Fast and Accurate Mosaicing Techniques for Aerial Images of Quasi-Planar Scenes

by

Alper Yıldırım

Submitted to the Graduate School of Sabancı University
in partial fulfillment of the requirements for the degree of
Master of Science

Sabancı University

August, 2014

Fast and Accurate Mosaicing Techniques for Aerial Images of Quasi-Planar Scenes

APPROVED BY:

Prof. Dr. Mustafa Ünel
(Thesis Advisor) .....

Assoc. Prof. Dr. Gözde Ünal .....

Assist. Prof. Dr. Hüsnü Yenigün .....

DATE OF APPROVAL: .....

Fast and Accurate Mosaicing Techniques for Aerial Images of Quasi-Planar Scenes

Alper Yıldırım

ME, Master's Thesis, 2014

Thesis Supervisor: Prof. Dr. Mustafa Ünel

Keywords: Image Mosaicing, Alignment, Separating Axis Theorem, Affine Refinement, Pose Estimation, Extended Kalman Filter

Abstract

Image mosaicing aims to increase visual perception by composing data from separate images since a mosaic image provides a more powerful scene description. 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 wide areas. 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.

In this thesis, novel techniques for creating fast and accurate aerial image mosaics of quasi-planar scenes are developed. First, a sequential mosaicing approach is proposed where all the past images intersecting the new image are used to estimate alignment of the new image. A tool from computer graphics, Separating Axis Theorem (SAT), is employed to detect image intersections. A new local affine refinement is introduced to provide global consistency throughout the mosaic. Second, a pose estimation based mosaicing technique is developed where the scene normal and the camera pose parameters are estimated through an Extended Kalman Filter (EKF). Mosaic is formed by using the homographies constructed from the estimated state vector. Using an EKF based approach provides a significant global consistency throughout the mosaic since all the parameters are updated by which error accumulations in the loop closing regions are compensated. Proposed

algorithm also provides localization and attitude information of the camera which might be beneficial for robotics applications. Both methods are verified through several experiments and comparisons with some state-of-the-art algorithms are presented. Results show that the developed algorithms work successfully as intended.

Düzlemsi Sahnelere Ait Havadan Çekilmiş Görüntüler İçin Hızlı ve Doğru Görüntü Mozaikleme Teknikleri

Alper Yıldırım

ME, Master Tezi, 2014

Tez Danışmanı: Prof. Dr. Mustafa Ünel

Anahtar Kelimeler: Görüntü Mozaikleme, Yerleşim, Ayırıcı Eksen Teoremi, Afin İyileştirme, Poz Kestirimi, Genişletilmiş Kalman Süzgeci

Özet

Görüntü mozaikleme, ayrı ayrı çekilmiş resimlerin bütünleştirilmesini ve bütünleşik resimlerin sahne hakkında daha iyi bir tanımlama sunmasından dolayı bu şekilde sahne hakkındaki görsel algının artırılmasını amaçlar. Mozaik resimlerden elde edilen durumsal farkındalık sivil ve askeri uygulamalar açısından önem taşır. Muhtemel sivil kullanım alanları, doğal felaketlerden dolayı hasar görmüş kentsel bölgelerin keşfi ve geniş dikili alanların incelenmesi olarak verilebilir. Askeri uygulamalar içinse, görüntü mozaikleme geniş alanda süregelen düşman aktiviteleri hakkında kritik bilgiler sağlayabilir. Literatürdeki farklı uygulamalar için geliştirilmiş çeşitli gerçek zamanlı görüntü mozaikleme çalışmalarına rağmen, birçok pratik uygulama için daha hızlı ve doğru sonuçlar veren yöntemlere duyulan ihtiyaç sebebiyle, konu hala gelişmeye açıktır.

Bu tezde, havadan alınmış düzlemsi sahnelere ait görüntülerin hızlı ve doğru şekilde mozaiklenmesini amaçlayan yeni yöntemler geliştirilmiştir. İlk olarak, yeni gelen bir resmin yerleşiminin belirlenmesi için bu resim ile kesişen bütün eski resimlerin kullanıldığı bir mozaikleme yaklaşımı geliştirilmiştir. Kesişen resimleri belirlemek için Bilgisayar Grafikleri literatüründe kullanılan Ayırıcı Eksen Teoremi kullanılmıştır. Mozaik görüntü üzerindeki global tutarlılığın artırımı için yeni bir afin iyileştirme yöntemi sunulmuştur. İkinci olarak, sahne normali ve kamera poz parameterelerinin Genişletilmiş Kalman Süzgeci ile kestirimine dayalı bir mozaikleme yöntemi önerilmiştir. Mozaik görüntü, durum vektörü parametrelerinden elde edilen homografiler yardımıyla oluşturulmaktadır. Bütün parametrelerin kestiriminin birlikte yapılması

ve bu sayede döngü kapanışlarındaki hataların kompanze edilmesinden dolayı, Genişletilmiş Kalman Süzgeci temelli bir yaklaşım kullanmak, mozaik görüntüye kayda değer oranda global tutarlılık sağlamaktadır. Önerilen metod ayrıca robotik uygulamalarda kullanışlı olabilecek kameranın yer ve duruş bilgisini de sağlamaktadır. İki yöntem de farklı durumlar için deneylere tabi tutulmuş ve bazı diğer gelişmiş mozaikleme algoritmaları ile karşılaştırılmaları sunulmuştur. Sonuçlar, geliştirilen yöntemlerin amaçlandığı gibi başarılı bir şekilde çalıştığını göstermektedir.

Acknowledgements

Foremost, I would like to express my gratitude to my thesis advisor Prof. Dr. Mustafa Ünel. First of all, I simply learnt how to do research from scratch under his guidance and support. I consider myself very lucky since I have an opportunity of selecting my research topic in an extensive number of different options in virtue of his competence in diverse areas of engineering and mathematics and also the freedom he provided generously during my studies which uncaged my imagination in every aspect.

I would gratefully thank my jury members, Assoc. Prof. Dr. Gözde Ünal and Assist. Prof. Dr. Hüsnü Yenigün for their favorable criticism, comments and suggestions for my thesis.

For their positive attitude and companionship, I am grateful to all of the members of Control, Vision and Robotics (CVR) research group, Taygun Kekeç, Barış Can Üstündağ, Caner Şahin, Sanem Evren, Talha Boz, Mehmet Ali Güney, Soner Ulun, Gökhan Alcan and Eren Demirel. I am also grateful to all of the other members of the Mechatronics Laboratory, especially to Kemal Adak for his continuous moral support during the process.

I also want to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for the generous financial support they provide within the BİDEB scholarship program.

Finally, I would like thank all of my family members, especially to my mother, for their endless support and patience all along my life. It would not be possible for me to stand where I am right now without their help and presence in my life.

Contents

1 Introduction **1**

1.1 Thesis Contributions and Organization 4

2 Background **6**

2.1 Motion Models 6

2.1.1 Translation 6

2.1.2 Euclidean 7

2.1.3 Similarity 7

2.1.4 Affine 8

2.1.5 Projectivity 9

2.2 Image Alignment 10

2.2.1 Direct Alignment 10

2.2.2 Feature Based Alignment 13

2.2.3 Advantages of Feature Based Alignment 16

2.3 Image Mosaicing 16

2.3.1 Homography 17

2.3.2 Homography Estimation 18

2.3.3 Homography Decomposition 22

3 A New Approach for Fast and Accurate Mosaicing of Aerial Images **25**

3.1 Image Mosaicing 26

3.2 Proposed Mosaicing Approach 29

3.2.1 Detection of Image Intersections by Using Seperating Axis Theorem 30

3.2.2 Homography Estimation Using Intersecting Images . . 33
3.2.3 Affine Refinement 33
3.3 Offline Enhancements 37
3.3.1 Gain Compensation 37
3.3.2 Multi-band Blending 37
3.4 Experimental Results 38
3.4.1 Numerical Comparisons 40

4 Pose Estimation Based Image Mosaicing via Extended Kalman Filter 52

4.1 Proposed Approach 54
4.1.1 Prediction 56
4.1.2 Measurement 59
4.2 Update 61
4.3 Mosaic Creation 63
4.4 Experimental Results 65
4.4.1 Small Village Image Sequence 66
4.4.2 Pteryx UAV-Volvo Factory Image Sequence 67
4.4.3 Bourget Airport Image Sequence 68
4.4.4 Construction site (France) Image Sequence 71

5 Conclusions 74

List of Figures

3.1 General structure of the proposed method 26

3.2 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. 27

3.3 An illustration of SAT. For a separating axis P_k, projected convex sets do not intersect. 32

3.4 Sample images from the aerial image datasets. 38

3.5 Mosaic image for the Czyste image sequence before and after postprocessing 39

3.6 Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Czyste image sequence 41

3.7 Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Munich Quarry image sequence . . . 42

3.8 Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Savona Highway image sequence . . . 43

3.9 Visual and numerical presentations of the spatial image relations in Czyste 45

3.10 Cumulative distribution of the residual error norms for Czyste image sequence 46

3.11 Visual and numerical presentations of the spatial image relations in Munich Quarry 48

3.12 Cumulative distribution of the residual error norms for Munich Quarry image sequence 49

3.13 Visual and numerical presentations of the spatial image relations in Savona Highway 50
3.14 Cumulative distribution of the residual error norms for Savona Highway image sequence 51
4.1 Flowchart of the proposed method 57
4.2 Initialization of the new image parameters from the previous image 58
4.3 Results of the proposed method and bundle adjustment for Small Village 67
4.4 Results of the proposed method and bundle adjustment for Small Village 69
4.5 Results of the proposed method and bundle adjustment for Bourget 70
4.6 Results of the proposed method and bundle adjustment for Construction site (France) 72

List of Tables

2.1 Possible Decompositions of the Homograpy 24
3.1 RMS values for the four cases in Czyste image sequence 44
3.2 RMS values for the four cases in Munich Quarry image sequence 47
3.3 RMS values for the four cases in Savona Highway image sequence 48
4.1 RMS values for Small Village 66
4.2 RMS values for Volvo 68
4.3 RMS values for Bourget 71
4.4 RMS values for Construction site (France) 71

Chapter I

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 merging several images to create a consistent and seamless composite image. This composite image can provide more information than spatially and temporally distinct individual images. Image mosaicing algorithms are frequently used for medical, personal and remote sensing applications. By using these algorithms, attractive panoramic images of the natural photos [1] can be obtained with from relatively cheap off-the-shelf cameras. In medical imaging, successful results are obtained from mosaicing of retinal images [2] and tissues [3]. Mosaicing algorithms

can be useful to create mosaics of microscopic [4] and fingerprint images [5]. These algorithms can also be useful in remote sensing applications where maps of an environment can be created using aerial [6] and underwater [7] images. They are also used as video compression and image stabilization purposes [8].

Finding the alignments of the images is the central part of all mosaicing algorithms. In literature, image alignment methods are usually categorized under two main categories: dense and sparse methods. These are known as direct and feature based alignment approaches [9]. In direct approaches, all the available data in the image is used instead of a set of sparse features in the images. Transformation parameters and pixel correspondences are estimated simultaneously in these approaches. These approaches provide a higher accuracy when compared to the feature based approaches since all the image information is exploited. Although this provides more accuracy, they require a close initialization to the true solution and a high degree of overlap between the images for the algorithm to converge. Pioneering work in this area is done by Lucas and Kanade [10]. An overview on historical progress and extensions of direct approaches can be found in [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] are available under projectivity. Projective homography is the most general linear transformation model for image mosaicing applications

where the scene is planar and the camera undergoes a rigid motion [9]. For the case of pure rotational camera motion, homography becomes the rotation matrix which is represented with a less number of parameters by which estimation becomes more stable [16, 1].

Several different frameworks have been proposed to create image mosaics. One approach is to consider the mosaicing problem under a recursive estimation framework [17] where homography parameters are treated 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 is considered 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.

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].

1.1 Thesis Contributions and Organization

In this thesis, two new mosaicing techniques capable of creating fast and accurate image mosaics of quasi-planar scenes are developed. Our contributions

can be highlighted as follows:

- A new mosaicing approach where alignments of the new images are computed by using all the previously aligned images intersecting the new image.
- To detect image intersections in an efficient manner, a tool from computer graphics, Separating Axis Theorem (SAT), is employed.
- A local affine refinement procedure is introduced to provide a better global consistency throughout the mosaic.
- A novel mosaicing technique based on camera pose estimation is developed where scene normal and camera pose parameters are updated by an Extended Kalman Filter (EKF). EKF handles error accumulations in the loop closing regions.

Organization of the thesis can be summarized as follows:
In Chapter 2, background information for image alignment and mosaicing is given. In Chapter 3, first mosaicing approach is presented. Visual and numerical results for this algorithm are provided with several experiments. In Chapter 4, our second mosaicing approach which is based on camera pose estimation is introduced. Algorithm is tested on some image datasets and visual and numerical results are presented. Finally, thesis is concluded in the Chapter 5 with some remarks.

