J. Vis. Commun. Image R. 25 (2014) 904–915

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

# Structure and motion from line correspondences: Representation, projection, initialization and sparse bundle adjustment

Lilian Zhang [a,b,*], Reinhard Koch [b]

[a] Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha, China
[b] Institute of Computer Science, University of Kiel, 24098 Kiel, Germany

ARTICLE INFO

*Article history:*
Received 20 July 2012
Accepted 9 February 2014
Available online 28 February 2014

*Keywords:*
Structure and motion
Line representation
Cayley representation
Line projection
Closed-form solution
Incremental initialization
Sparse bundle adjustment
Unconstrained optimization

ABSTRACT

We address the problem of structure and motion from line correspondences, which ranges from the representation of lines, their projections and the initialization procedure to the final adjustment. The Cayley representation of spatial lines is developed, which is a nonlinear minimal parametrization circumventing the tiresome Plücker constraint. The relationships between different line representations are given. Based on these relationships, we derive a novel line projection function which is consistent with the previous results. After building the line observation model, we employ a closed-form solution for the first image triplet, then develop an incremental initialization approach to initialize the motion and structure parameters. Finally, the sparse bundle adjustment (SBA) is applied to refine the parameters, which updates the spatial lines by using the Cayley representation with an unconstrained optimization engine. The experiments show that the proposed algorithm outperforms the previous works both in efficiency and accuracy.



## 1. Introduction

Structure and motion from multiple images is a classical and recurrent topic in computer vision. It has received a lot of attention and a number of algorithms have emerged in application areas such as scene modeling, augmented reality and robot navigation. This paper addresses the problem of estimating camera motion and recovering the three-dimensional structure of a scene composed of straight line segments from a set of line correspondences across multiple views. This is useful, as line features are prominent in most of the man-made environments, and a map of line segments gives a higher level of relevant information on the structure of the environment than point features.

Unlike point primitives with a simple representation and projection rule, there are two intrinsic difficulties when dealing with lines. Firstly, there is no global linear and minimal parametrization for 3D lines representing their four degrees of freedom by four parameters [1]. Secondly, depending on the representation, it may be non-trivial to project a 3D line into the image plane [2].

To circumvent the tiresome Plücker constraints as discussed in Section 8.2 of [3], the Cayley representation of lines is defined in this paper. The transformation between the Cayley representation and the Plücker coordinates is trivial. Besides, we present a novel derivation of the line projection function using the relationship between the Plücker coordinates and the Plücker matrix [4], which is consistent with the function derived from the dual relationship between points and lines. The resulting projection function is linear in terms of the Plücker coordinates. However, it is not well-adapted for nonlinear optimization because the Plücker coordinates have two degrees of internal gauge freedoms and are subject to the Plücker constraints. Instead, we adopt the Cayley representation to update line parameters during the optimization procedure to achieve an efficient nonlinear optimization approach with an unconstrained optimization engine.

For the scene reconstruction and camera motion estimation, we propose an initialization approach to boot the nonlinear optimization procedure by starting from the image triplet reconstruction [5] and incrementally adding new views and new line segments. Then we employ the sparse bundle adjustment (SBA) for nonlinear estimation. The SBA algorithm implementation [6] was originally designed for point-based reconstruction. We redesign the SBA library to make it appropriate for adjusting the line parameters.

To summarize, this paper addresses the problem of structure and motion from line correspondences. Compared to the traditional multi-view algorithm (such as [7,8]), our contributions are as follows: (i) The Cayley representation of 3D lines is introduced in Section 3 which can be updated without Plücker constraints.

* Corresponding author at: Institute of Computer Science, University of Kiel, 24098 Kiel, Germany.

*E-mail addresses:* lz@mip.informatik.uni-kiel.de (L. Zhang), rk@mip.informatik.uni-kiel.de (R. Koch).

http://dx.doi.org/10.1016/j.jvcir.2014.02.013
1047-3203/

(ii) A novel line projection function is derived in Section 4 for general camera models. (iii) A robust initialization approach is proposed in Section 5 to boot the nonlinear optimization procedure. (iv) The sparse bundle adjustment algorithm is extended to solve the line feature-based reconstruction problems in Section 6 which improves the speed significantly. We validate our algorithm and compare it to the existing works to show its efficiency and accuracy in Section 7. Finally, the conclusion is drawn in Section 8.

## 2. Related work and problem statement

This section briefly reviews the related works on line representations, line projection functions and reconstruction procedures. Then the statement of the problem is given.

### 2.1. The 3D line representation

Previous works on line representations are divided into two groups: the nonlinear minimal 4-parameter representation, and the linear over-parametrization representation. For the first group, Roberts [9] used two direction cosines together with the 2D coordinates to represent the direction and position of a line. Ohwovoriole [10] represented a line as the intersection of two planes which are parallel to the *X*-axis and the *Y*-axis respectively. For the linear over-parametrization group, Smith et al. [11] and Andrew and Walterio [12] represented a line with its two endpoints. Hartley and Zisserman [1] introduced two dual representations: a pair of points and a pair of planes. Montiel et al. [13] and Eade and Drummond [14] used the midpoint and the direction of the segment to represent a 3D line. Weng et al. [5] represented a 3D line by its closest point to the origin together with its direction. Pottmann et al. [15] and Seo and Hong [16] used the Plücker coordinates which include the moment of a line to the origin and the line direction in space.

