosaic image for the Czyste image sequence before and after postprocessing.

N_{ij} is number of pixels in the overlapping region of the images i and j. I_{ij} and I_{ji} denote the mean intensity value of the overlapping regions for the images i and j, respectively. Parameters σ_N and σ_g are variances of normalized intensity errors and gains respectively. g_i denotes the gain value for the image i. Gains of the images are obtained from the minimization of the cost function which can easily be done in closed form.

4.2. Multi-band blending

Seams caused by illumination differences can be reduced with gain compensation. However, there are also some misregistrations on the mosaic image because of the violation of the assumption of scene planarity and error accumulations in the loop closing regions of the mosaic. We propose to improve the mosaic image with multi-band blending algorithm [42] by which we aim to attenuate these visual artifacts. Algorithm given in Brown et al. [1] is used to blend the mosaic image.

In the blending operation, information from multiple images are fused by using a weight function. The weight function can be denoted as $W(x, y) = w(x)w(y)$ where $w(x)$ and $w(y)$ vary linearly between 0 at the edges and 1 at the image center. These weight functions are also resampled by warping them on the mosaic frame which is denoted as $W^i(x, y)$ (see Fig. 4). Now



(a) Proposed method.

(b) Bundle adjustment.

(c) Gracias' method.

Fig. 7. Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Czyste image sequence.

the problem is to fuse all images given corresponding weight functions.

The blending weights for each image are initialized by finding the set of points for which image i is most responsible.

$$W^i_{max}(x,y) = \begin{cases} 1 & \text{if } W^i(x,y) = argmax_j W^j(x,y) \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

The weight maps are further blurred for each band to form weights in each band. A high pass version of the image can be formed by

$$B^i_\sigma(x,y) = I^i(x,y) - I^i_\sigma(x,y) \quad (13)$$

$$I^i_\sigma(x,y) = I^i(x,y) * g_\sigma(x,y) \quad (14)$$

where $g_\sigma(x,y)$ is the Gaussian filter with standard deviation σ. The image $I^i(x,y)$ is convolved with a Gaussian and subtracted from itself to form high pass version of the image, which preserves less details. $B^i_\sigma(x,y)$ represents spatial frequencies in the range of wavelengths $\lambda \in [0,\sigma]$. For this band, the images must be convolved with corresponding *max* weight functions

$$W^i_\sigma(x,y) = W^i_{max}(x,y) * g_\sigma(x,y) \quad (15)$$

where $W^i_\sigma(x,y)$ represents blending weight for the wavelength $\lambda \in [0,\sigma]$. In what follows, each subsequent band $k \geq 1$ is blended using previous lower frequency band images and weights

$$B^i_{(k+1)\sigma} = I^i_{k\sigma} - I^i_{(k+1)\sigma} \quad (16)$$

$$I^i_{(k+1)\sigma} = I^i_{k\sigma} * g_{\sigma'} \quad (17)$$

$$W^i_{(k+1)\sigma} = W^i_{k\sigma} * g_{\sigma'} \quad (18)$$

where the Gaussian standard deviation of next band is set as $\sigma' = \sqrt{2k+1}\sigma$. As a result, subsequent bands have the same range of

(a) Proposed method.

(b) Bundle adjustment.

(c) Gracias' method.

Fig. 8. Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Munich Quarry image sequence.

wavelengths. The final composite for each band is formed as:

$$I^{multi}_{k\sigma}(x,y) = \frac{\sum_{i=1}^{N} B^{i}_{k\sigma}(x,y) W^{i}_{k\sigma}(x,y)}{\sum_{i=1}^{N} W^{i}_{k\sigma}(x,y)}. \quad (19)$$