Chapter II

2 Background

Image mosaicing process involves aligning the images captured from different camera poses to each other. The fundemental part of all the mosaicing algorithms is to find the alignments between images. Finding the alignments include obtaining a mathematical mapping between the pixel coordinates of these images.

2.1 Motion Models

Several different parametric models can be used for the purpose of image alignments. We can summarize these models as translation, Euclidean, similarity, affine and projective models.

2.1.1 Translation

Translation between the the pixel coordinates of two images can be given as:

$$\mathbf{x}' = \mathbf{x} + \mathbf{t} \tag{1}$$

where the $\mathbf{x}'$ and $\mathbf{x}$ denote the pixel coordinates of the images. This can be expressed with a linear transformation by using homogeneous coordinates as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & t_x \\ 0 & 0 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2}$$

2.1.2 Euclidean

Euclidean model includes a 2D translation and 2D rotation between images. Given a 2D rotation $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$ and translation $\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$, Euclidean motion between the homogeneous coordinates of two images can be given as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

Euclidean motion preserves the magnitude and relative angle proporties of the lines in space. It has 3 degrees of freedom (DOF) as the 2D rotation has one DOF and the translation has 2 DOF.

2.1.3 Similarity

Similarity transformation is a motion model which is composed of an isometric scaling and Euclidean motion. For a scaling $\mathbf{S} = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix}$, it can be given

as follows:

$$\mathbf{x}' = \mathbf{SRx} + \mathbf{t} \tag{4}$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_x \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{5}$$

Similarity transformation has four DOF. These are three DOFs of the Euclidean motion and a scaling factor for the isometric scaling denoted with s. It is a shape preserving transformation where angles between lines and ratio of the line lengths remain unchanged. A similarity transformation can be calculated from 2 point correspondences.

2.1.4 Affine

Affine model includes a six DOF linear transformation which can be written in terms of homogeneous pixel coordinates as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

Affine transformations preserve the parallelism. Area ratios are also invariant under this transformation. Ratio of the lengths of the line segments are not preserved except the case where lines are parallel to each other.

2.1.5 Projectivity

Projectivity is the most general linear transformation that is defined with a 3×3 nonsingular matrix. A projective transformation can be given as below in terms of homogeneous coordinates:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{21} & h_{22} & h_{23} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

where the transformation matrix includes nine elements. A 3×3 projective transformation mapping homogeneous pixel coordinates to each other is also called as homography. It differs from an affine transformation by its last row which includes extra three elements. However since ratio of these elements to each other matters because of the homogeneous coordinates, it has eight degrees of freedom where any nonzero multiple of the matrix implies the same transformation. In terms of pixel coordinates, this mapping can be given with the following nonlinear equation:

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \tag{8}$$

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{9}$$

Cross ratio of the collinear points is an invariant of the projective transformation. Parallelism is not usually preserved under projective transformations.

2.2 Image Alignment

After a suitable motion model is chosen, parameters of this model must be estimated. Since it is not usually possible to find a perfect alignment between images because of the uncertainties such as noise, illumination differences and parallax, this problem is usually expressed as an optimization problem where 'best' possible alignment between images is found. There are two main approaches in the literature based on the utilized cost function to find the alignment parameters of images. These are the direct (pixel based) and feature based approaches. Both approaches have their advantages and disadvantages.

2.2.1 Direct Alignment

This approach includes warping the image on top of the other and trying to find the parameters by which the overlapping pixels of both images agree. This problem is defined on several different properties of the images [9]. The simplest approach is to find alignments parameters by minimizing intensity differences between images. Assume that we want to find alignment between two images by using a translational motion model. Cost function based on the intensity differences can be given by the following equation:

$$E_{SSD}(u) = \sum_i e_i^2 = \sum_i \left[I_1(x_i + u) - I_0(x_i)\right]^2 \tag{10}$$

where u is the displacement and $I_1(x_i)$ denotes the intensity value of the image at x_i. However, it is possible that a bias and a scale differences exist in the image intensities. To handle these illumination differences, a bias and a scale parameter can be added to the cost function [10]. Updated cost

function can be given as:

$$E_{SSD}(u) = \sum_i e_i^2 = \sum_i \left[I_1(x_i + u) - (1+\alpha) I_0(x_i) - \beta\right]^2 \tag{11}$$

where β and α denotes the bias and gain parameters respectively. Since squared diffences of the intensities are used in the optimization problem, outliers can dramatically affect the results of the problem. To reduce the affects of these outliers, robust cost functions are proposed in the literature. For example, it is possible to use sum of absolote differences (SAD) of the intensities instead of using a least square scheme which can be given as:

$$E_{SAD} = \sum_i \|e_i\| = \sum_i \|I_1(x_i + u) - (1+\alpha) I_0(x_i) - \beta\| \tag{12}$$

However, this function is not suitable to be used with the optimization techniques where Jacobians are utilized as it is not differentiable at the origin. Using a differentiable function which does not grow as fast as square function can be a possible option. For example, Huber robust error function [34] is given as:

$$h(x) = \begin{cases} \|x\|^2 & x < \sigma \\ 2\sigma\|x\| - \sigma^2 & x \geq \sigma \end{cases} \tag{13}$$

This cost function has both the fast convergence properties of L_2 norm and robustness of a L_1 norm [1]. If this kind of a robust error function is used,

the cost function is given as:

$$E_{SAD} = \sum_i h\left(e_i\right) = \sum_i h\left(I_1\left(x_i + u\right) - (1+\alpha) I_0\left(x_i\right) - \beta\right) \tag{14}$$

It should be noted that, in direct alignment, a hierarchical estimation scheme [35] is usually employed to speed up the convergence of the problem. This is done by using an image pyramid where estimation is first performed on coarser level and results of this estimation is used in a finer level for initialization.

Direct alignment can also be performed for other motion models other than pure translation. In this case, instead of using a translation vector u, a spatially varying motion field which is a function of x_i parameterized by a small size parameter vector (parameters of the motion model) is employed. As a result, new cost function can be given as:

$$E_{SSD}\left(u\right) = \sum_i e_i^2 = \sum_i \left[I_1\left(f(x_i, p)\right) - I_0\left(x_i\right)\right]^2 \tag{15}$$

where f is the function that maps a given point x_i according to the motion model parametrized by p vector.

The biggest advantage of the direct approaches is that they can use all the information in the image which provides accurate registration results. Also, these methods can be used for the cases where the amount of the texture in the images (distinctive features) is insufficient. Their biggest disadvantage is they have a limited range of convergence [9].

2.2.2 Feature Based Alignment

Feature based registration is another approach that is used to align images. These approaches are based on utilizing sparse distinctive features of the images and using them to estimate the alignment parameters. To find the alignment between two images, distinctive features are extracted from both images and feature matching is employed after finding the feature correspondences. Feature based approaches are available in the literature for a long time. Some old studies employing these approaches are [36] and [37].

Several different image features can be used for image alignment. Recent feature detectors (keypoint detectors) have good invariance properties that can be used to find point matches between images. This provides robustness to the large point-of-view changes in the images. For example, some feature detectors have good scale ([38]) and affine invariance properties ([39], [40] and [41]). It is also possibe to use some other kind of features for image alignment. For example, line features can be exploited as in [42] and [43]. Tuytelaars and Van Gool [44] propose to use affine invariant regions to detect correspondences between images.

After the features are detected from images, it is important to find the feature matches between images. For some cases e.g. video sequences [45], local motion around the point features can be assumed to be translational where equation (10) can be utilized to compare the small patches around feature points. For the situations where features are tracked over long image sequences, appearances of the features may change dramatically. In this case, it is more reasonable to use an affine motion model. For example, Shi and Tomasi [45] compare patches by using a translational model between temporally neighbour frames where after location estimation obtained from this

procedure, an affine registration between two frames are performed between the patches of the current and base frames. This kind of detect-than-track approaches are suitable for video sequences where locations of the features can be accurately predicted in the next frame.

Another possible feature matching scheme is the detect-and-match approach which is suitable for the cases in which temporal and geometric relations between images are unknown [46] and [47]. For these situations, features can easily appear in different scales and orientations which makes use of view invariant features more important. Some recently developed view invariant features are analysed and their performances are evaluated in [48]. For the usual cases, it is observed that Scale Invariant Feature Transform (SIFT) [38] usually performs the best.

The simplest way of matching features between image pairs is to compare all features of one image with the those of the other image. However, this approach becomes infeasible for some cases as its computational complexity becomes quadratic with the number of the features. As a result, to handle feature matching more efficiently, different indexing schemes which are usually based on finding neigbours in high dimensional spaces are proposed. As an example, a Best-Bin-First (BBF) algorithm is proposed by Beis and Lowe [49]. It should be noted that, efficient detection of feature matches between images is still considered as a problem which is far from being solved [9].

After a set of feature correspondences are computed, the problem is to estimate the alignment parameters from this set of features. A possible approach is to use a least-squares estimation for this task. However, it is possible that there are some false matches between images which can seriously spoil the quality of the estimations especially if a least-squares scheme

is used. For a more robust estimation, it is better to perform some procedures to eliminate these false matches which do not suit to the considered model. There are two widely used solutions to this problem which are known as RANdom SAmple Consensus (RANSAC) [28] and least median of squares (LMS) [50]. For both techniques, first a set of correspondences that are enough to define the model is chosen and model is estimated by using these correspondences. Estimated model is tested on all of the feature correspondences to specify its fitting performance. Residuals of all the features are calculated with respect to the estimated model which is given as:

$$r_i = x'_i - g\left(x_i, p\right) \tag{16}$$

where p is the parameters of the given model that is mapping point x_i to x'_i. For RANSAC, features whose residul norm is within a given interval are assumed to be inliers. Procedure is repeated S times and model with the maximum number of inliers are chosen as the final solution. To ensure that a robust model of the given correspondances are obtained, enough number of trials must be performed. Let the chance of a feature correspondence to be valid is p and P be the total probability of success after S trials. Probability of a trial which uses only inlier features becomes p^k where k is the minimum number of the correspondences needed to estimate the model parameters. Probability of failure to find set of features composed of only inlier features is given as:

$$1 - P = \left(1 - p^k\right)^S \tag{17}$$

As a result, required minimum number of trials needed is given by the fol-

lowing equation:

$$S = \frac{\log(1-P)}{\log(1-p^k)} \tag{18}$$

For LMS, median of the residual norms of a given model is considered. Model which has the smallest median value is chosen to be the final solution.

2.2.3 Advantages of Feature Based Alignment

Feature based alignment methods have become very popular lately as a result of successful keypoint detectors which have very good scale and affine invariant properties. As a result, alignment of the images from completely different point of view and scale become possible which provides robustness to the image alignment process since feature based methods do no not need close initialization as in direct methods.

2.3 Image Mosaicing

Image mosaicing is the process of composing several images of a scene to create a large field of images of the scene. This is done by aligning all the images on the same reference frame by their estimated alignments. Both direct or feature based methods can be used to find the alignments of the images. However, feature based method become popular lately since they have attractive invariance properties which makes mosaicing of images from very different perspectives possible and ability to recognize if two images have common texture [47].

Image mosaicing is possible with different motion models which were detailed in 2.1. Most common motion models which can be used for mosaic-

ing are similarity, affine and homography (a subset of projective transformations). Homography is the most general and popular motion model for image mosaicing since it is the most general linear transformation on the homogeneous image coordinates which is capable of representing perspective distortions between images.

2.3.1 Homography

For two different camera frames, coordinates of the 3D points can be related with a rotation and translation. For the coordinates of a point with respect to the two camera frames, $\mathbf{X}_1$, $\mathbf{X}_2$, coordinate transformation between two frames can be given as:

$$\mathbf{X}_2 = R\mathbf{X}_1 + T \tag{19}$$

This transformation can be expressed as a homogeneous linear transformation when some additional constraints hold. For example, if the camera translation is zero (pure rotational motion), transformation becomes as:

$$\mathbf{X}_2 = R\mathbf{X}_1 \tag{20}$$

where homography is the rotation matrix. Coordinates of the points can also be transformed to each other with a linear transformation for a general euclidean motion when the scene is planar [51]. Let N be the unit normal of the plane with respect to the first camera frame. Distance of the point $\mathbf{X}_1$ to the camera is given as:

$$d = N^\intercal \mathbf{X}_1 = n_1 X + n_2 Y + n_3 Z \tag{21}$$

By using (20) and (21), we obtain

$$\frac{1}{d}N^{\intercal}\mathbf{X}_1 = 1 \tag{22}$$

$$\mathbf{X}_2 = R\mathbf{X}_1 + T \tag{23}$$

$$\mathbf{X}_2 = R\mathbf{X}_1 + T\frac{1}{d}N^{\intercal}\mathbf{X}_1 \tag{24}$$

$$\mathbf{X}_2 = \left(R + T\frac{1}{d}N^{\intercal}\right)\mathbf{X}_1 \tag{25}$$

$$H = R + T\frac{1}{d}N^{\intercal} \tag{26}$$

As a result, mapping between image coordinates between two camera frames can be expressed with a homography for the cases where camera undergoes a pure rotation in a general scene or an Euclidean motion where scene is planar.

