Pattern Recognition Letters 32 (2011) 1210–1221

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Robust refinement methods for camera calibration and 3D reconstruction from multiple images

Maw-Kae Hor [a,1], Cheng-Yuan Tang [b,*], Yi-Leh Wu [c,2], Kai-Hsuan Chan [a,3], Jeng-Jiun Tsai [a,4]

[a] Dept. of Computer Science, National Chengchi University, Taipei, Taiwan, ROC
[b] Dept. of Information Management, Huafan University, Taipei, Taiwan, ROC
[c] Dept. of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 25 January 2010
Available online 12 March 2011
Communicated by H.H.S. Ip

Keywords:
Refinement
Outlier
Camera calibration
3D reconstruction
Bundle adjustment

A B S T R A C T

This paper proposes robust refinement methods to improve the popular patch multi-view 3D reconstruction algorithm by Furukawa and Ponce (2008). Specifically, a new method is proposed to improve the robustness by removing outliers based on a filtering approach. In addition, this work also proposes a method to divide the 3D points in to several buckets for applying the sparse bundle adjustment algorithm (SBA) individually, removing the outliers and finally merging them. The residuals are used to filter potential outliers to reduce the re-projection error used as the performance evaluation of refinement. In our experiments, the original mean re-projection error is about 47.6. After applying the proposed methods, the mean error is reduced to 2.13.

1. Introduction

In computer vision, an interesting research topic, obtaining 3D models of scenes from images, has been studied for a few decades. Obtaining 3D models (or 3D reconstruction) from images is still a difficult task (Brown and Burschka, 2003; Faugeras, 1992). Obtaining 3D models from images can be separated into a few sub-problems, such as features matching (correspondence problems), structure from motion, uncalibrated structure from motion, self-calibration, dense stereo matching, and 3D reconstruction systems. Tang et al. (2008c) showed a method to modify the Scale Invariant Feature Transform (SIFT) descriptors for image matching. Cornelius et al. (2004) proposed to use the bundle adjustment (BA) technique (Lourakis and Argyros, 2008) to optimize the accuracy of 3D points with uncalibrated images. Steinicke et al. (2006) implemented a system to construct building models artificially in virtual reality with multiple views. Aliaga et al. (2007) showed a method to reduce the estimation error of the camera parameters mathematically to construct high accuracy building model. Merrell et al. (2007) proposed a method to use GPU to increase efficiency and reduce noise. Furukawa and Ponce (2007a,b, 2008) used bundle adjustment to reconstruct accurate object models automatically. Mellor (2003) proposed a method to combine camera, GPS, inertial sensors, and inclinometers to shoot a large number of photos to detect high accuracy 3D points. And he calibrates by noise and map texture to construct building models. Kamberov et al. (2006) also showed a method to reconstruct 3D models by uncalibrated image which are taken from random views.

However, the existence of spatial outliers remains a problem for any 3D reconstruction method. Some researches work on the removal of outliers in images; e.g., James and Dimitrijev (2010) proposed a method to detect outliers in images which can combine the template matching methods for image recognition. Hu and Sung (2004) showed a method to evaluate the spatial outlier factor by local trimmed mean. Shekhar et al. (2003) proposed a general definition of S-outliers for spatial outliers and presented scalable spatial outlier detection algorithms.

Given multiple images obtained from multi-camera capturing systems, the methods for camera calibration and 3D reconstruction from these images are highly interested for many researchers. This work first follows the method proposed by Furukawa and Ponce (2008). However, by applying the bundle adjustment directly to optimize the 3D models, the outliers will greatly influence the optimized results. To alleviate the above problem, a filtering approach is adapted to filter out the outliers. First the 3D points are divided into several buckets and the SBA (Lourakis and Argyros, 2008) are applied to each bucket individually. The residual error is then

* Corresponding author. Tel.: +886 2 26632102x4357.
E-mail addresses: hor@cs.nccu.edu.tw (M.-K. Hor), cytang@cc.hfu.edu.tw (C.-Y. Tang), ywu@csie.ntust.edu.tw (Y.-L. Wu), d9601@cs.nccu.edu.tw (K.-H. Chan), g9602@cs.nccu.edu.tw (J.-J. Tsai).
[1] Tel.: +886 2 29393091x62297.
[2] Tel.: +886 2 27301221.
[3] Tel.: +886 0968 581 401.
[4] Tel.: +886 0963 308 367.

0167-8655/$ - see front matter Crown Copyright © 2011 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.03.007

used to filter potential outliers to reduce the overall re-projection error.

The main contributions of this work are as follows. First, this work proposes to employ hierarchical and local search with Zero mean Normalized Cross-Correlation (ZNCC) value to refine the corresponding points. Second, this work proposes to divide a set of points into buckets for outlier removal with the SBA.

The rest of this paper is structured as follows. Section 2 briefly discusses the 3D reconstruction process and the method proposed by Furukawa and Ponce (2008). Section 3 introduces the robust estimate techniques by line fitting. Section 4 elaborates the proposed refine method. Experiment results are shown in Section 6. Section 7 concludes this paper and lists future research directions.