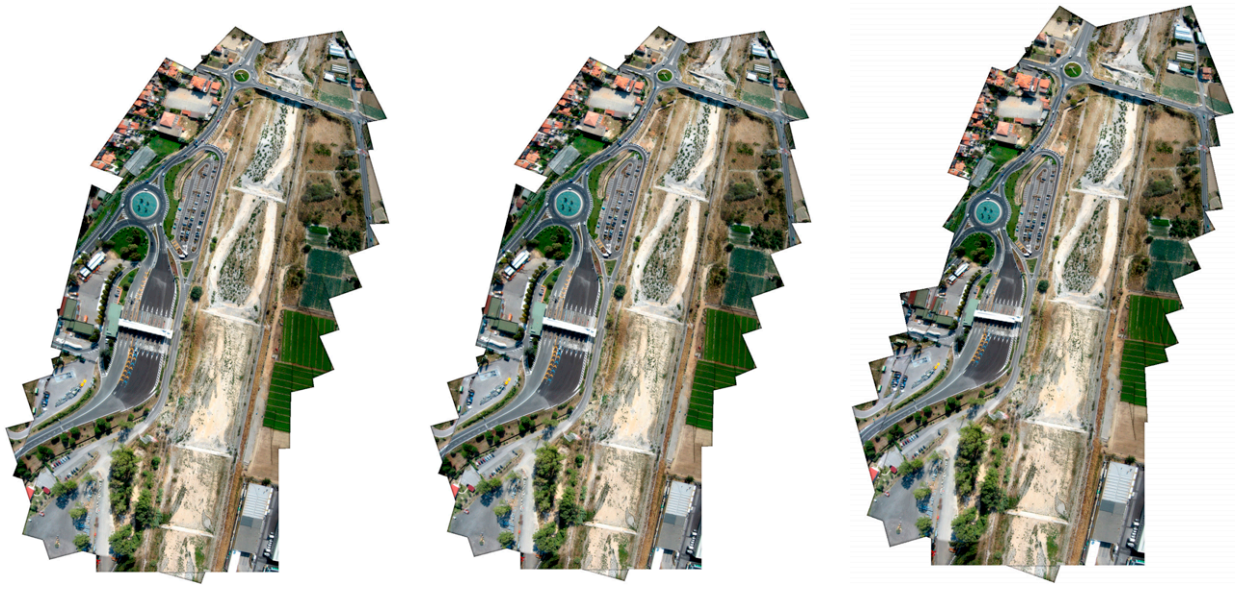
The multi-band mosaic is obtained by summing images of all subsequent bands. This multi-band blending approach allows high frequency bands to be blended over short ranges and low frequency bands to be blended over long ranges.

5. Experimental results

We tested our mosaicing approach on the images of three publicly available datasets. These are Czyste [43], Munich Quarry [44] and Savona Highway [44]. A set of sample images selected from these datasets are depicted in Fig. 5.

Our method is run for two different cases: with and without affine refinement. For the case with affine refinement, procedure is chosen to be run for once in every ten step of the algorithm on the most recent thirty images. We postprocess the results by using gain compensation, multi-band blending and contrast stretching. Results we obtain before and after post processing are shown in Fig. 6 for Czyste. It can be observed that illumination differences are eliminated and seams caused by misregistrations are attenuated which provides visually pleasing results. However, it should be noted that the raw mosaic image is also satisfactory since it provides a sufficient scene awareness which is one of the main purposes of aerial image mosaicing.

To compare the performance of the proposed method, we also performed experiments with some other methods in the literature. One of these is the study of Gracias et al. [38] where



(a) Proposed method. (b) Bundle adjustment. (c) Gracias' method.

Fig. 9. Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Savona Highway image sequence.

a real-time affine mosaicing technique is proposed based on recursive least-squares estimation. We also implemented the bundle adjustment algorithm[22] where optimization is run on the homography parameters of the images. Homographies of all the images are estimated by minimizing the total feature reprojection error between image pairs. This nonlinear optimization problem is solved using Levenberg–Marquardt algorithm.

Visual results of proposed method (with affine refinement), the bundle adjustment and Gracias' method for Czyste, Munich Quarry and Savona Highway datasets are given in Fig. 7, Fig. 8, Fig. 9, respectively. It is apparent from Fig. 7 that mosaic results of the proposed method and the bundle adjustment are similar to each other and these results are quite different than the one created with Gracias' method. For the Munich Quarry and Savona Highway datasets in Figs. 8 and 9, it is observed that image mosaics created from the proposed method are indistinguishable from the results of the bundle adjustment. Results of the Gracias' method are also similar to those of the proposed method and the bundle adjustment. However, some differences are visible in the results of this method when the mosaic images are carefully examined.

5.1. Numerical comparisons

Since visual comparisons can be subjective, a numerical evaluation of the algorithms is also necessary. To evaluate the algorithm performances, feature reprojection errors present in the results of each method are calculated. We use the root mean square (RMS) of the norm of feature reprojection errors as our performance metric.