2.3.2 Homography Estimation

For a set of inlier point correspondences between two images, a Direct Linear Transformation (DLT) algorithm [15] can be used to compute the homography between these images. Let the mapping between the coordinates of two images be given as:

$$\mathbf{x}'_i = H\mathbf{x}_i \tag{27}$$

Since this is a homogeneous transformation, $\mathbf{x}'$ vector is an up to a scale multiple of $H\mathbf{x}$, relation between these two vector can be expressed by the

following equation:

$$\mathbf{x}_i' \times (H\mathbf{x}_i) = \mathbf{0} \tag{28}$$

as cross product of collinear vectors equal to zero vector. $H\mathbf{x}_i$ can be written as follows:

$$H\mathbf{x}_i = \begin{bmatrix} \mathbf{h}^1\mathbf{x}_i \\ \mathbf{h}^2\mathbf{x}_i \\ \mathbf{h}^3\mathbf{x}_i \end{bmatrix} \tag{29}$$

where $\mathbf{h}^j$ denotes the j^{th} row of H. Cross product in (28) can be written explicitly as:

$$\mathbf{x}_i' \times (H\mathbf{x}_i) = \begin{bmatrix} y_i'\mathbf{h}^3\mathbf{x}_i - w_i'\mathbf{h}^2\mathbf{x}_i \\ w_i'\mathbf{h}^1\mathbf{x}_i - x_i'\mathbf{h}^3\mathbf{x}_i \\ x_i'\mathbf{h}^2\mathbf{x}_i - y_i'\mathbf{h}^1\mathbf{x}_i \end{bmatrix} \tag{30}$$

This expression is decomposed as a matrix vector product as follows:

$$\begin{bmatrix} \mathbf{0} & -w_i'\mathbf{x}_i^\intercal & y_i'\mathbf{x}_i^\intercal \\ w_i'\mathbf{x}_i^\intercal & \mathbf{0} & -x_i'\mathbf{x}_i^\intercal \\ -y_i'\mathbf{x}_i^\intercal & x_i'\mathbf{x}_i^\intercal & \mathbf{0} \end{bmatrix} \begin{bmatrix} h^{1\intercal} \\ h^{2\intercal} \\ h^{3\intercal} \end{bmatrix} = \mathbf{0} \tag{31}$$

Since this is a skew-symmetric matrix, it has two independent rows. After the third row is omitted, equations become:

$$\begin{bmatrix} \mathbf{0} & -w'_i\mathbf{x}_i^\intercal & y'_i\mathbf{x}_i^\intercal \\ w'_i\mathbf{x}_i^\intercal & \mathbf{0} & -x'_i\mathbf{x}_i^\intercal \end{bmatrix} \begin{bmatrix} h^{1\intercal} \\ h^{2\intercal} \\ h^{3\intercal} \end{bmatrix} = \mathbf{0} \tag{32}$$

This equation can be written for all point correspondences where each point gives two independent equations ($A_i h = 0$). By concatenating A_i matrices vertically for n point correspondences, total number of $2n$ equations are obtained where a system of equations are given as :

$$Ah = 0 \tag{33}$$

where size of A is $2n \times 9$. For exact point correspondences, A has a one dimensional nullspace. However, because of the noise involved in the point coordinates, this homogeneous system of equations must be solved by using least-squares where optimization problem is stated as:

$$\min_h \|Ah\|^2 \quad \text{subject to} \quad \|h\|^2 = 1 \tag{34}$$

The solution is found to be the eigenvector of $A^T A$ corresponding to its smallest eigenvalue which can be obtained from Singular Value Decomposition (SVD) of A.

It should be noted that during the estimations, algebraic error is minimized. However, it is more sensible to minimize geometric error since alignment quality is related to this quantity. To decrease the differences between results of algebraic and geometric error minimization, a normalization is

necessary for the pixel coordinates of the images. Normalization can be performed with following steps [15] :

1. Feature coordinates of the first image ($\mathbf{x_i}$) are normalized. First, a translation ($\mathbf{T}$) is performed on all the points which map the centroid of the points to the origin. After this mapping, an isotropic scaling ($\mathbf{S}$) is performed on the points such that average distance of the feature points to the origin is $\sqrt{2}$. Final transformation becomes $\mathbf{K} = \mathbf{ST}$.

2. A similar procedure is also performed for the feature coordinates of the other image ($\mathbf{x}'$). Let transformation applied on these features to be $\mathbf{K}' = \mathbf{S}'\mathbf{T}'$.

3. DLT algorithm is performed on the normalized feature coordinates. Let the estimated homography be $\mathbf{H}_n$. Homography between the original feature coordinates can be recovered as $\mathbf{H} = (\mathbf{K}')^{-1}\mathbf{H}_n\mathbf{K}$.

Another advantage of the normalization is that it provides invariance to the chosen coordinate frame. Normalization is stated as an essential step for homography estimation which should not be thought as optional [15].

After an estimation is performed, it is also important to determine its accuracy. Covariance matrix of the homography can be calculated as:

1. Given the point correspondances for two images ($\mathbf{x}'_\mathbf{i}$ and $\mathbf{x_i}$) where homogeneous feature coordinates are mapped to each other with $\mathbf{x}'_\mathbf{i} = H\mathbf{x_i}$, Jacobian of x' is calculated with respect to the homography parameters for all the correspondances. This can be calculated from (8) and (9).

2. These Jacobians are concatenated vertically and **J** is formed which includes all the individual jacobians. Covariance matrix of the homography is obtained from J by the following equation:

$$\Sigma_H = A\left(A^\intercal J^\intercal \Sigma^{-1} J A\right)^{-1} A^\intercal \tag{35}$$

where A is any 9×8 matrix whose columns are orthogonal to H. Σ is the covariance matrix formed from the covariances of the feature coordinates which is a $2n \times 2n$ matrix. Since we can assume that the components of the feature coordinates are independent from each other, this matrix can be chosen as a multiple of identity (λI).

2.3.3 Homography Decomposition

Relative rotation and translation between two camera frames can be extracted from the estimated homography between images (H_L) [51]. To extract these quantities, homography is normalized with its second largest eigenvalue which is given as:

$$H = \frac{H_L}{\sigma_2\left(H_L\right)} = \pm\left(R + \frac{1}{d} T N^T\right) \tag{36}$$

A sign ambiguity is presented with the normalized homography. This ambiguity is eliminated by imposing positive depth constraint to the equation. For the depth values of the scene ($\lambda_1,\ \lambda_2$) of two camera frames, mapping between camera coordinates of the points are given as:

$$\lambda_1 \mathbf{x}_1 = \pm\lambda_2 H \mathbf{x}_2, \quad \lambda_1, \lambda_2 > 0 \tag{37}$$

Since scene depths take positive values, positive depth constrain can be imposed as follows:

$$\mathbf{x}_2^T H \mathbf{x}_1 > 0 \tag{38}$$

As a result, correct sign of the normalized homography is obtained. To decompose this homography, SVD of $H^T H$ is calculated such that

$$H^T H = V \Sigma V^T \tag{39}$$

$$\Sigma = diag\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right) \tag{40}$$

$$V = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \tag{41}$$

It should be noted that translation can only be extracted up to a scale factor since there is an inherent depth ambiguity in (36). As a result, we can only expect a scaled translation from a homography. To extract $\{R, \frac{1}{d}T, N\}$, the following steps must be followed:

1. $\mathbf{u}_1$ and $\mathbf{u}_2$ vectors are defined as follows:

$$\mathbf{u}_1 = \frac{\sqrt{1-\sigma_3^2}v_1 + \sqrt{\sigma_1^2 - 1}v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}} \tag{42}$$

$$\mathbf{u}_2 = \frac{\sqrt{1-\sigma_3^2}v_1 - \sqrt{\sigma_1^2 - 1}v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}} \tag{43}$$

Table 2.1: Possible Decompositions of the Homograpy

Solution 1	Solution 2	Solution 3	Solution 4
$R_1 = W_1 U_1^T$	$R_2 = W_2 U_2^T$	$R_3 = R_1$	$R_4 = R_2$
$N_1 = \hat{v}_2 u_1$	$N_2 = \hat{v}_2 u_2$	$N_3 = -N_1$	$N_4 = -N_2$
$\frac{1}{d} T_1 = (H - R_1) N_1$	$\frac{1}{d} T_2 = (H - R_2) N_2$	$\frac{1}{d} T_3 = -\frac{1}{d} T_1$	$\frac{1}{d} T_3 = -\frac{1}{d} T_2$

2. U_1, U_2, W_1 and W_2 are defined as follows:

$$U_1 = [\mathbf{v_2}, \mathbf{u_1}, \hat{\mathbf{v}}_2 \mathbf{u}_1] \tag{44}$$

$$W_1 = [H\mathbf{v}_2, H\mathbf{u}_1, (H\mathbf{v}_2) \times (H\mathbf{u}_1)] \tag{45}$$

$$U_2 = [\mathbf{v_2}, \mathbf{u_2}, \hat{\mathbf{v}}_2 \mathbf{u}_2] \tag{46}$$

$$W_2 = [H\mathbf{v}_2, H\mathbf{u}_2, (H\mathbf{v}_2) \times (H\mathbf{u}_2)] \tag{47}$$

3. There are four possible triples $\left(R, \frac{1}{d}T, N\right)$ which results in the same homography. Possible Solutions are given in Table 2.1.

4. The dot product of the unit plane normal with the homogeneous image coordinates ($N^T \mathbf{x}$) is equal to the plane-camera distance which must take a positive value for physically possible cases. At most two of the possible solutions can fulfill this condition. It is also possible that only one of the possible solutions meet this requirement. However, it is not the usual situation [52].

Chapter III

3 A New Approach for Fast and Accurate Mosaicing of Aerial Images

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 a fast and accurate manner. 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. General structure of the proposed method is depicted in Figure 3.1. 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. Our work is accepted to be published as [53].

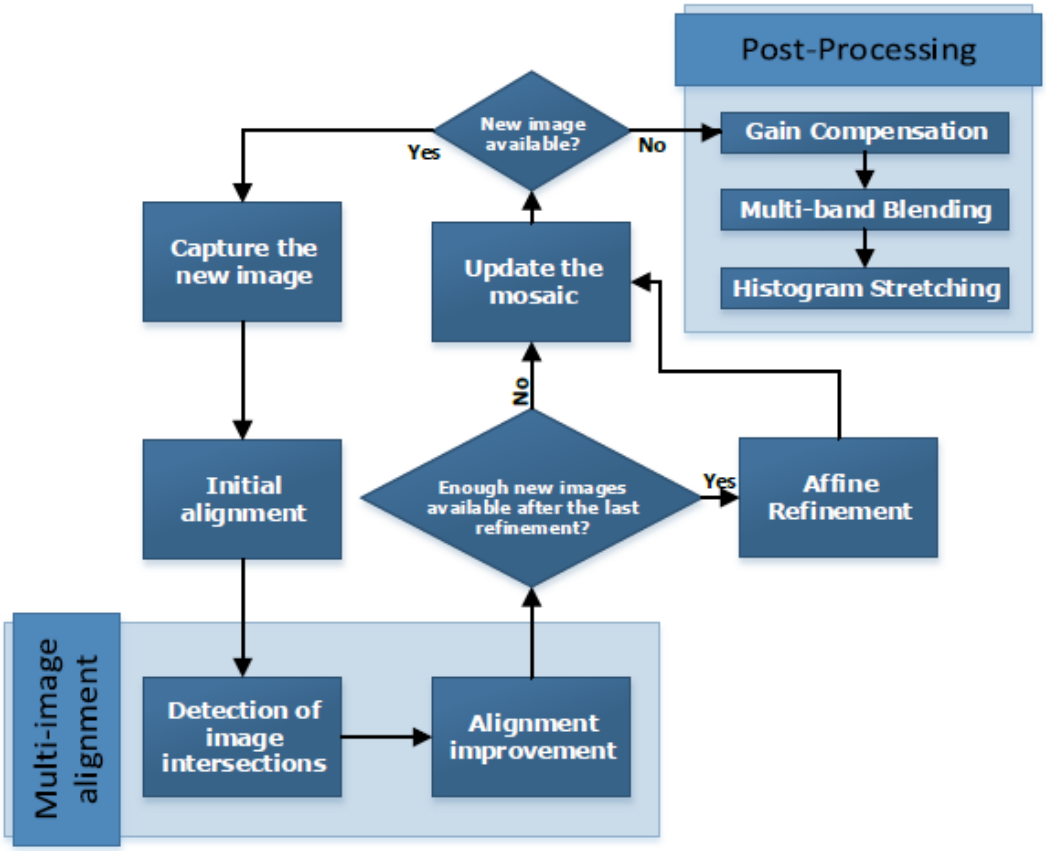


Figure 3.1: General structure of the proposed method

3.1 Image Mosaicing

Image mosaicing includes aligning images which are captured from different camera poses and registering them on a image plane (mosaic plane or reference frame). The easiest way to register images captured 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 $I_0, I_1, I_2..., I_{n-1}$ from a planar scene and pairwise homographies H_{01}, H_{12}, H_{23} ..., $H_{(n-2)(n-1)}$

between image pairs are known where H_{ij} is the homography which aligns I_j to I_i, homography between the new (I_n) and the reference image (I_r) can be calculated as:

