Robotics and Autonomous Systems 62 (2014) 1755–1767

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

A new approach to real-time mosaicing of aerial images

Taygun Kekec, Alper Yildirim, Mustafa Unel *

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

HIGHLIGHTS

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

ARTICLE INFO

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

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

ABSTRACT

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

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2. Image mosaicing

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