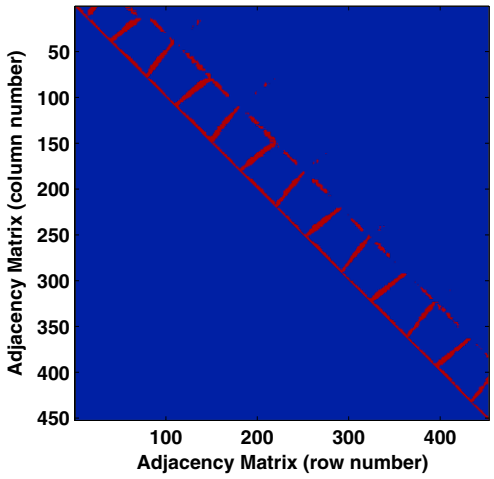
5.1.1. Czyste image sequence

For the Czyste image sequence, 453 images are used during the experiments. We calculate the error for the proposed approach with/without affine refinement. Results for the implementation of Gracias et al. (2004) and the bundle adjustment are also calculated. Spatial relations between images are given in Fig. 10. Matched image pairs are depicted as red points in the adjacency matrix which is shown in Fig. 10(a). The number of matching images versus image indices is plotted in Fig. 10(c). Camera trajectory for the dataset is sketched in Fig. 10(b). Total number of 645 272 pairwise feature matches are used during the computations. All of these feature matches are utilized during the operation of each method. Results are given in Table 1. It can be inferred from the table that RMS value is the smallest for the bundle adjustment which is an expected result since bundle adjustment is supposed to give the lower bound of the sum of squared errors. It is also apparent from the table that affine refinement improves the error characteristics of the image mosaic by 10.2% in terms of RMS value when compared to the case without affine refinement. For this experiment, Gracias' method gives slightly better results than our algorithm. It is partly because this method utilizes a recursive estimation scheme where motion parameters of all the images are estimated in every step of the algorithm. This provides a better global consistency to the Gracias' method. However, success of the algorithm is mainly because of the accuracy of the affine motion model for the given images. For an image sequence where perspective distortions between the images and the reference image are negligible, the algorithm can give successful results since the affine motion model handles such cases effectively.

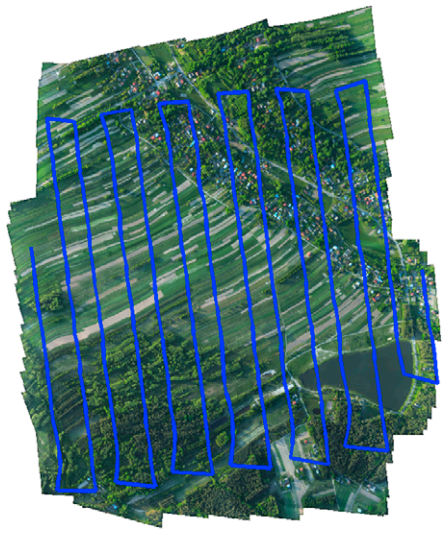
Table 1
RMS values for the four cases in Czyste image sequence.

Case	Total matches	RMS(pix)
Algorithm (without Affine Refinement)	645 272	4.4520
Algorithm (with Affine Refinement)	645 272	3.9971
Gracias et al. (2004)	645 272	3.8636
Bundle Adjustment	645 272	0.8390

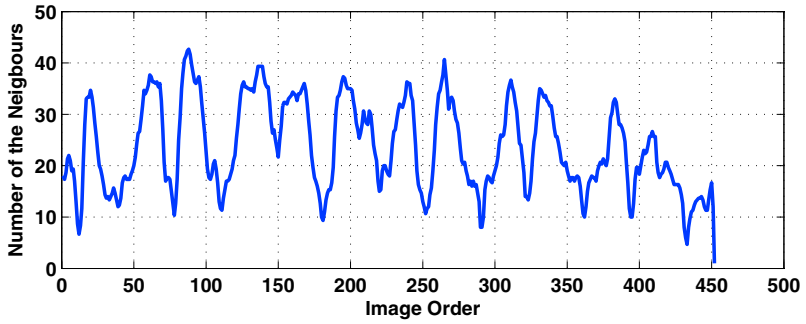
Cumulative distributions of the error for all the methods are plotted in Fig. 11. It can be observed from Fig. 11 that for the same residual norm value, cumulative distribution value for the case with affine refinement is less for small pixel values and more for larger pixel values when compared to the case without affine refinement. This implies that affine refinement decreases the norm of the large residuals at the expense of increasing the small ones which means the error norms are more uniformly distributed. Same behavior is also observed between the proposed technique and Gracias' method. Our algorithm outperforms Gracias' algorithm for small residual values and underperform for large residuals which causes the RMS value of this method to be smaller than our algorithm since the large residuals have a leverage effect on the sum-of-squared errors.