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

Although this approach is straightforward, because of its multiplicative nature, errors accumulate with every new image which causes a drift in the mosaic in time. Drift of the images in the mosaic are depicted in Figure 3.2. Since a Normalized Direct Linear Transformation (NDLT) algorithm is used

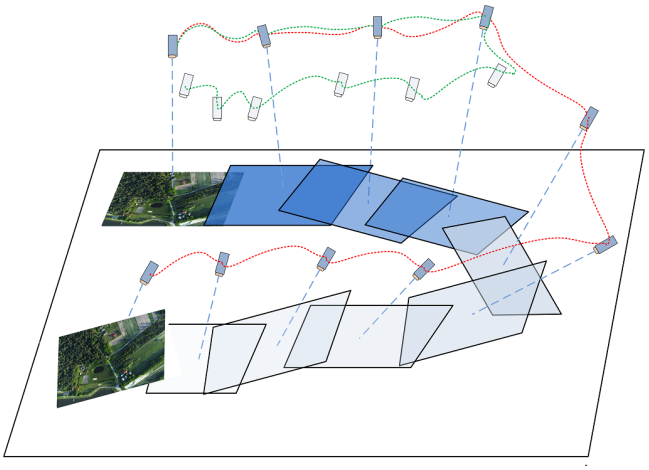


Figure 3.2: Drift caused by estimation errors. UAV returns to the same area and snaps the same image from the initial position. True and estimated trajectories are shown with green and red dashed curves respectively.

during the estimations of the the pairwise homographies, minimization of the algebraic error is sufficient [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 of 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.

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 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.

3.2 Proposed Mosaicing Approach

The homography estimation process discussed in Section 3.1 includes the estimation between two images. However, estimating the homography by using only the previous image can lead to errors in 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 'Seperating Axis Theorem' to detect the previous images intersecting the new image since only aligned images intersecting each other are supposed to have common features.

3.2.1 Detection of Image Intersections by Using Seperating Axis Theorem

Separating Axis Theorem (SAT) is a popular tool in computer graphics which can be used to detect collisions between objects [54]. 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:

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

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 Figure 3.3.

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:

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 Eqn. (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 Eqn. (5) is checked for all $\vec{d}$ directions.

4. If there exist 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.

Using SAT provides efficient operations in the proposed approach. However, it should be noted that the number of the images increases linearly with

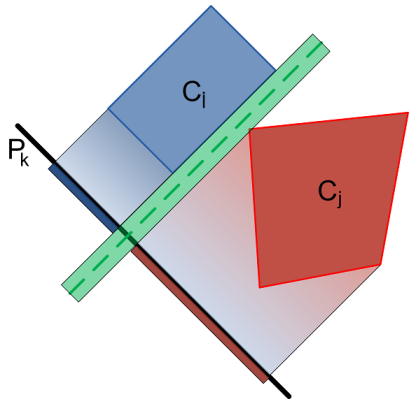


Figure 3.3: An illustration of SAT. For a separating axis P_k, projected convex sets do not intersect.

the number of the previous images in the mosaic. This might pose some problems to the algorithm when number of the images in the mosaic takes larger values. It is possible to further reduce number of the images used with SAT. For example, a sorting algorithm can be employed to sort the location of the aligned images in the mosaic. Every new image can be added to this list with a logarithmic computational complexity. Assume that we obtain a new image which is on the right side of a previous image in the sorted list and know that it does not intersect this previous image. We can directly eliminate a large number of other images in the list which stay on the left side of this previous image (the ones approximately at the same level with it in the up-down direction). This can dramatically reduce the number of the needed trials. In our experiments, we did not utilize such an approach since SAT required negligible amount of computational power even for very large number of images.

3.2.2 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:

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

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 MLE-SAC [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.

3.2.3 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},...,$H_{r(n-1)}$ 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 increas-

ing 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. [55] 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. [56] use affine model for image motions and update all the parameters at each time step via recursive least-squares estimation. Pizarro and Singh [7] 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 [57] and Davis [58] propose partial global optimization procedures similar to ours in their estimations for robotics and mosaicing applications, respectively. Sibley [57] 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 [58] 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 homogra-

phies 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 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 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. Equations (7), (8) and (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.

3.3 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 [59]. 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.

3.3.1 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 [1]. 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. Gains of the images are obtained from the minimization of the cost function which can be solved in closed form.

3.3.2 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

(a) Czyste

(b) Munich Quarry

(c) Savona Highway

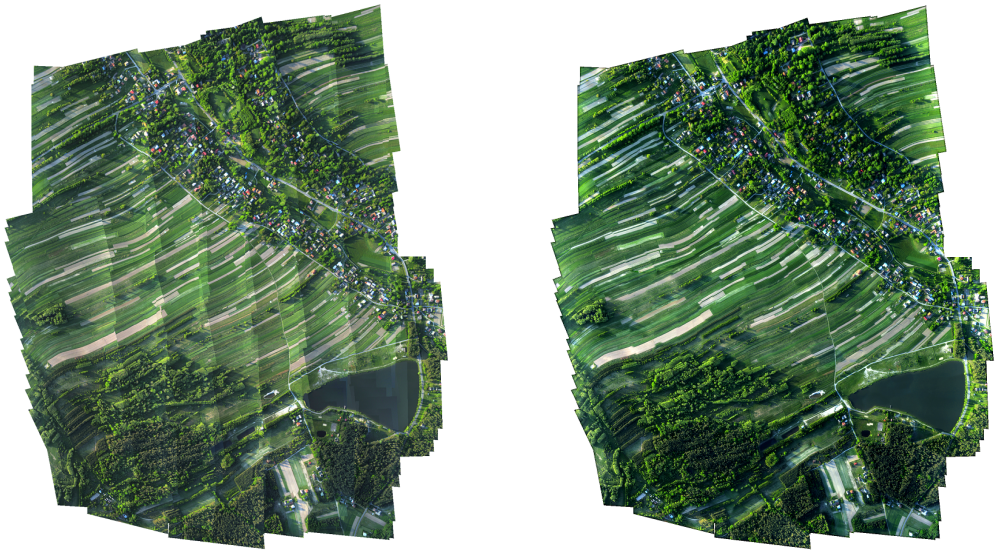
Figure 3.4: Sample images from the aerial image datasets.

the mosaic image with multi-band blending algorithm [59] by which we aim to attenuate these visual artifacts. Algorithm given in Brown et al. [1] is used to blend the mosaic image.

3.4 Experimental Results

We tested our mosaicing approach on the images of three publicly available datasets. These are Czyste [60], Munich Quarry [61] and Savona Highway [61]. A set of sample images selected from these datasets are depicted in Figure 3.4.

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.



(a) Before Postprocessing (b) After Postprocessing

Figure 3.5: Mosaic image for the Czyste image sequence before and after postprocessing

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 Figure 3.5 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. [56] where 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

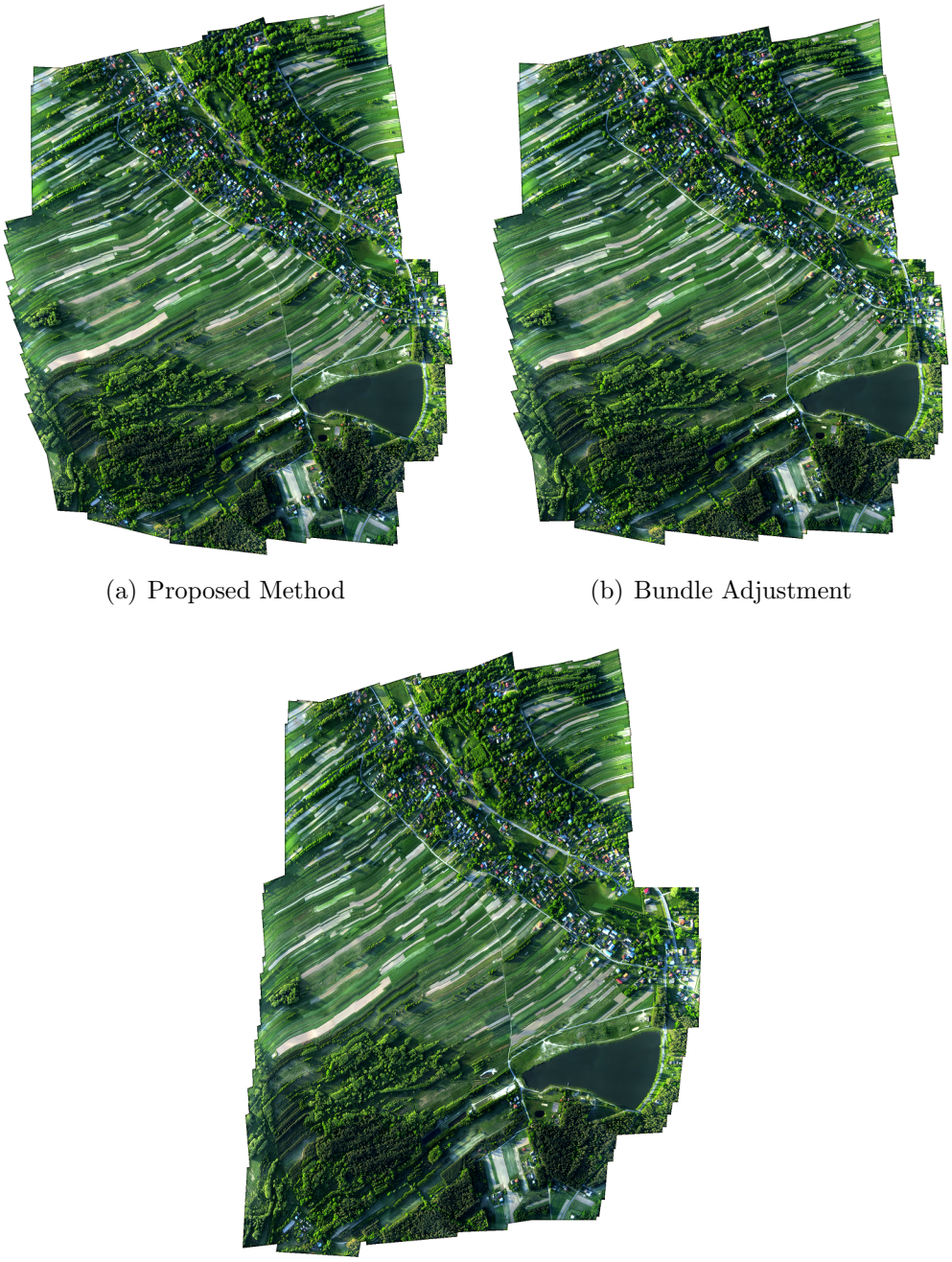
estimated by minimizing the total feature reprojection error between image pairs. This nonlinear optimization problem is solved using Levenberg-Marquard algorithm.

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 Figure 7, Figure 3.7, Figure 3.8, respectively. It is apparent from Figure 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 Figure 3.7 and 3.8, 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.

3.4.1 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.

