Signal Processing 89 (2009) 2152–2158

Contents lists available at ScienceDirect

Signal Processing

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

A practical new approach to 3D scene recovery

Qingbo Zhu *, Hongyuan Wang, Wen Tian

Institute of Digital Video Technology and Communication, Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, PR China

ARTICLE INFO

Article history:
Received 31 October 2008
Received in revised form
4 April 2009
Accepted 24 April 2009
Available online 5 May 2009

Keywords:
3D scene reconstruction
Aspect ratio
Self-calibration
Maximum-likelihood estimation
Metric calibration
Uncalibrated image sequence

ABSTRACT

An efficient new approach to 3D scene reconstruction from an uncalibrated image sequence without anything known about the scene and the cameras is proposed. Present algorithms cannot deal with cameras with aspect ratio deviated from 1.0 well. Camera intrinsic parameters are skillfully handled in the cost function of the self-calibration phase to dispose the variable aspect ratio. Compared with existing algorithms, both intrinsic parameters of cameras and feature points of the scene, can be recovered efficiently, even for the cameras with aspect ratio greatly deviated from 1.0, like 0.6 and 2.3 in practice. Both synthetic and real data have been used to test the proposed method. The method is verified to be more accurate and practical than state-of-the-art algorithms, even if the aspect ratio is 2.5. Moreover, the reconstruction does not require any further non-linear optimization for general applications.

1. Introduction

Recovery of a scene structure has been studied for several decades. The classic method is presented in [3], of which camera self-calibration is a key and difficult problem and is the main subject of this paper. Much work has been done on camera self-calibration so far. All the algorithms can be divided into batch and stratified approaches according to the existence of projective reconstruction. Representatives of the former are the classic approaches proposed by Maybank et al. [1,2], Mendonca and Cipolla [9], while the methods proposed by Armstrong [3], Hartley et al. [4,5], Heyden [7], Triggs [8], Pollefeys et al. [14,15], Nistér [11], Chandraker et al. [16] and so on belong to the latter. Of those methods, [5,14] require that affine reconstruction is obtained first and used as the initialization towards Euclidean reconstruction. A more detailed discussion of stratification can be referred to [18].

Kruppa equations [1] based on the absolute conic are known to be difficult to be solved numerically, especially with large number of views. A variant of that approach is proposed in [7] where epipolar components are removed in the equations, and the sensitivity to noise is reduced. Afterwards, Triggs [8] firstly used absolute quadric to encode the camera parameters in the algorithm framework, and it is more general and gauge free. Lately Pollefeys et al. [15] used it to calibrate cameras with variable intrinsic parameters, but the initial value is proven to be a problem, and the camera intrinsic parameters matrix (parameter model) used for test is empirical and special.

The modulus constraints have been used to locate the infinity plane in [14], but it can only permit variance of focus length. The algorithm proposed in [5] follows the rigorous stratified approach, but our experiments have verified the execution time is very long at the step of searching for the plane at infinity. The possible reason is the SVD decomposition operation within the iteration process with large number of views.

* Corresponding author.
E-mail addresses: zhuqb.plan@gmail.com (Q. Zhu), wythywl@public.wh.hb.cn (H. Wang), drowterran@gmail.com (W. Tian).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.04.030

Nistér [11] has modified cheirality [6] constraints with respect to camera centers only and utilized a cost function reflecting the prior likelihood of camera intrinsic parameters. The convergent speed of the method is fast, but the accuracy is significantly reduced as the ratio of scale factors (aspect ratio) largely deviated from 1.0. It is unable to be applied to some cameras, like Texas CCD TC-244/245 with aspect ratio 2.3, and Sony CCD ICX055 with aspect ratio 0.6 presented in [23].

A method using the singular values of essential matrix has been proposed in [9]. It does not require a consistent set of weakly calibrated camera matrices. Some experiments in [10] have verified the convergence in several cases, but the requiring calculation of the fundamental matrix induces the non-linear method will not converge in some camera motion cases. The accuracy of the result is not good since it is very sensitive to noise.

Recently, Habed and Boufama [17] converted Kruppa equations into a set of bivarate polynomial equations for constant intrinsic parameters in order to simplify and speed up the process of solution. However, the method still consists of two drawbacks: one is that it can not use prior information of the camera parameters, and the other is the low accuracy.