(a) Adjacency matrix.



(b) Approximate trajectory of the UAV sketched on the mosaic.



(c) Number of the neighbor images for a given image.

Fig. 10. Visual and numerical presentations of the spatial image relations in Czyste. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

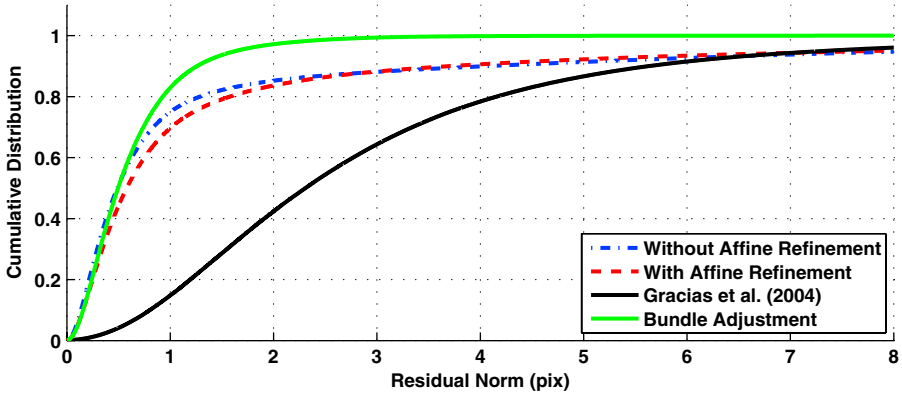


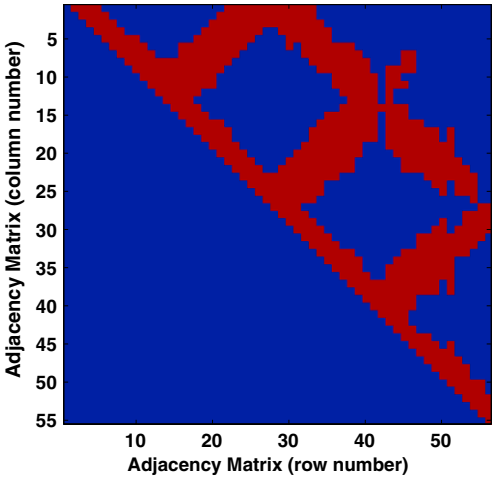
Fig. 11. Cumulative distribution of the residual error norms for Czyste image sequence.

5.1.2. Munich quarry image sequence

For the Munich Quarry image sequence, 56 images are used during the experiments. Spatial relations of the images in the mosaic are depicted in Fig. 12. RMS values are given in Table 2 for different methods. It can be inferred from the table that RMS values for the proposed approach with/without affine refinement are both close to the RMS value of the bundle adjustment. Cumulative distributions of the residuals are also similar to each other for these cases which is clear from Fig. 13. There is a 6.1% decrease in the RMS value when affine refinement is activated. It is an important improvement as the difference between the proposed approach without affine refinement and the bundle adjustment is 9.7%. For the results of the Gracias' method, RMS value is found to be larger than other methods.

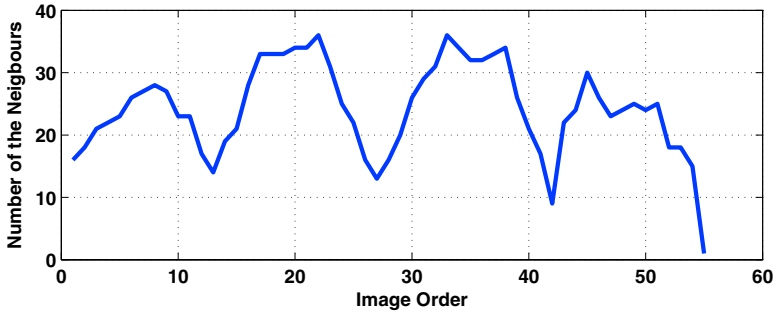
5.1.3. Savona highway image sequence

30 images are used during the experiments. Spatial relations of the images are given in Fig. 14. RMS values are provided in Table 3 for different methods. It can be inferred from the table that performance of the proposed approach with affine refinement is nearly equal to the results of the bundle adjustment. Gracias'



(a) Adjacency matrix.