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

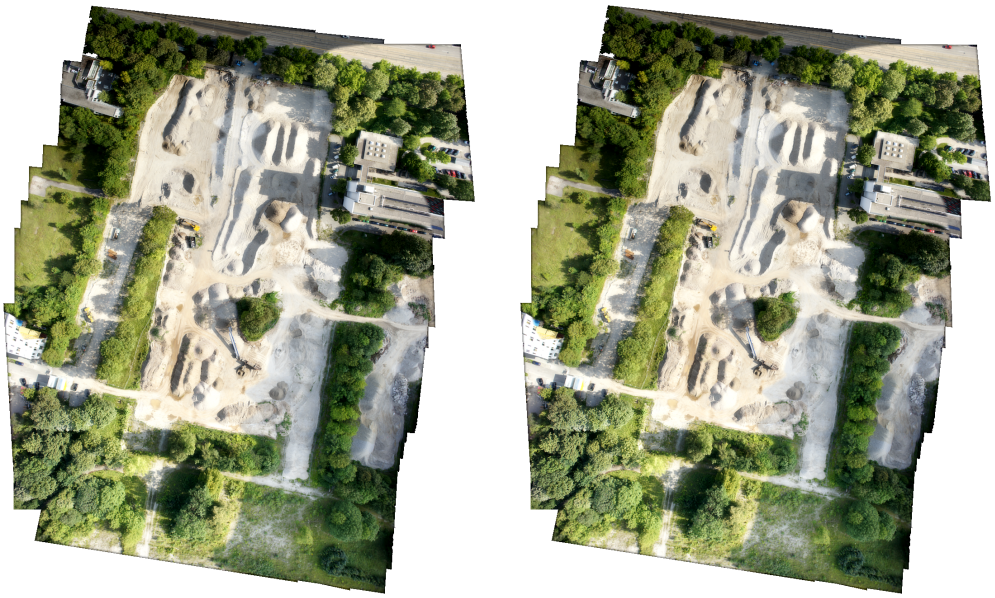


(a) Proposed Method

(b) Bundle Adjustment

(c) Gracias' Method

Figure 3.6: Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Czyste image sequence



(a) Proposed Method

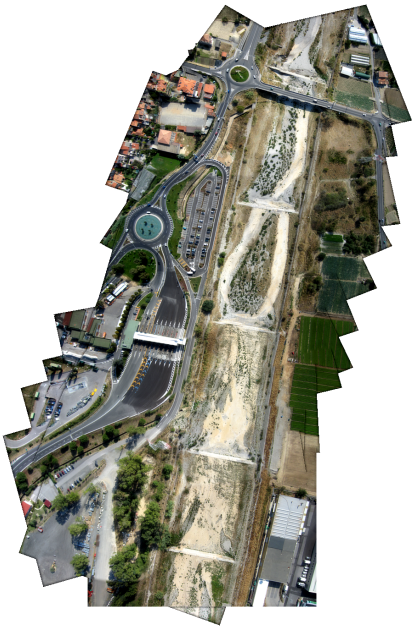
(b) Bundle Adjustment

(c) Gracias' Method

Figure 3.7: Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Munich Quarry image sequence

(a) Proposed Method

(b) Bundle Adjustment



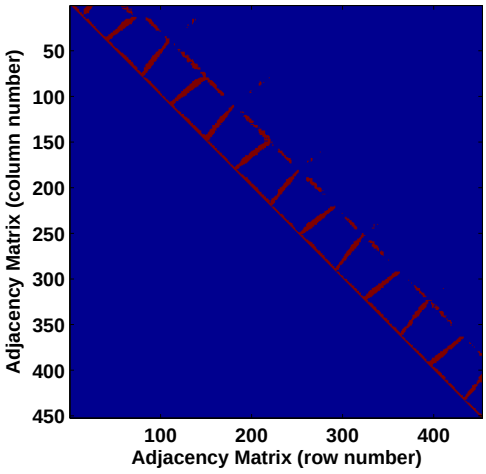
(c) Gracias' Method

Figure 3.8: Mosaic images of the proposed method, the bundle adjustment and Gracias' method for Savona Highway image sequence

Table 3.1: RMS values for the four cases in Czyste image sequence

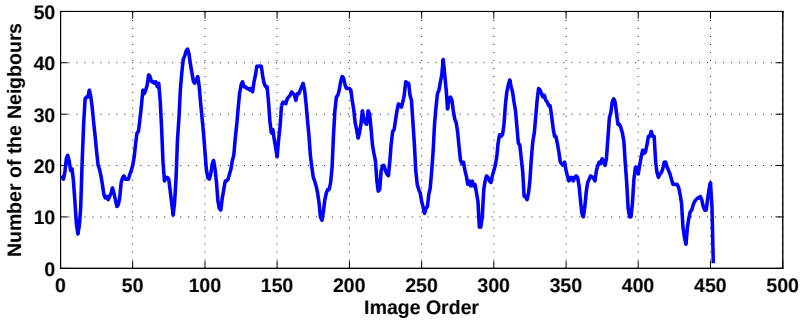
Case	**Total Matches**	**RMS(pix)**
Algorithm (without Affine Refinement)	645272	4.4520
Algorithm (with Affine Refinement)	645272	3.9971
Gracias et al. (2004)	645272	3.8636
Bundle Adjustment	645272	0.8390

given in Figure 3.9. Matched image pairs are depicted as red points in the adjacency matrix which is shown in Figure 9(a). The number of matching images versus image indices is plotted in Figure 9(c). Camera trajectory for the dataset is sketched in Figure 9(b). Total number of 645272 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 the Table 3.1. 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.



(a) Adjacency Matrix

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



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

Figure 3.9: Visual and numerical presentations of the spatial image relations in Czyste

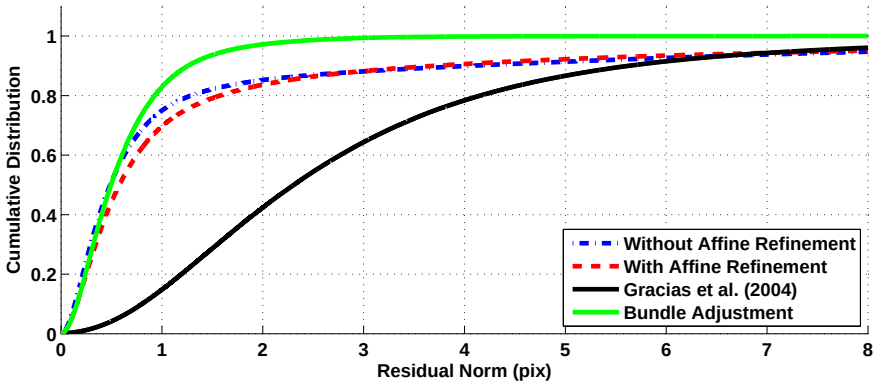


Figure 3.10: Cumulative distribution of the residual error norms for Czyste image sequence

Cumulative distributions of the error for all the methods are plotted in Figure 3.10. It can be observed from Figure 3.10 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.

For the Munich Quarry image sequence, 56 images are used during the experiments. Spatial relations of the images in the mosaic are depicted in Figure 3.11. RMS values are given in Table 3.2 for different methods. It can be inferred from the table that RMS values for the proposed approach

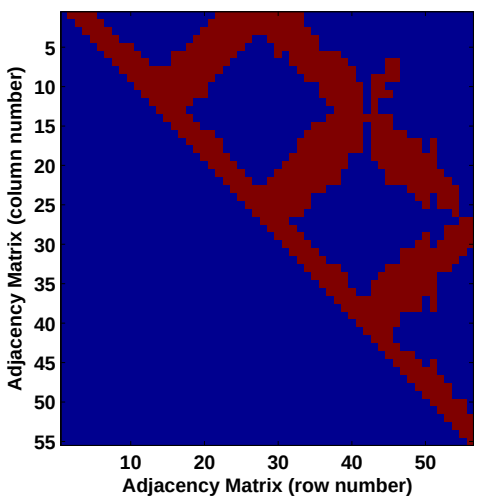
Table 3.2: RMS values for the four cases in Munich Quarry image sequence

Case	**Total Matches**	**RMS(pix)**
Algorithm (without Affine Refinement)	69149	1.3497
Algorithm (with Affine Refinement)	69149	1.2676
Gracias et al. (2004)	69149	3.1742
Bundle Adjustment	69149	1.2185

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 Figure 3.12. 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.

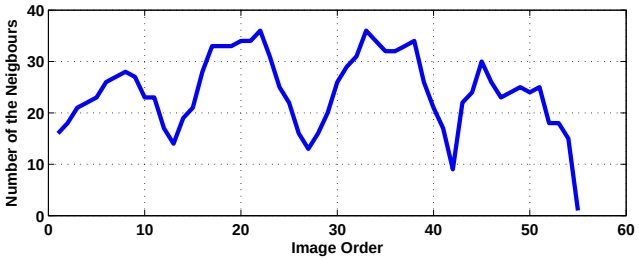
30 images are used during the experiments. Spatial relations of the images are given in Figure 3.13. RMS values are provided in Table 3.3 for different methods. It can be inferred from the table that performance of the proposed approach with affine refinement is nearly equal to the results of the bundle adjustment. Gracias' 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 Figure 3.14. 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 Figure 3.8 (see Section 3.1).

It is apparent from the visual and numerical results that numerical results can be quite different for various cases where visual differences are negligible.



(a) Adjacency Matrix

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



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

Figure 3.11: Visual and numerical presentations of the spatial image relations in Munich Quarry

Table 3.3: RMS values for the four cases in Savona Highway image sequence

Case	**Total Matches**	**RMS(pix)**
Algorithm (without Affine Refinement)	72509	1.4402
Algorithm (with Affine Refinement)	72509	1.2252
Gracias et al. (2004)	72509	4.4611
Bundle Adjustment	72509	1.2137

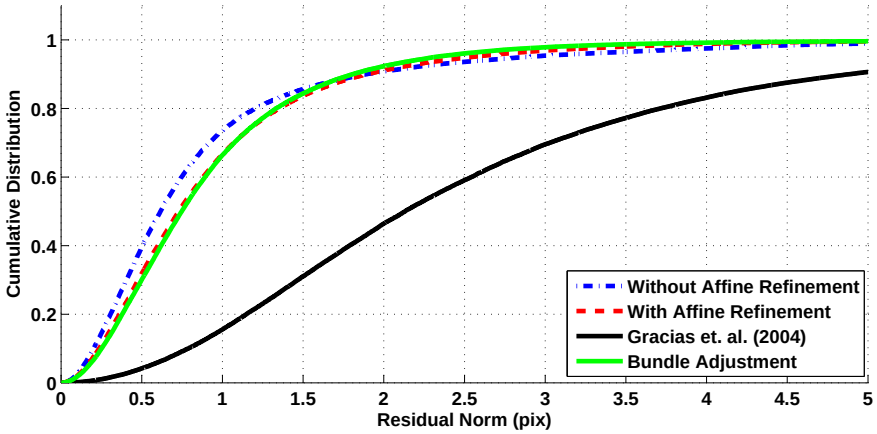
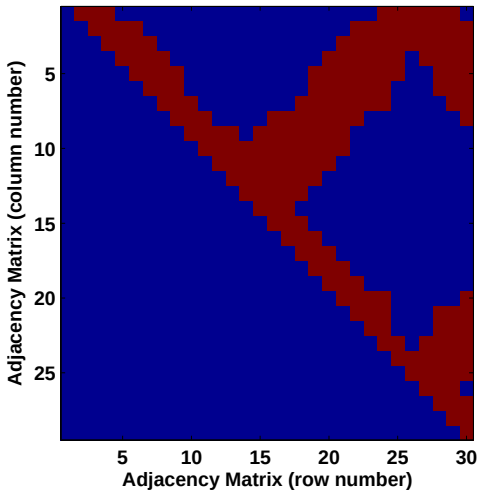


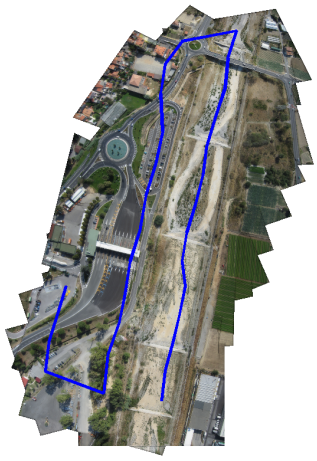
Figure 3.12: Cumulative distribution of the residual error norms for Munich Quarry image sequence

As a result it can be questioned that if using RMS error of the point features as the performance metric is a good idea. Since we mostly observe the seams of the mosaic at the edges of the shapes, using the property of these edge-like structures, e.g. line or curve continuity, could provide a better measure for the mosaic quality. We did not use such a metric for two main reasons. First, new generation of point features, e.g. SIFT or SURF, are usually detected in large numbers and well-spreaded to the whole image which implies that all parts of the scene are represented approximately in equal weight. Second, using edge-like features can be tricky in the sense of feature description and matching because of some well known problems they suffer from, e.g. aperture problem and weak invariance to the point of view changes.