Hartley and Zisserman [1] and Bartoli and Sturm [8] gave a summary of the 3D line representation. The nonlinear minimal representations use only four parameters which are equal to the degrees of freedom of a 3D line, hence there is no internal gauge freedom nor consistency constraint. This property makes the minimal representations well-adapted to the nonlinear optimization. However, it is difficult to explicitly express the line projection function because of the non-linearity. Contrarily, those linear over-parametrization representations can be easily expressed in the projection function but with internal gauge freedom which may induce numerical instabilities. Taylor and Kriegman [7] used the closest point and the direction (in total six parameters) to build the projection function, while using two direction cosines and 2D coordinates (in total four parameters) to update during the nonlinear optimization. Bartoli and Sturm [8] used the Plücker coordinates to build the projection function and exploited the orthonormal representation, which includes a $2 \times 2$ and a $3 \times 3$ orthogonal matrix to update during the bundle adjustment.

### 2.2. The line projection function

The line projection functions are dependent on the representation of lines. The simplest one is the same as the point projection function for those representing lines by their endpoints [13,14,17]. Suffering from the problem of unstable detection of endpoints, the triangulation of two endpoint pairs may not correspond to the correct spatial line. For calibrated cameras, Weng et al. [5], Taylor and Kriegman [7], Goddard [18] and Schindler et al. [19] derived a line projection function from rigid transformation in metric space. Faugeras and Mourrain [20], Martinec and Pajdla [21] and Bartoli and Sturm [2] derived a $3 \times 6$ projection matrix for the Plücker coordinates under the perspective camera model. This linear projection function is similar to the point projection function but with higher dimension, which means it is computationally more costly. Hartley and Zisserman [1] introduced a line projection function for the Plücker matrix representation, which is in quadratic form with respect to the $3 \times 4$ camera projective matrix.

Our previous work [4] presented a novel derivation of the line projection function based on the relationship between the Plücker coordinates and the Plücker matrix. In this paper, we show that it is consistent with another derivation, which is based on the dual relationship between points and lines.

### 2.3. The reconstruction procedure

For the problem of camera motion estimation and scene structure recovering from line correspondences, the existing approaches are subject to three categories: linear solutions, the Extended Kalman Filter (EKF) based methods, and nonlinear optimization methods. Given a set of at least thirteen line segments viewed in three frames, it is possible to determine the motion and structure. Liu and Huang [22] and Spetsakis and Aloimonos [23] developed a linear algorithm for this problem based on these observations. Weng et al. [5] presented a closed-form solution and established the uniqueness of the solution. The above approaches only work on three views. By using matrix factorization, Triggs [24], Morris and Kanade [25] and Martinec and Pajdla [21] presented an improved linear algorithm without the limitation of only three views. However, the algorithm requires all lines to be visible in all views. Nevertheless, they can offer a good initial guess for those iterative methods. Several approaches based on the EKF are proposed in the literature [16,26,27]. These EKF based algorithms suffer from the drawback of the EKF itself, which limits the number of update parameters. Nonlinear algorithms are almost always the last step to yield a reliable result, especially in the presence of noise. Yen and Huang [28] and Liu and Huang [29] iteratively solved a set of nonlinear equations about the motion parameters for three views. Montiel et al. [13] described a nonlinear approach which considers any number of images greater than or equal to two.

Some works in the literature reconstruct 3D lines from line drawings [30,31]. Generally, the 3D reconstruction from line drawing uses clean user input of a single view. It assumes that all edges of the planar faced object are visible and edges are connected at end points to form clear wire frame of the object. In contrast to that, the line feature based reconstruction assumes a noisy line detection result from multiple images and lines may be visible only in parts of images.

The most closely related two works using bundle adjustment are [7,8]. Taylor and Kriegman [7] initialized the algorithm from the partially known camera rotation parameters, otherwise, they booted their nonlinear algorithm by sampling the subset of the parameter space, which is computationally expensive and does not guarantee the correct convergence of their algorithm. Bartoli and Sturm [8] first used a linear algorithm to reconstruct lines associated with each triplet of consecutive images and then registered these triplets using the method proposed in [2]. The alignment of two groups of 3D lines is not robust when the extracted lines are contaminated by image noise in the two triplets. In contrast, we develop an incremental initialization procedure to boot the SBA algorithm, which is also suitable for online reconstruction.

### 2.4. Problem statement

Consider a situation in which a set of 3D line features $\{\Gamma_j\}$ is viewed by a set of cameras $\{\Upsilon^i\}$. Denote by $\ell_j^i$ the observation of the *j*th 3D line as seen by the *i*th camera. Lines may be viewed only