(b) Approximate trajectory of the UAV sketched on the mosaic.



(c) Number of the neighbor images for a given image.

Fig. 12. Visual and numerical presentations of the spatial image relations in Munich Quarry.

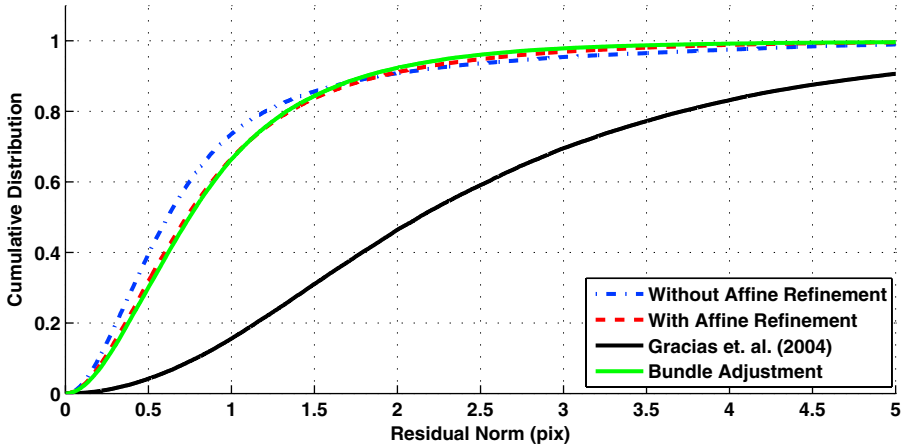


Fig. 13. Cumulative distribution of the residual error norms for Munich Quarry image sequence.

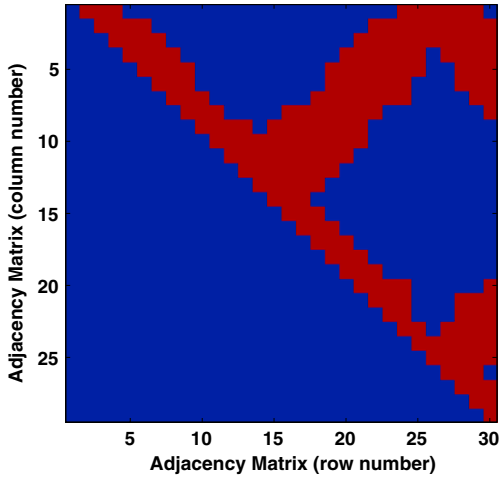
Table 2
RMS values for the four cases in Munich Quarry image sequence.

Case	Total matches	RMS(pix)
Algorithm (without Affine Refinement)	69 149	1.3497
Algorithm (with Affine Refinement)	69 149	1.2676
Gracias et al. (2004)	69 149	3.1742
Bundle Adjustment	69 149	1.2185

Table 3
RMS values for the four cases in Savona Highway image sequence.

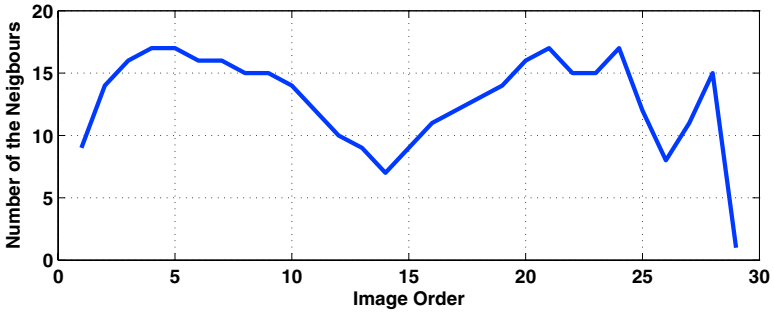
Case	Total matches	RMS(pix)
Algorithm (without Affine Refinement)	72 509	1.4402
Algorithm (with Affine Refinement)	72 509	1.2252
Gracias et al. (2004)	72 509	4.4611
Bundle Adjustment	72 509	1.2137

method has the largest RMS value among all methods which is again due to the affine motion model where large perspective distortions cause the method to underperform. Cumulative distributions of the error are plotted in Fig. 15. It is obvious from this figure that cumulative distributions are also very similar for the proposed method and the bundle adjustment. It should be noted that because of the selection of the reference image, growth and shrink of the images are apparent in Fig. 9 (see Section 2).