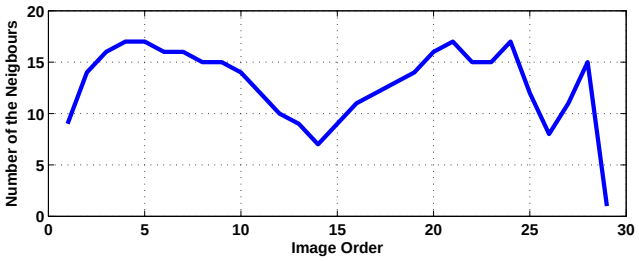
It should be noted that the improvement achieved by affine refinement will be useful for cases where navigational requirements are more stringent. However, it can be deactivated for cases where only visual appearance is the



(a) Adjacency Matrix



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



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

Figure 3.13: Visual and numerical presentations of the spatial image relations in Savona Highway

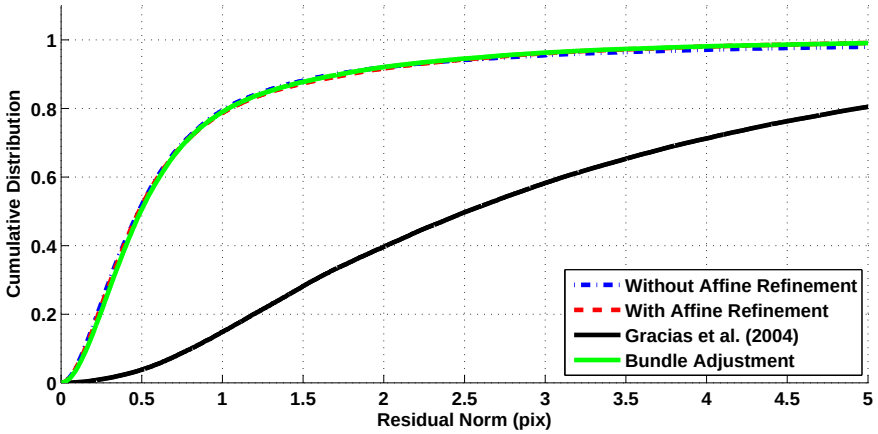


Figure 3.14: Cumulative distribution of the residual error norms for Savona Highway image sequence

prime concern.

Chapter IV

4 Pose Estimation Based Image Mosaicing via Extended Kalman Filter

In the previous chapter, we proposed a sequential mosaicing approach where new images were aligned to the mosaic by using the alignments of all the previous images intersecting the new image. This provides a good consistency to the mosaic as all the possible information available in the mosaic is considered during the operation. However, since previous images are fixed and alignment of the new image is performed under this condition, only locally optimal results can be obtained. It is clear that, a joint estimation including the new image and all the previous images would yield more successful results in the sense of global consistency. However, this is not a scalable option as operation time for estimating the alignments of all the previous images increases with the number of the images. To strike a bargain between the global consistency and computational expense, we introduced a partial global estimation where only alignments of a limited number of recent images were updated. Experimental results validated that this procedure brought some extra global consistency to the mosaic which is apparent from the decrease in the RMS values of feature reprojection errors presented in the related section.

However, it is possible to use recursive estimation techniques available in the literature for image mosaicing where their recursive nature provides a

computationally efficient estimation. It is also possible to update the alignment parameters of all the images in every step of the algorithm in a scalable manner by this option. There are some studies in the literature where these kinds of estimation schemes are employed. Gracias et. al [56] propose an RLS (recursive least-squares) filter approach for mosaicing of underwater images. An affine mosaicing approach is considered in this study to get a linear estimation. Since estimations are directly based on the minimization of the vector composed of the feature reprojection errors, scalability of this approach with the number of images can be questionable. An EKF based estimation of the pairwise homographies between image pairs are proposed in [17]. In this study, an EKF loop is employed for every image set which forms a loop. After the loop is closed, all of the pairwise homographies between consecutive images (members of the loop) are updated via EKF. Error is propagated to all of the pairwise homographies in this manner. Problem of this approach is that estimation updates are limited to the images that are the members of the loop where the estimation lacks a full-state covariance matrix including the stochastic relations between all of images available in the mosaic. Such an approach is proposed in Civera et. al [18]. In this work, a Simultaneous Localization and Mapping (SLAM) based approach is proposed for the mosaicing of the images captured from a pure rotational camera. State vector includes the pose of the last camera and global coordinates of all features extracted from the images. It is reasonable to use feature parameters directly for the pure rotational camera motion since a limited number of features are available for this case. However, this approach is not suitable for planar scene mosaicing since the number of available features can be unbounded. As a result, this method lacks scalability for aerial image

mosaicing applications. We also propose a new method based on recursive parameter estimation. In our method, estimation of a full state vector and its covariance matrix imply more accurate results. Also for each image, we only need to add six new parameters to the estimation which makes our method more scalable than all the available studies in the literature.

We develop a novel method for creating image mosaics of quasi-planar scenes based on Extended Kalman Filter (EKF) framework. It includes a state space approach where the state vector is composed of the scene normal and camera extrinsic parameters (rotation and translation). A joint estimation is performed on all the image parameters when a new image is included into the estimation. This is handled with a low computational effort thanks to the efficient nature of the EKF update equations and sparse structure of the spatial image relations. Utilization of EKF provides a good global consistency between images since it can handle the accumulated error at the loop closing regions by propagating the error to the whole mosaic. Sparse nature of image relations implies small size measurement equations which provide a computationally efficient operation and make real-time operation possible. We tested our algorithm on some publicly available datasets. Results are promising both visually and numerically. Our study will appear as [62].

4.1 Proposed Approach

We use classical EKF loop to update the mosaic with every new image. State vector includes scene parameters and global camera poses which are obtained from the relative rotation, relative translation and plane normal parameters extracted from pairwise homographies between image pairs. Rotations are parameterized with a vector of Euler angles which is denoted with $\mathbf{\Phi}_i =$

$\begin{bmatrix}\gamma_i & \beta_i & \alpha_i\end{bmatrix}^\intercal$ (for the parameters of i^{th} image, I_i). Rotation matrix related to $\mathbf{\Phi}_i$ can be expressed as:

$$R_i = {}^{i}_{1}R = \begin{bmatrix}\cos\alpha_i & -\sin\alpha_i & 0\\ \sin\alpha_i & \cos\alpha_i & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\cos\beta_i & 0 & \sin\beta_i\\ 0 & 1 & 0\\ -\sin\beta_i & 0 & \cos\beta_i\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\gamma_i & -\sin\gamma_i\\ 0 & \sin\gamma_i & \cos\gamma_i\end{bmatrix} \tag{1}$$

which encodes the rotation of the first camera frame with respect to the i^{th} camera frame. Translation parameters of the camera frames are also included in the estimation. However, we use scaled camera translations in our estimations since we can only expect a scaled translation from a homography. For I_i, scaled translation is denoted as $\mathbf{t}_i$ which is a three parameter vector representing the translation of the first camera frame with respect to the i^{th} camera frame. Scene is modeled as a plane and to represent this plane, two parameters, θ and ψ, angle of the plane normal with respect to the first camera frame, are used. Unit normal vector of this plane can be written in terms of these parameters as:

$$\mathbf{n} = {}^{1}n = \begin{bmatrix}\sin\psi\sin\theta\\ \sin\psi\cos\theta\\ \cos\psi\end{bmatrix} \tag{2}$$

where $\mathbf{n}$ is the plane unit plane normal with respect to the first camera frame. State vector of the EKF after I_i is included to the estimation can be defined

as:

$$\mathbf{x} = \left[\theta, \quad \psi, \quad \mathbf{\Phi}_2^\intercal, \quad \mathbf{t}_2^\intercal, \quad \ldots \quad \mathbf{\Phi}_i^\intercal, \quad \mathbf{t}_i^\intercal\right]^\intercal \tag{3}$$

Proposed algorithm can be outlined as follows:

1. To include a new image in the estimation, its pairwise homography with the previous image is estimated (denoted with H_{ij} for the homography between new image i and previous image j).

2. Relative rotation $\left({}_j^iR\right)$ and scaled translation $({}^it_{ij})$ are extracted from this homography and used to initialize new state vector variables $(\mathbf{\Phi}_i, \mathbf{t}_i)$.

3. By using the approximate location of the new image in the mosaic, which will be determined during the prediction step, previously aligned images which intersect the new image are identified. Homography estimation is performed between these images and the new image. These pairwise homographies are utilised as the measurements of the estimation.

4. State vector is updated via EKF update equations.

A flowchart of the proposed method is given in Figure 4.1.

4.1.1 Prediction

To include a new image to the estimation process, its approximate location in the mosaic must be predicted. This is achieved by a homography estimation performed between the new and previous image. Relative pose of the camera where new image is captured can be extracted from this pairwise homography. As the state parameters for the previous image are known from

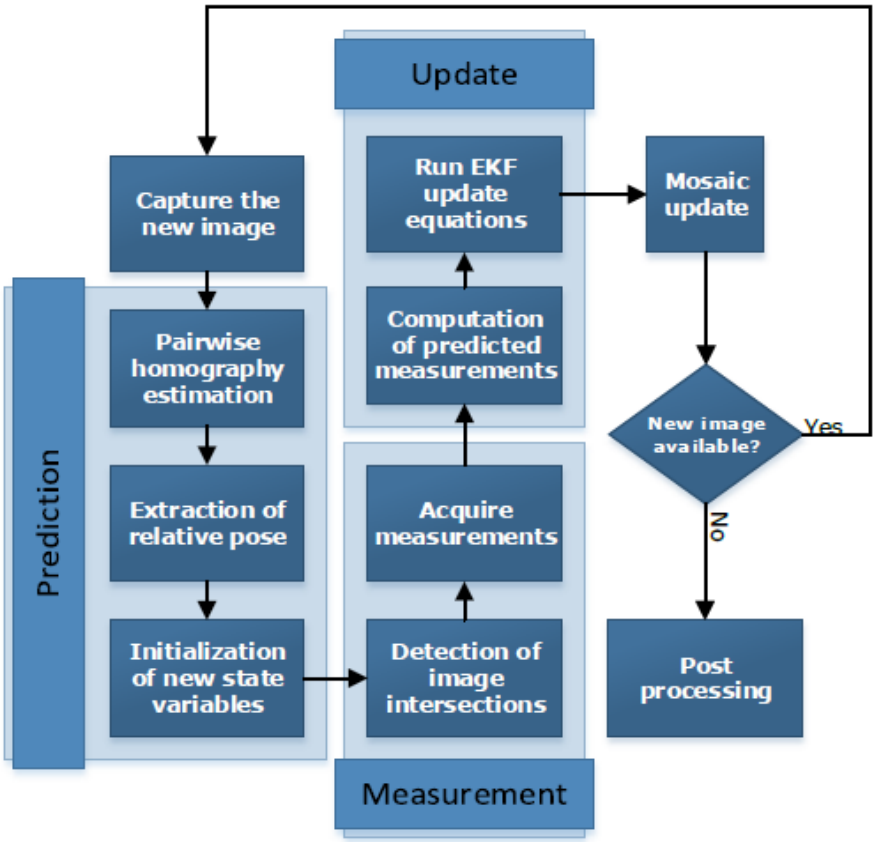


Figure 4.1: Flowchart of the proposed method

the previous time step, predicted parameters can be obtained for the new camera frame. The predicted camera orientation, $\hat{\mathbf{\Phi}}_i^-$ is extracted from $\hat{R}_i$, which is computed as:

$$\hat{R}_i = {}^{i}_{j}\hat{R}\,{}^{j}_{1}\hat{R} \tag{4}$$

where ${}^{i}_{j}\hat{R}$ is the relative rotation between i^{th} and j^{th} camera frame extracted from H_{ij}. As the relative translation extracted from the pairwise homography is a scaled translation, a small adjustment is necessary to calculate the

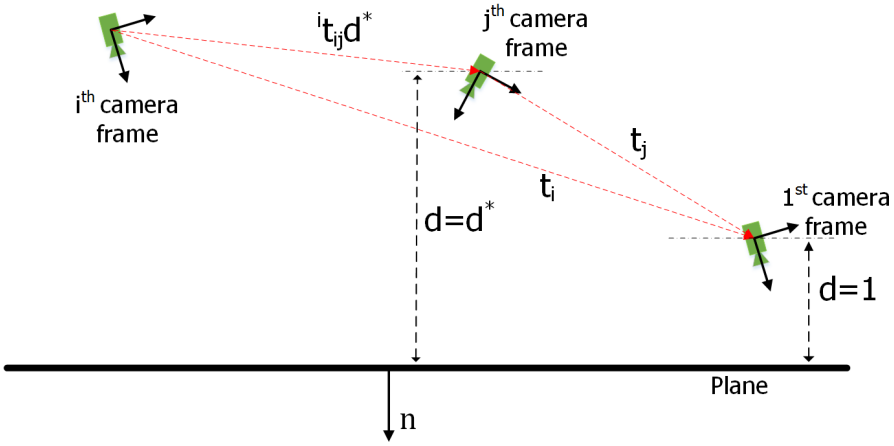


Figure 4.2: Initialization of the new image parameters from the previous image

predicted value of scaled translation ($\hat{\mathbf{t}}_i^-$) which is given as:

$$\hat{\mathbf{t}}_i^- = {}_j^i\hat{R}\mathbf{t}_j + {}^i\hat{t}_{ij}d^* \tag{5}$$

$$d^* = 1 + \mathbf{n}^\intercal R_j^\intercal \mathbf{t}_j \tag{6}$$

where $\mathbf{t}_j$, R_j are the translation and rotation of the previous camera frame obtained from state vector. Camera-plane distance is assumed to be unity for the first camera frame. Related variables and important quantities are depicted in Figure 4.2. After $\hat{\mathbf{\Phi}}_i^-$ and $\hat{\mathbf{t}}_i^-$ are obtained and included into the state vector, covariances of these predicted states are also needed. To find the covariances of these parameters, Jacobian of these parameters with respect to the state vector and relative pose parameters must be calculated.