2. Camera calibration from multi-view stereo and bundle adjustment

We now briefly discuss the reconstruction process over multiple views and the method proposed by Furukawa and Ponce (2008). We first consider the simplest case of two input images. We consider the following geometry relationships: the epipolar geometry between view pairs, represented by the fundamental matrix (Faugeras, 1992; Hartley et al., 1992) and the trifocal geometry between view triplets (Spetsakis and Aloimonos, 1990; Hartley, 1995; Aliaga et al., 2007). The above geometry relationships and the image correspondences can be computed automatically from images as described below.

2.1. Geometry relationships

The simplest case is the geometry of two perspective views. Both views can be acquired simultaneously as in a stereo setting. Or the images can be acquired sequentially, e.g., by using a moving camera. The above two situations are considered geometrically equivalent. Let P and P' be the camera matrices for each view where the $'$ indicates the second view. A point X in 3D space is projected as $x = PX$ and $x' = P'X$ in the first and second view respectively. The following three geometric relationships (Hartley and Zisserman, 2004) stand:

(1) **Correspondence geometry**: What are the constraints of the position of the correspondence point x' in the second view, given an image point x in the first view?
(2) **Camera geometry**: How to find the cameras matrix P and P' for the two views, given a set of corresponding image points $\{x_i, x_i'\},\ i = 1, \ldots, n,$?
(3) **Scene geometry**: What is the 3-D position of X, given corresponding image points pair (x, x') and cameras matrix P, P'?

The epipolar geometry of the two views can be used to compute the image correspondences automatically. The details are referred to Furukawa and Ponce (2008).

For two views, the basic algebraic entry is the fundamental matrix. For three views, the trifocal tensor is employed. The trifocal tensor is a $3 \times 3 \times 3$ matrix that relates the coordinates of the corresponding points or lines in three views. Just as the fundamental matrix is determined by the two camera matrices, the trifocal tensor is determined by three camera matrices to summarize the relative projective geometry of three cameras.

The tensor-based methods can be extended to a quadrifocal tensor in four views. However, the computation a quadrifocal tensor using a non-iterative method is possible but oftentimes staggering and the reconstruction from more than four views is even more difficult. Fig. 1 shows the recovered point structure and surrounding camera views.

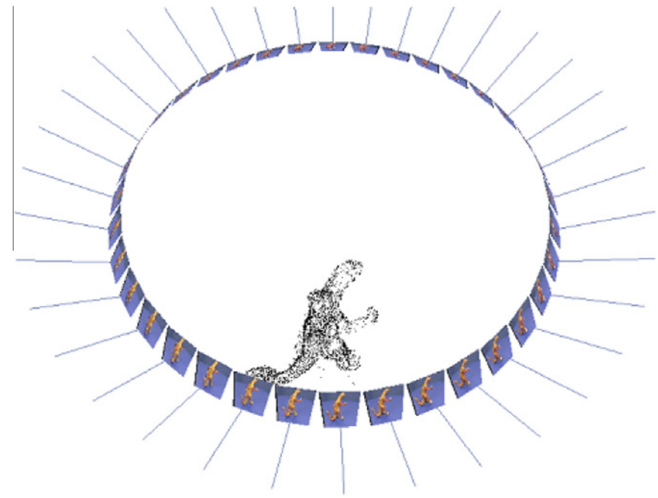

Fig. 1. Dinosaur: 3D point structure and camera positions for the Dinosaur sequence. (Image source: Furukawa and Ponce (2008).)

Many methods have been considered for reconstruction from more than four views and one of the most accepted methods was proposed by Furukawa and Ponce (2008).

2.2. 3D points refinement

This section discusses the method proposed by Furukawa and Ponce (2008) and the implementation processes.

2.2.1. Initializing and sub-sampling

To get more accurate 3D model, multiple images around the 3D object to be reconstructed are captured. The expected re-projection error for refining 3D model is defined. The distance between feature points and epipolar line is shown in Fig. 2. And the expected re-projection error Er is defined as $(\sum_i^N d_i^2 + d_i'^2)/N$, where N is the number of corresponding points, and d is the distance between the feature points and the epipolar lines (Furukawa and Ponce, 2008). To obtain more reliable corresponding points, image pyramids are built for all input images, as shown in Fig. 3. And then the level $L = \lceil \log_2 E_r \rceil$ of the pyramids is employed to run the PMVS (Furukawa and Ponce, 2007b). After the process, all the 3D points and a set of images in which the points is visible are obtained. Fig. 4 shows the relationship among the 3D points, image projection points, and a set of visible images V.

The above information is employed to construct the 3D points' projection in images, and sub-sample the points (Fig. 5). The image is divided into 10×10 blocks, and at most ε features in each block are randomly selected. Finally, the number of the points after sub-sampling is about 10–20% of the original points. The remaining points are employed to refine the correspondence points. Sub

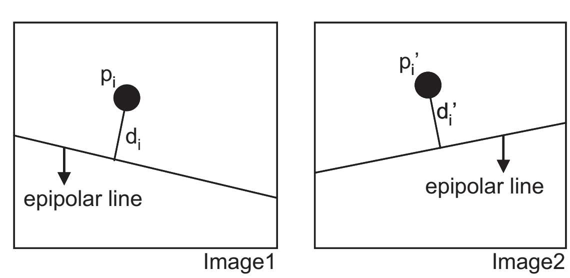

Fig. 2. The re-projection error is the distance between feature points and the epipolar line.