(a) Adjacency matrix.

(b) Approximate trajectory of the UAV sketched on the mosaic.



(c) Number of the neighbor images for a given image.

Fig. 14. Visual and numerical presentations of the spatial image relations in Savona Highway.

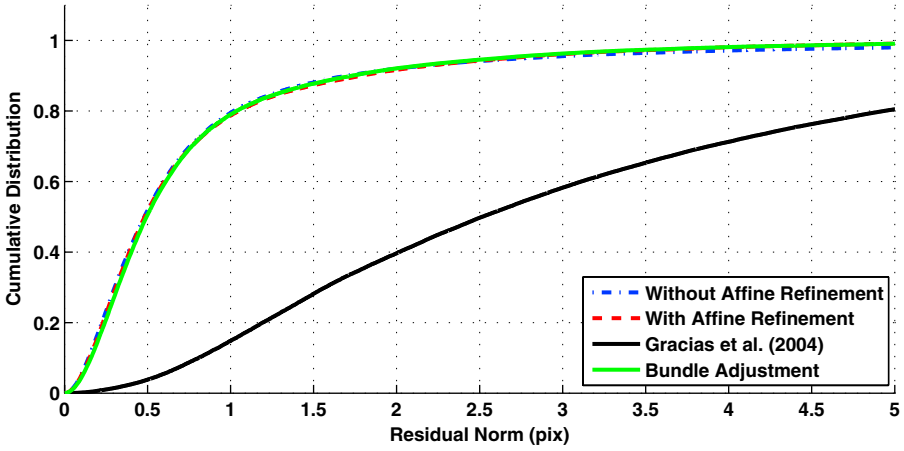


Fig. 15. Cumulative distribution of the residual error norms for Savona Highway image sequence.

It should be noted that the improvement achieved by affine refinement will be useful for cases where navigational requirements are more stringent. However, it can be deactivated for cases where only visual appearance is the prime concern.

6. Conclusion

We have now developed a new image mosaicing approach which can create large image mosaics from a set of ordered aerial images in real-time. First, all of the previous images which have feature matches with the new image are taken into account during the homography estimation by utilizing Separating Axis Theorem (SAT). Second, a novel affine refinement is performed on the aligned images to improve the global consistency of the mosaic. Furthermore, gain compensation and multi-band blending procedure are suggested as offline steps to attenuate the seams in the raw mosaic image. Visually appealing mosaic results are obtained on some publicly available datasets. Our method is also compared visually and numerically with two state-of-the-art algorithms in the literature. Performance evaluations in terms of RMS values of the residuals show success of our algorithm.

As future works, we plan to develop a more meaningful way of selecting images used in the affine refinement procedure instead of only using the temporally recent ones.

