Optik 126 (2015) 3010–3015

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

An efficient method for organizing unordered images in 3D reconstruction

Boshen Ding [1], Zhenzhong Wei *, Sirui Ling [2]

Key Laboratory of Precision Opto-Mechatronics Technology of Ministry of Education, Beihang University, Room 532, Tower B, New Main Building, Beijing, China

ARTICLE INFO

Article history:
Received 5 June 2014
Accepted 9 July 2015

Keywords:
Structure from motion
Unordered views
3D reconstruction
Pairwise matching

ABSTRACT

Modeling 3D objects from unordered image data sets is being researched for several years. In this paper we investigate the problem of inefficiency in organizing unordered image sets into clusters of related. With the increase in number of images, computation for pairwise matching becomes more expensive. Practically considerable algorithms are more focused on the organization of images. Instead of saving time from the process of arrangement of views, we emphasize on the time-consuming part "comparison". This process is ahead of establishment of the relationship among views which includes building of the spanning tree or skeletal graph. The proposed algorithm achieves similar results to that of pairwise matching algorithm. More importantly it computationally saves more time by the introduction of a requirement before pairwise comparison which reduces computation time for two-view matching. The method we present has been tested on image sets both on the Internet and shot by ourselves. And it has been observed that more widely separated views lead to much better results.

1. Introduction

Structure from motion has been a popular research topic for years. Current two main approaches for sequence reconstruction are batch method [1–3] and incremental method [4,5]. Batch method using factorization framework has been improved to prior-free technique recently [3]. Incremental method tends to use trifocal tensor [6–8] for more accurate reconstruction and combines with prior information [9] for faster processing. Meanwhile, identification of a subset of several images from unordered image sets [9–12] is in practice for both methods.

In recent few years, there has been a growing concentration on developing robust view-ordering and view-grouping algorithms in the case of reconstruction from unordered set of views. The practical value leads the trend of unordered sets reconstruction to maximize the accuracy and completeness with a minimized computation time. Several methods exist which address the mentioned problems. Schaffalitzky and Zisserman proposed an approach for organizing unordered image sets [10] which used a greedy algorithm guided by hashing to establish the graph. Yao et al. made an improvement in defining view-similarity value which gets rid of the process of checking of existence of a cycle for added views in building of a spanning tree [13]. The following progress is made by Snavely et al. [12]. They identified a subset of views that could represent the completeness of reconstruction which reduced the runtime. Using prior information from GPS/INS sensors is also an efficient view selection strategy [14]. Besides, a great deal of attention on the development of robust and scalable vision algorithms applied to Internet photo collections has been aroused. Accordingly, several systems were set up aiming at reconstruction of unordered image sets [11,15–17].

Recently, the maturity of feature detection and matching [18] called SIFT(Scale Invariant Feature Transform) leads computer vision to a new level. Considering the robustness of SIFT, many researchers established their systems [7,12,16,19], which demonstrates its competence in both quantity and efficiency at application level. Similarly, proposed reconstruction basically uses it to extract features and match. The following subsections of ordering and reconstruction are both based on it.

In this paper, we investigate the issue of the unordered images reconstruction. As is known to all, the computational complexity of comparison will be $O\left(n^2\right)$ based on increase in number of images. We propose a relatively simple way for the organization of the

* Corresponding author. Tel.: +86 13520257650; fax: +86 10 82316930.
E-mail addresses: dingboshen@163.com (B. Ding), zhenzhongwei@buaa.edu.cn (Z. Wei), lingsirui@126.com (S. Ling).
[1] Tel.: +86 15101075179.
[2] Tel.: +86 15810541727.

http://dx.doi.org/10.1016/j.ijleo.2015.07.061
0030-4026/

views. Before figuring out the relationship between each pair of views, we make a judgment is made on the two nominated ones by using previous comparison consequences. As a result, all pairs of views do not require to be compared, which causes reduction of the runtime of comparison step greatly and thus saving computation effort. Moreover, we present experimental results on organization. Hence, proving most ordering results are similar to the pairwise matching's, but with much less time. Finally, we accomplish the reconstruction and testify the correctness of the ordering algorithm. Thus, the remainder of the paper will introduce the efficient method of views organization and the procedure of sparse reconstruction. The reconstruction section is not the emphasized part, so we describe it in brief.

2. Structure and motion

In general, sequential reconstruction is based on binocular stereo vision. Hence initializing structure and camera pair comes first, and then goes to the procedure of adding a view to the previous reconstruction each time. The framework is based on the assumption that the order information is provided ahead. We will describe the ordering approach in the following subsection.

The intrinsic matrix is easily determined if EXIF (Exchangeable Image File) tags embedded with digital images are available. Considering the inaccuracy in some cases using EXIF tags, we adopt Zhang's calibration method [20] to obtain the relative precise matrix value and distortion in the case of fixed focal length.