Jacobian can be given as:

$$J = \frac{\partial \begin{bmatrix} \hat{\mathbf{\Phi}}_i^- & \hat{\mathbf{t}}_i^- \end{bmatrix}}{\partial \begin{bmatrix} \mathbf{x}_{old} & \mathbf{\Phi}_{ij} & {}^i\mathbf{t}_{ij} \end{bmatrix}} \tag{7}$$

where $\mathbf{\Phi}_{ij}$ and ${}^i\mathbf{t}_{ij}$ are the pairwise rotation and translation parameters obtained from pairwise homography. This calculation can be performed easily by using (4) and (5). By using this Jacobian, new covariance matrix of the state vector is computed as:

$$P_{\mathrm{k}} = \left[\begin{array}{c|c} I & \mathbf{0} \\ \hline & J \end{array}\right] \begin{bmatrix} P_{\mathrm{k,old}} & \mathbf{0} \\ \mathbf{0} & C_{ij} \end{bmatrix} \left[\begin{array}{c|c} I & \mathbf{0} \\ \hline & J \end{array}\right]^{\mathsf{T}} \tag{8}$$

where $P_{\mathrm{k,old}}$ is covariance matrix obtained from the previous time step and C_{ij} is the covariance matrix of the new parameters. C_{ij} is assumed to be a multiple of identity, i.e λI in the estimations.

4.1.2 Measurement

Pairwise homography between a new and the previous image is calculated and used to initilize $\mathbf{\Phi}_i$ and $\mathbf{t}_i$. However, it is also possible that the new image has some common features with some other previously aligned images in the mosaic. To provide a better global consistency to our estimation, we should also include the pairwise homographies between the new and these previous images to the estimation. Homographies of these images are incorporated to the estimation as measurements.

It is necassary to detect the previous images which have common features with the new image (intersecting the new image) efficiently since performing

matching trials for all the past images would be computationally expensive. List of the previous images is narrowed by using Separating Axis Theorem (SAT) which is previously explained. Pairwise homographies between the new image and previous images are used to construct measurements. Before using them as measurements, all the estimations are normalized since homography is a redundant parametrization where an arbitrary nonzero multiple of the transformation implies the same transformation. Our normalized homography (h_{ij}) can be given by the following equation:

$$h_{ij} = \text{sgn}\left(\det H_{ij}\right) \frac{H_{ij}}{\|H_{ij}\|} \tag{9}$$

where $\|.\|$ is the Frobenious norm. Covariance matrices of the pairwise homographies are also required for estimation. However, we need to calculate the inverse of the covariance matrices during the inversion of the innovation covariance. Because of the redundant nature of a homography, its covariance matrix is not invertible. As a result, a linear transformation on h_{ij} is utilized to construct measurements by which inversion of the innovation covariance is possible. We choose our measurements as:

$$z_{ij} = A_{ij} h_{ij} \tag{10}$$

where A_{ij} is a 8×9 matrix whose rows are orthogonal to h_{ij} and each other. To obtain the covariance matrix for our measurement h_{ij}, the procedure detailed in Chapter 2. Estimation is performed under the two steps.

1. Jacobian of the feature matches are calculated with respect to the measurement parameters. During the calculations, features coordinates of the first image (x) is assumed to be correct and error is assumed to be

only in the second image. For example, let the homography matrix be h_{ij}. For i^{th} feature match, using the equation $X_i' = h_{ij}X_i$ where X' and X are homogeneous coordinates of x' and x, Jacobian of the feature match with respect to the measurement is calculated from:

$$J_i = \frac{\partial x'}{\partial h_{ij}} \frac{\partial h_{ij}}{\partial z_{ij}} \text{ where } \frac{\partial h_{ij}}{\partial z_{ij}} = A_{ij}^{\intercal} \tag{11}$$

For every feature match, J_i is calculated and by concatenating these jacobians J is found to be $J = \left(J_1^{\intercal}, J_2^{\intercal}, \ldots, J_n^{\intercal}\right)^{\intercal}$ for n feature matches, which is size of $2n \times 9$.

2. Covariance matrix of the measurement is given as:

$$\Sigma_{z_{ij}} = \left(J^{\intercal}\Sigma_{x'}^{-1}J\right)^{-1} \tag{12}$$

where $\Sigma_{x'}$ is a block diagonal matrix whose diagonal elements are the covariance matrices of the feature coordinates. We take this matrix as identity for all of our estimations since we can assume that feature errors are independent.

4.2 Update

Measurements are used to update the predicted state estimates $\left(\hat{x}_k^-\right)$ obtained from the prediction step. Assuming that there are n measurements acquired from the pairwise homography estimates, the measurement vector (z) and

its covariance matrix (R_z) are defined as:

$$z = \left(z_{i1}^{\intercal}, \quad z_{i2}^{\intercal}, \quad \ldots, \quad z_{in}^{\intercal} \right)^{\intercal} \tag{13}$$

$$R_z = \mathrm{diag}\left(C_{z_{i1}}, C_{z_{i2}}, \ldots, C_{z_{in}} \right) \tag{14}$$

Predicted homographies $\hat{h}^-$ can be calculated from the predicted state estimates as:

$$\hat{h}^- = g\left(\hat{\mathbf{x}}_k^- \right) \tag{15}$$

where g is the nonlinear homography function. Assume we want to calculate the predicted homography between i^{th} and j^{th} camera frames (H_{ij}). We use the parameters $\mathbf{n}$, $\hat{\mathbf{\Phi}}_i^-$, $\hat{\mathbf{t}}_i^-$, $\hat{\mathbf{\Phi}}_j^-$, $\hat{\mathbf{t}}_j^-$ which are available in the state vector. $\hat{R}_i^-$ and $\hat{R}_j^-$ is obtained from $\hat{\mathbf{\Phi}}_i^-$ and $\hat{\mathbf{\Phi}}_j^-$, respectively. Predicted homography is given as follows:

$$\hat{R}_{ij}^- = \hat{R}_i^- \left(\hat{R}_j^- \right)^T \tag{16}$$

$$\hat{t}_{ij}^- = \frac{\hat{\mathbf{t}}_i^- - \hat{R}_{ij}^- \hat{\mathbf{t}}_j^-}{1 + \mathbf{n}^T \left(\hat{R}_j^- \right)^T \hat{\mathbf{t}}_j^-} \tag{17}$$

$$\hat{n}_{ij}^- = \hat{R}_j^- n \tag{18}$$

$$\hat{H}_{ij}^- = \hat{R}_{ij}^- + \hat{t}_{ij}^- \left(\hat{n}_{ij}^- \right)^T \tag{19}$$

$$\hat{h}_{ij}^- = \mathrm{sgn}\left(\det \hat{H}_{ij}^- \right) \frac{\hat{H}_{ij}^-}{\| \hat{H}_{ij}^- \|} \tag{20}$$

Predicted homographies are also transformed by the same transformation matrices used to transform pairwise homographies as in (10); i.e.

$$\hat{z} = A\hat{h}^{-} \tag{21}$$

$$A = \text{diag}(A_{i1}, A_{i2}, \dots,) \tag{22}$$

Update equations for the Kalman filter are given as:

$$S_k = Z_k P_k^{-} Z_k^{\intercal} + R_z \tag{23}$$

$$K_k = P_k^{-} Z_k^{\intercal} S_k^{-1} \tag{24}$$

$$x_k = x_k^{-} + K_k \left(z - \hat{z}\right) \tag{25}$$

$$P_k = \left(I - K_k Z_k\right) P_k^{-} \tag{26}$$

where Z_k is the jacobian of the measurement function with respect to the state variables calculated at $x = \hat{x}_k^{-}$. It can be obtained by using chain rule as:

$$Z_k = \frac{\partial z}{\partial x} = \frac{\partial z}{\partial g}\frac{\partial g}{\partial x} = A\frac{\partial g}{\partial x} \tag{27}$$

4.3 Mosaic Creation

Mosaic image can be obtained from the homographies obtained from the state variables. Assume that we want to calculate the homography between I_i and the first image. In terms of state vector parameters, homography

between I_i and the first image can be given as:

$$G_{i1} = K H_{i1} K^{-1} \tag{28}$$

$$= K \left(R_i + \mathbf{t}_i \mathbf{n}^\intercal \right) K^{-1} \tag{29}$$

which maps the points from the first image to I_i. To create the mosaic image, we must align all the images on a common reference plane. For example, to align I_i on the first image, we need G_{1i} which is the inverse of G_{i1} calculated in (29).

It is straightforward to align all the images on the first image since our state vector parameters are in terms of the parameters of the first camera frame ($\mathbf{n} = {}^1n$, $R_i = {}^i_1R$, $\mathbf{t}_i = {}^it_1$). However, a more reasonable idea is to align the images on a virtual reference frame in which the plane normal has only z-axis component (${}^vn = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\intercal$). We choose this new camera frame as its origin is coincident with the origin of the first camera frame. It is necessary to determine its rotation with respect to the first camera frame. Since we know the plane normal with respect to the first camera frame and want it to be mapped to another frame as ${}^vn = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\intercal$, we need to determine the rotation matrix which maps these two vectors to each other. Rodrigues' formula is used to determine this rotation matrix between two

vectors.

$$^{v}n = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\intercal} \tag{30}$$

$$\theta = \operatorname{atan2}\left(\mathbf{n}^{\intercal} \cdot {}^{v}n, \|n \times {}^{v}n\|\right) \tag{31}$$

$$k = \left(n \times {}^{v}n\right) / \sin(\theta) \tag{32}$$

$$R_{\delta} = I + \Big[k\Big]_{\times} \sin\theta + \Big[k\Big]_{\times}^{2} \left(1 - \cos\theta\right) \tag{33}$$

We express the state variables in terms of this virtual camera frame. Rotation parameters of i_{th} camera frame is transformed as:

$$R_i^{new} = R_i R_{\delta}^{\intercal} \tag{34}$$

Translation is not changed since virtual camera frame is coincident with the first camera frame. As a result, image homographies which transfer the points from the virtual camera frame to the i_{th} camera frame are found as:

$$H_{iv} = R_i^{new} + \mathbf{t}_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^{\intercal} \tag{35}$$

We again use the inverse of the homography calculated in (35) to align the images on our virtual camera plane.

4.4 Experimental Results

We tested our algorithm on some aerial image databases in which scenes are quasi-planar. We use the SIFT features [38] extracted from the images during the experiments. Bundle adjustment results are also obtained for the same

Table 4.1: RMS values for Small Village

Case	**Total Matches**	**RMS(pix)**
Proposed Approach	891409	3.2703
Bundle Adjustment	891409	2.8836

images. Optimization is run only on the homography parameters and not on the camera intrinsic parameters since we assume a calibrated camera where these parameters are constant. During the optimization, homographies of the images which minimize sum of squares of the feature reprojection errors between images are obtained by using Levenberg-Marquard algorithm. Root mean square (RMS) value for the residual norms (norm of feature reprojection error) is given in all experiments for both the proposed method and the bundle adjustment. Visual results are also provided for both cases. Results are also blended by using a multi-band blending [59] technique detailed in [1] to get higher quality mosaics.

4.4.1 Small Village Image Sequence

Image sequence is provided in [63]. Total number of 280 images are used in this experiment which are captured from a Canon IXUS 125HS camera. Size of the images is 4608×3456. Images are resized to the size of 1152×864 during the experiments. Altitude of the camera is reported as 162 m. Visual results for the proposed approach and the bundle adjustment are given in Figure 4.3. Numerical performances of both proposed approach and the bundle adjustment are given in terms of RMS of the residuals in Table 4.1.

It can be concluded that there are no obtrusive differences between the visual results of the proposed method and the bundle adjustment for Small

(a) Result of the proposed method

(b) Result of the bundle adjustment

Figure 4.3: Results of the proposed method and bundle adjustment for Small Village

Village image sequence. Numerical results are also close to each other which also show the success of the proposed approach.

4.4.2 Pteryx UAV-Volvo Factory Image Sequence

Images are provided in [64]. 364 images are used during the experiments. Images are captured from a Canon PowerShot S90 camera and are size of 3648×2736. During the experiments, images are resized to 912×684. Visual

results are shown in Figure 4.4. Altitude of the camera is reported as 200 m [65]. Numerical results are provided in Table 4.2.