References

[1] Matthew Brown, David G. Lowe, Automatic panoramic image stitching using invariant features, Int. J. Comput. Vision 74 (1) (2007) 59–73.
[2] Tae Eun Choe, Isaac Cohen, Munwai Lee, Gerard Medioni, Optimal global mosaic generation from retinal images, in: Proceedings of the 18th International Conference on Pattern Recognition - vol. 03, ICPR '06, IEEE Computer Society, Washington, DC, USA, 2006, pp. 681–684.
[3] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec, Nicholas Ayache, Robust mosaicing with correction of motion distortions and tissue deformation for in vivo fibered microscopy, Med. Image Anal. 10 (5) (2006) 673–692.
[4] Ludovico Carozza, Alessandro Bevilacqua, Filippo Piccinini, Mosaicing of optical microscope imagery based on visual information, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE, 2011, pp. 6162–6165.
[5] Heeseung Choi, Kyoungtaek Choi, Jaihie Kim, Mosaicing touchless and mirror-reflected fingerprint images, Information Forensics and Security, IEEE Transactions on 5 (1) (2010) 52–61.
[6] Yuping Lin, G. Medioni, Map-enhanced uav image sequence registration and synchronization of multiple image sequences, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007, pp. 1–7.
[7] Pizarro, Oscar, Hanumant Singh, Toward large-area mosaicing for underwater scientific applications, IEEE J. Ocean. Eng. 28 (4) (2003) 651–672.
[8] Michal Irani, Steve Hsu, P. Anandan, Video compression using mosaic representations, Signal Process., Image Commun. 7 (4–6) (1995) 529–552.
[9] Richard Szeliski, Image alignment and stitching: a tutorial, Found. Trends. Comput. Graph. Vis. 2 (1) (2006) 1–104.
[10] Bruce D. Lucas, Takeo Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th international joint conference on Artificial intelligence - vol. 2, IJCAI'81, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 1981, pp. 674–679.
[11] Simon Baker, Iain Matthews, Lucas-kanade 20 years on: A unifying framework, Int. J. Comput. Vision 56 (3) (2004) 221–255.
[12] David G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the International Conference on Computer Vision-Vol. 2 - Vol. 2, ICCV '99, IEEE Computer Society, Washington, DC, USA, 1999, p. 1150.
[13] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, Speeded-up robust features (surf), Comput. Vis. Image Underst. 110 (3) (2008) 346–359.
[14] Tinne Tuytelaars, Luc J. Van Gool, Matching widely separated views based on affine invariant regions, Int. J. Comput. Vis. 59 (1) (2004) 61–85.
[15] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, 2004, ISBN: 0521540518.
[16] Steven Lovegrove, Andrew J. Davison, Real-time spherical mosaicing using whole image alignment, in: Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III, ECCV'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 73–86.
[17] Fernando Caballero, Luis Merino, Joaquin Ferruz, Anbal Ollero, Homography based kalman filter for mosaic building. applications to uav position estimation, in: ICRA'07, 2007, pp. 2004–2009.
[18] Javier Civera, Andrew J. Davison, Juan A. Magallón, J.M. Montiel, Drift-free real-time sequential mosaicing, Int. J. Comput. Vision 81 (2) (2009) 128–137.
[19] Eun-Young Kang, I. Cohen, G. Medioni, A graph-based global registration for 2d mosaics, in: Pattern Recognition, 2000. Proceedings. 15th International Conference on, vol. 1, 2000, pp. 257–260.
[20] Armagan Elibol, Nuno Gracias, Rafael Garcia, Fast topology estimation for image mosaicing using adaptive information thresholding, Robot. Auton. Syst. 61 (2) (2013) 125–136.
[21] Dae Woong Kim, Ki-Sang Hong, Real-time mosaic using sequential graph, J. Electronic Imaging 15 (2) (2006) 023005.
[22] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, Andrew W. Fitzgibbon, Bundle adjustment - a modern synthesis, in: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, ICCV '99, Springer-Verlag, London, UK, 2000, pp. 298–372.
[23] M.I. A. Lourakis, A.A. Argyros, SBA: A Software Package for Generic Sparse Bundle Adjustment, ACM Trans. Math. Software 36 (1) (2009) 1–30.
[24] Kurt Konolige, Sparse sparse bundle adjustment, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2010, pp. 102.1–102.11. http://dx.doi.org/10.5244/C.24.102.
[25] Changchang Wu, Sameer Agarwal, Brian Curless, Steven M. Seitz, Multicore bundle adjustment, in: In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 3057–3064.
[26] Guoqing Zhou, Near real-time orthorectification and mosaic of small uav video flow for time-critical event response, IEEE T. Geoscience and Remote Sensing (2009) 739–747.
[27] R. Ross, J. Devlin, A. De Souza-Daw, Mobile robot mosaic imaging of vehicle undercarriages using catadioptric vision, in: Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on, 2012, pp. 247–252.
[28] Martin A. Fischler, Robert C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[29] Chang Yuan, Gerard Medioni, Jinman Kang, Isaac Cohen, Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints, IEEE Trans. Pattern Anal. Mach. Intell. 29 (9) (2007) 1627–1641.
[30] P.H.S. Torr, A. Zisserman, Mlesac: A new robust estimator with application to estimating image geometry, Comput. Vis. Image Underst. 78 (2000) 2000.
[31] Ben Tordoff, David W. Murray, Guided sampling and consensus for motion estimation, in: Proceedings of the 7th European Conference on Computer Vision-Part I, ECCV '02, Springer-Verlag, London, UK, UK, 2002, pp. 82–98.
[32] O. Chum, J. Matas, Matching with prosac - progressive sample consensus, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, 2005, pp. 220–226.
[33] Ondrej Chum, Jir Matas, Josef Kittler, Locally optimized ransac, in: Bernd Michaelis, Gerald Krell (Eds.), Pattern Recognition, in: Lecture Notes in Computer Science, vol. 2781, Springer, Berlin, Heidelberg, 2003, pp. 236–243.
[34] T. Suzuki, Y. Amano, T. Hashizume, Vision based localization of a small uav for generating a large mosaic image, in: SICE Annual Conference 2010, Proceedings of, 2010, pp. 2960–2964.
[35] D. Blake Barber, Joshua D. Redding, Timothy W. Mclain, Randal W. Beard, Clark N. Taylor, Vision-based target geo-location using a fixed-wing miniature air vehicle, J. Intell. Robotics Syst. 47 (4) (2006) 361–382.
[36] Philip J. Schneider, David Eberly, Geometric Tools for Computer Graphics, Elsevier Science Inc., New York, NY, USA, 2002.
[37] Sawhney, S. Harpreet, Steve Hsu, Rakesh Kumar, Robust video mosaicing through topology inference and local to global alignment. Computer Vision–ECCV'98, Springer, Berlin, Heidelberg, 1998, 103-119.
[38] Gracias, Nuno, Costeira, Joao Paulo and Victor, JS. Linear global mosaics for underwater surveying 5th IFAC Symposium on Intelligent Autonomous Vehicles, Vol. 1. 2004.
[39] Pizarro, Oscar, Singh, Hanumant, Toward large-area mosaicing for underwater scientific applications, IEEE J. Ocean. Eng. 28 (4) (2003) 651–672.
[40] Sibley, Gabe, Relative bundle adjustment. Department of Engineering Science, Oxford University, Tech. Rep 2307.09 (2009).
[41] Davis, James, Mosaics of scenes with moving objects. Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on. IEEE, 1998.
[42] Peter J. Burt, Edward H. Adelson, A multiresolution spline with application to image mosaics, ACM Trans. Graph. 2 (4) (1983) 217–236.
[43] AerialRobotics Dataset. ftp://www.aerialrobotics.eu/ accessed: 2014-03-30.
[44] UAVPeople Dataset. http://www.uavpeople.com/download/dataset accessed: 2014-03-25.