2.1. Feature matching of images

Views are related by corresponding image features which indicate the inherent epipolar geometry. We also weigh the view-similarity value by the two-view putative matches' status. Therefore, the first step is to find correspondences between two images. We use SIFT algorithm considering of its robust characteristic [18].

By using the difference-of-Gaussian function, potential interest points are identified. One can be found in different scale space for not only once, which leads to duplicate points. We record them once for the practical use. After acquiring exact keypoint locations, SIFT provides a local descriptor for each keypoint. The matching criteria, the vector's Euclidean distance ratio of their closest to second-closest neighbors may cause indiscriminative situation that one point matches with several points. If one feature in image *I* matches more than one feature in image *II*, instead of removing all of these matches [15], we set them down for further use. The rest correspondences consist of putative set of consideration. We will remove latent incorrect correspondences in epipolar geometry computation step.

2.2. Ordering images

In this subsection we present a simplified view-ordering algorithm, in the case that no prior information is provided. In most mainstream algorithms, using pairwise initial matches to establish a hashing table is inevitable. Whereas, we make a modification in the pairwise relating process that avoids the expensive computation caused by exhaustive pairwise feature matching.

We first search for all the images' features by SIFT algorithm mentioned in Section 2.1. Instead of comparing each pair among all the views, we first use one view to match with all the rest of views. As a consequence, we obtain the relationship between the first one and the others. In some circumstances, we could obtain

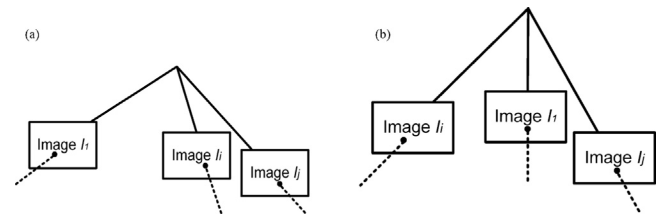

Fig. 1. (a) Shows images I_i and I_j are both matched "well" with image I_1. (b) Shows images I_i and I_j are both matched "badly" with image I_1.

the relationship between each pair picked from the rest of views. In other words, if two images I_i and I_j are both matched "well" or "badly" with the first one, it could not help to explain the relationship between I_i and I_j. Because we cannot distinguish if the "distance" between these views is near. Fig. 1 can show the reason straightforwardly.

Fig. 1(a) shows the situation that both images I_i and I_j have great many of correspondences with image I_1. But the correlation between I_i and I_j is implicit. They might share little overlap or the matching algorithm could filter few correspondences because of viewpoint changing, especially visual angle's alteration [18]. Similarly, Fig. 1(b) indicates the situation that both images I_i and I_j have few correspondences with image I_1. We cannot make a conclusion that I_i and I_j share few correspondences either. As shown in Fig. 1(b), on the contrary, the distance between images I_i and I_j could be small. Therefore, in the case of these two circumstances we cannot make any modification.

Nevertheless, when the third circumstance occurs that I_1 is "well" matched with one of I_i and I_j and "badly" matched with the other one, we could almost draw a conclusion. The last circumstance indicates that I_1 is "near" to one image in I_i and I_j, but "far" from the other. Then I_i and I_j could not be "near" except for one situation shown in Fig. 2.

Generally, I_i and I_j could be deemed to have no overlap in the third circumstance. But in Fig. 2's case, images I_i and I_j actually share a little overlap. It is because both I_i and I_1 are nearer to the object than I_j. They almost move along $I_{j''}$s ray axis. Therefore, their scenes are magnified results of different parts of image I_j. Fig. 2(b) shows the example of minority situation visually.

In fact, we can judge the minority situation by using the standard variance of I_j and I_1's matching points. Obviously, I_1's result of standard variance will be larger. Then we set a threshold on I_1/I_j

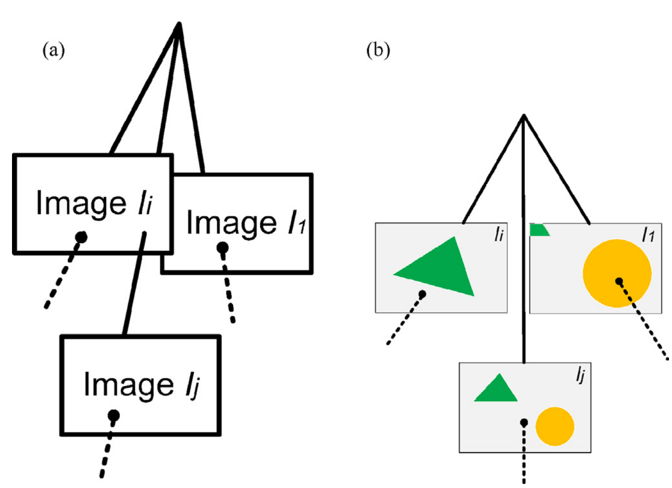

Fig. 2. The situation that images I_i and I_1 are matched "badly", while images I_j and I_1 are matched "well". However, I_i and I_j, are relevant whose correspondences are considerable. (b) Is the example of this minority situation.