Besides, rank degeneracy and positive semidefiniteness of the dual quadric have been enforced as part of the estimation procedure in [16], resulting in the complication of non-linear solution. The initial value and convergence are two difficult problems.

The performances of the actual experiments following the methods proposed by Pollefeys et al. [15] and Nistér [11,12] have determined the features and relations of the camera intrinsic parameters should be paid more attention, so that the solution will be more acceptable.

We propose a stratified new approach, leave some parts of parameter model known, and extend current cost function. The algorithm consists of three steps. Firstly, projective reconstruction is implemented by factorization of the measurement matrix or decomposition of the fundamental matrix, and bundle adjustment [20] refinement afterwards. Secondly, find the projective transformation to upgrade to metric reconstruction through Levenberg–Marquardt (LM) iterative method with a new cost function based on the prior work of Nistér's. The last step is the global metric bundle adjustment with all the camera parameters, which is optional for the requirement of the accuracy.

The rest of the paper is organized as follows. Section 2 establishes some notations and basic principals. Section 3 introduces the new cost function we proposed. Section 4 outlines the method we proposed. Section 5 presents experiments on synthetic and real data, and conclusions are drawn in Section 6.

2. Basic principals and notations

A multi-view perspective camera can be modeled through the following equation:

$$\lambda_{ij}x_{ij} = P_i X_j \quad (i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, N), \tag{1}$$

where M, N indicate the number of views and 3D points separately, P_i is the ith camera's projection matrix, $X_j = [x\ y\ z\ 1]^T$ is the homogenous coordinates of the jth 3D point in the world reference frame, $x_{ij} = [u\ v\ 1]^T$ is the homogenous coordinates of the ith image point in the jth view, and λ_{ij} is the projective depth.

It has been shown that it is only able to obtain a projective calibration from an uncalibrated image sequence through matches of pixel feature points and epipolar geometry [18].

After projective reconstruction, the remaining task is mainly to find the projective transform H to upgrade the results to metric reconstruction $\{\tilde{P}_i, \tilde{X}_j, K_i, R_i, t_i\}$. The relations can be expressed as follows:

$$\tilde{P}_i = P_i H^{-1}, \tag{2}$$

$$\tilde{X}_j = H X_j, \tag{3}$$

and $\tilde{P}_i$ can be written as

$$\tilde{P}_i = K_i R_i [I \ \ -t_i], \tag{4}$$

with

$$K_i = \begin{bmatrix} a_u(i) & s(i) & u_0(i) \\ & a_v(i) & v_0(i) \\ & & 1 \end{bmatrix}, \tag{5}$$

K_i encodes the intrinsic parameters of the ith camera, within which $a_u(i)$ and $a_v(i)$ indicate horizontal and vertical scale factors, respectively, $a_v(i)/a_u(i)$ is the aspect ratio, $s(i)$ is the skew and $(u_0(i), v_0(i))$ is the principal point.

(R_i, t_i) denotes a rigid transformation (motion parameters) of the ith camera, where R_i is a rotation matrix and t_i is a translation vector.

Assume $u_0(i) = w_i/2$ and $v_0(i) = h_i/2$, where (w_i, h_i) is the (*width*, *height*) of the ith camera.

Let

$$Kn(i) = \begin{bmatrix} 1/(w_i+h_i) & & -w_i/2(w_i+h_i) \\ & 1/(w_i+h_i) & -h_i/2(w_i+h_i) \\ & & 1 \end{bmatrix}. \tag{6}$$

It is feasible to assume $s(i) = 0$ as $s(i) \to 0$ for most applications.

Multiply K_i with $Kn(i)$ to the left, and the result is

$$K'_i = Kn(i)K_i = \begin{bmatrix} a_u(i)/(w_i+h_i) & & \\ & a_v(i)/(w_i+h_i) & \\ & & 1 \end{bmatrix}. \tag{7}$$

It is very simple compared with K_i, and it is easier to find the two unknown parameters in the diagonal. $Kn(i)$ is a known upper triangular matrix, which will not affect the solution of motion parameters. When K'_i is found, K_i will be obtained through left multiplied with $Kn(i)^{-1}$ to the left.

3. Maximum-likelihood estimation

As in majority of modern digital camera systems, the image center is the center of the retinal plane, skew is almost 0, and the varying parameters are the scale factors.