Taygun Kekec received his B.S. degree in Computer Engineering from Yildiz Technical University, Istanbul, Turkey in 2010. He received his M.S. degree in Mechatronics Engineering from Sabanci University, Istanbul, Turkey, in 2013. He is currently a Ph.D. student at University of Jean Monnet, Saint-Etienne, France. His research interests are computer vision and machine learning.

Alper Yildirim received his B.S. degree in Mechanical Engineering from TOBB University of Economics and Technology, Ankara, Turkey, in 2011. He is currently a graduate student in Control, Vision and Robotics (CVR) research group in the Mechatronics Engineering Program at Sabanci University, Istanbul, Turkey. His research interests are computer vision, vision based control and robotics.

Mustafa Unel received the B.S. degree in Electrical and Electronic Engineering and the M.S. degree in Systems and Control Engineering from Bogazici University, Istanbul, Turkey in 1994 and 1996, respectively. He received a second M.S. degree in Applied Mathematics and the Ph.D. degree in Electrical Sciences and Computer Engineering from Brown University, RI, USA in 1998 and 1999, respectively. During 1999–2001, he was a postdoctoral fellow at Brown and Yale Universities. In Spring 2001, he joined the Department of Computer Engineering at Gebze Institute of Technology, Turkey as an assistant professor, and directed Vision, Graphics and Human–Computer Interaction Lab. Since Fall 2004, he has been with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey where he is currently a full professor and leads the Control, Vision and Robotics (CVR) research group. His research interests are in computational vision, visual servoing, unmanned aerial vehicles, motion control, robotics and microsystems.