Table 4.2: RMS values for Volvo

Case	**Total Matches**	**RMS(pix)**
Proposed Approach	1009580	1.8353
Bundle Adjustment	1009580	1.6191

For Volvo Factory image sequence, we again obtained similar visual and numerical results for proposed approach and the bundle adjustment. However, some fractures and seams are available in the mosaic image. Since these problems are also available for the results obtained by the bundle adjustment, we can conclude that these inconveniences are due to the violation of planar scene assumption which is the main assumption for all mosaicing algorithms where homography is used as the motion model and Euclidean camera motion is present.

4.4.3 Bourget Airport Image Sequence

Images are provided in [63]. 251 images are used during the experiments. Images are captured from a Canon IXUS 125HS camera. Altitude of the camera is reported as 120 m. Size of the images are 4608×3456 and they are resized to 1152×864 during the experiments. Results of the Bourget dataset for the proposed approach and the bundle adjustment are given in Figure 4.5. Numerical results are presented in Table 4.3.

There are apparent defects in the mosaic for Bourget Airport image sequence for both proposed method and the bundle adjustment. This is again related to the violation of the planar scene assumption. Defects are more

(a) Result of the proposed method

(b) Result of the bundle adjustment

Figure 4.4: Results of the proposed method and bundle adjustment for Small Village

(a) Result of the proposed method

(b) Result of the bundle adjustment

Figure 4.5: Results of the proposed method and bundle adjustment for Bourget

Table 4.3: RMS values for Bourget

Case	Total Matches	RMS(pix)
Proposed Approach	240325	3.6772
Bundle Adjustment	240625	1.7528

apparent since there are high buildings, towers, planes in the airport and camera altitude is low which causes a more serious violation. When carefully inspected, it can be noticed that some defects in the mosaic created by the proposed algorithm is corrected with the bundle adjustment. A relatively large difference between RMS of the feature residuals between the proposed algorithm and the bundle adjustment also validates this observation.

4.4.4 Construction site (France) Image Sequence

Images are provided in [63]. Total number of 28 images are used during the experiments. Images are captured from a Canon IXUS 220HS camera. Size of the images are 4000×3000. Altitude of the camera is reported as 82 m. Images are resized to 912×684 during the experiments. Visual results for the image sequence are given in Figure 4.6. Numerical results are presented in Table 4.4.

For a numerical comparison, RMS of the residuals are tabulated in Table 4.4. For this image sequence, results of the proposed algorithm is almost

Table 4.4: RMS values for Construction site (France)

Case	Total Matches	RMS(pix)
Proposed Approach	27393	2.5042
Bundle Adjustment	27393	2.2284

identical, both visually and numerically, to the results of the bundle adjust-

(a) Result of the proposed method

(b) Result of the bundle adjustment

Figure 4.6: Results of the proposed method and bundle adjustment for Construction site (France)

ment. This is partly because of the large intersection ratios between images which can be inferred from the compact structure of the mosaic image.

Chapter V

5 Conclusions

We have now developed two different mosaicing approaches to create image mosaics of the planar scenes. In the first method presented in Chapter 3, we proposed a sequential mosaicing approach where alignment of new images were computed by using all the previous images intersecting the new image. To detect image intersections, a computer graphics tool, namely 'Separating Axis Theorem' (SAT) was employed. This theorem uses basic geometric procedures which provide an efficient operation. Since alignments of the previous images were assumed to be fixed during the alignment estimation of the new image which provides locally optimal estimates, we proposed a novel refinement procedure to enhance the global consistency of the mosaic by which alignments of the recent images were updated jointly. Experimental results show the success and potential of our algorithm when it is compared to some the other state-of-the art methods in the literature.

In the second method presented in Chapter 4, we proposed a new image mosaicing technique based on recursive estimation of the alignment parameters of the images. Parameters of all the images were updated at each time step by using Extended Kalman Filter. This was handled quite efficiently thanks to the recursive structure of the estimation and sparse nature of the image relations which provides small-size measurement equations. Several

experiments on publicly available datasets were conducted to assess the performance of our proposed algorithm. Results show that our algorithm produces satisfactory image mosaics which are visually and numerically close to the results of the bundle adjustment.

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. We also plan to use a computationally cheaper detect-and-track based feature matching approach as in [18].

References

[1] Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision **74** (2007) 59–73

[2] Choe, T.E., Cohen, I., Lee, M., Medioni, G.: Optimal global mosaic generation from retinal images. In: Proceedings of the 18th International Conference on Pattern Recognition - Volume 03. ICPR '06, Washington, DC, USA, IEEE Computer Society (2006) 681–684

[3] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust Mosaicing with Correction of Motion Distortions and Tissue Deformation for In Vivo Fibered Microscopy. Medical Image Analysis **10** (2006) 673–692

[4] Carozza, L., Bevilacqua, A., Piccinini, F.: 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) 6162–6165

[5] Choi, H., Choi, K., Kim, J.: Mosaicing touchless and mirror-reflected fingerprint images. Information Forensics and Security, IEEE Transactions on **5** (2010) 52–61

[6] Lin, Y., Medioni, G.: Map-enhanced uav image sequence registration and synchronization of multiple image sequences. In: CVPR '07. IEEE Conference on. (2007) 1–7

[7] Pizarro, O., Singh, H.: Toward large-area mosaicing for underwater scientific applications. Oceanic Engineering, IEEE Journal of **28** (2003) 651–672

[8] Irani, M., Hsu, S., Anandan, P.: Video compression using mosaic representations. Signal Processing: Image Communication **7** (1995)

[9] Szeliski, R.: Image alignment and stitching: a tutorial. Found. Trends. Comput. Graph. Vis. **2** (2006) 1–104

[10] Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2. IJCAI'81, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1981) 674–679

[11] Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. Int. J. Comput. Vision **56** (2004) 221–255

[12] Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2. ICCV '99, Washington, DC, USA, IEEE Computer Society (1999) 1150–

[13] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. **110** (2008) 346–359

[14] Tuytelaars, T., Gool, L.J.V.: Matching widely separated views based on affine invariant regions. International Journal of Computer Vision **59** (2004) 61–85

[15] Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)

[16] Lovegrove, S., Davison, A.J.: 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, Berlin, Heidelberg, Springer-Verlag (2010) 73–86

[17] Caballero, F., Merino, L., Ferruz, J., Ollero, A.: Homography based kalman filter for mosaic building. applications to uav position estimation. In: ICRA'07. (2007) 2004–2009

[18] Civera, J., Davison, A.J., Magallón, J.A., Montiel, J.M.: Drift-free real-time sequential mosaicing. Int. J. Comput. Vision **81** (2009) 128–137

[19] Kang, E.Y., Cohen, I., Medioni, G.: A graph-based global registration for 2d mosaics. In: Pattern Recognition, 2000. Proceedings. 15th International Conference on. Volume 1. (2000) 257–260 vol.1

[20] Elibol, A., Gracias, N., Garcia, R.: Fast topology estimation for image mosaicing using adaptive information thresholding. Robot. Auton. Syst. **61** (2013) 125–136

[21] Kim, D.W., Hong, K.S.: Real-time mosaic using sequential graph. J. Electronic Imaging **15** (2006) 023005

[22] Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment - a modern synthesis. In: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice. ICCV '99, London, UK, Springer-Verlag (2000) 298–372

[23] Lourakis, M.A., Argyros, A.: SBA: A Software Package for Generic Sparse Bundle Adjustment. ACM Trans. Math. Software **36** (2009) 1–30

[24] Konolige, K.: Sparse sparse bundle adjustment. In: Proceedings of the British Machine Vision Conference, BMVA Press (2010) 102.1–102.11 doi:10.5244/C.24.102.

[25] Wu, C., Agarwal, S., Curless, B., Seitz, S.M.: Multicore bundle adjustment. In: In IEEE Conference on Computer Vision and Pattern Recognition (CVPR, IEEE (2011) 3057–3064

[26] Zhou, G.: 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] Ross, R., Devlin, J., De Souza-Daw, A.: Mobile robot mosaic imaging of vehicle undercarriages using catadioptric vision. In: Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on. (2012) 247–252

[28] Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24** (1981) 381–395

[29] Yuan, C., Medioni, G., Kang, J., Cohen, I.: Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 1627–1641

[30] Torr, P.H.S., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding **78** (2000) 2000

[31] Tordoff, B., Murray, D.W.: Guided sampling and consensus for motion estimation. In: Proceedings of the 7th European Conference on Computer Vision-Part I. ECCV '02, London, UK, UK, Springer-Verlag (2002) 82–98

[32] Chum, O., Matas, J.: Matching with prosac - progressive sample consensus. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1. (2005) 220 – 226 vol. 1

[33] Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In Michaelis, B., Krell, G., eds.: Pattern Recognition. Volume 2781 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2003) 236–243

[34] Huber, P.: Robust Statistics. Springer Berlin Heidelberg (2011)

[35] Anandan, P., Bergen, J., Hanna, K., Hingorani, R.: Hierarchical model-based motion estimation. In: Motion Analysis and Image Sequence Processing. Springer (1993) 1–22

[36] Moravec, H.P.: The Stanford cart and the CMU rover. Springer (1990)

[37] HANNAH, M.: Test results from sri's stereo system. In: Science Applications International Corp, Proceedings: Image Understanding Workshop,. Volume 2. (1988)

[38] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60** (2004) 91–110

[39] Baumberg, A.: Reliable feature matching across widely separated views. In: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Volume 1., IEEE (2000) 774–781

[40] Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: Computer Vision, ECCV 2004. Springer (2004) 228–241

[41] Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets, or 'how do i organize my holiday snaps?2. In: Computer Vision, ECCV 2002. Springer (2002) 414–431

[42] Zoghlami, I., Faugeras, O., Deriche, R.: Using geometric corners to build a 2d mosaic from a set of images. In: Computer Vision and Pattern Recognition, 1997., IEEE (1997) 420–425

[43] Bartoli, A., Coquerelle, M., Sturm, P.: A framework for pencil-of-points structure-from-motion. In: Computer Vision-ECCV 2004. Springer (2004) 28–40

[44] Tuytelaars, T., Van Gool, L.: Matching widely separated views based on affine invariant regions. International journal of computer vision **59** (2004) 61–85

[45] Shi, J., Tomasi, C.: Good features to track. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94. (1994) 593–600

[46] Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets, or "how do i organize my holiday snaps?". ECCV '02, London, UK, UK, Springer-Verlag (2002) 414–431

[47] Brown, M., Lowe, D.G.: Recognising panoramas. In: ICCV. Volume 3. (2003) 1218

[48] Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence, IEEE Transactions on **27** (2005) 1615–1630

[49] Beis, J.S., Lowe, D.G.: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, IEEE (1997) 1000–1006

[50] Rousseeuw, P.J.: Least median of squares regression. Journal of the American statistical association **79** (1984) 871–880

[51] Ma, Y.: An invitation to 3-d vision: from images to geometric models. Volume 26. springer (2004)

[52] Malis, E., Vargas, M., et al.: Deeper understanding of the homography decomposition for vision-based control. (2007)

[53] Kekec, T., Yildirim, A., Unel, M.: A new approach to real-time mosaicing of aerial images. Robotics and Autonomous Systems (2014)

[54] Schneider, P.J., Eberly, D.: Geometric Tools for Computer Graphics. Elsevier Science Inc., New York, NY, USA (2002)

[55] Sawhney, H.S., Hsu, S., Kumar, R.: Robust video mosaicing through topology inference and local to global alignment. In: Computer Vision, ECCV 98. Springer (1998) 103–119

[56] Gracias, N., Costeira, J.P., Victor, J.: Linear global mosaics for underwater surveying. In: 5th IFAC Symposium on Intelligent Autonomous Vehicles. Volume 1. (2004)

[57] Sibley, G.: Relative bundle adjustment. Department of Engineering Science, Oxford University, Tech. Rep **2307** (2009)

[58] Davis, J.: Mosaics of scenes with moving objects. In: Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on, IEEE (1998) 354–360

[59] Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. ACM Trans. Graph. **2** (1983) 217–236

[60] : AerialRobotics Dataset. (`ftp://www.aerialrobotics.eu/`)

[61] (`http://www.uavpeople.com/download/dataset`)

[62] Yildirim, A., Unel, M.: Image mosaicing by camera pose estimation based on extended kalman filter. In: ICIAR. (2014)

[63] (`https://www.sensefly.com/examples-of-postflight-processing.html`) Accessed: 2014-06-18.

[64] (`ftp://aerialrobotics.eu/2012-04-30\%20VOLVO/`) Accessed: 2014-06-18.

[65] (`http://dronemapper.com/sample_data`) Accessed: 2014-06-18.