Image and Vision Computing 28 (2010) 1369–1376

Contents lists available at ScienceDirect

Image and Vision Computing

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

A branch-and-bound algorithm for globally optimal camera pose and focal length ☆

Kyuhyoung Choi, Subin Lee, Yongduek Seo *

Department of Media Technology, Sogang University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 18 August 2009
Received in revised form 17 December 2009
Accepted 3 February 2010

Keywords:
Global optimization
Branch-and-bound algorithm
Camera pose and focal length

ABSTRACT

This paper considers the problem of finding the global optimum of the camera rotation, translation and the focal length given a set of 2D–3D point pairs. The global solution is obtained under the *L*-infinity optimality by a branch-and-bound algorithm. To obtain the goal, we firstly extend the previous branch-and-bound formulation and show that the image space error (pixel distance) may be used instead of the angular error. Then, we present that the problem of camera pose plus focal length given the rotation is a quasi-convex problem. This provides a derivation of a novel inequality for the branch-and-bound algorithm for our problem. Finally, experimental results with synthetic and real data are provided.

1. Introduction

Recent development in the L_∞-norm minimization method shows that globally optimal solutions of some geometric vision problems can be computed via the bisection algorithm based on the techniques of linear programming (LP) or second-order cone programming (SOCP) [1–3]. The L_∞ method provides a novel way of obtaining the global optima of diverse problems, but its applicable area has been somewhat limited to the cases, where the residual function can be expressed in the form of a quasi-convex function, for which the camera rotation is usually assumed to be known *a priori*.

Recently, Hartley and Kahl [4] showed that the assumption of known rotation may be overcome through a branch-and-bound algorithm, making it possible to find the rotation through an efficient search in the rotation space. The branch-and-bound (BnB) algorithm tests every cubic sub-domain of the rotation space for the possibility that it may provide a better (smaller residual) solution by solving a feasibility problem. The size of the sub-domain is then reduced repeatedly through a top-down approach. It is shown that the method can find the optimal solutions to the problems of camera pose and two-view relative motion under the L_∞ sense. The error metric adopted for the BnB search is the *angular* error metric similar to some of the previous works such as [1,5]. That is, the residual is defined as the angle between the ray of the image measurement and the ray of its parametric projection.

This paper presents that the image space error metric (pixel distance) can be used in the branch-and-bound algorithm instead of the angular error metric used in Hartley and Kahl [4]. By far the most important benefit of this extension is that we can directly deal with the problem in the unit of pixel distance which is indeed more intuitive than the angular distance. Based on the pixel distance formulation, we derive a new bound for the inequality of the feasibility test, which becomes the key engine of the BnB search algorithm. The new bound is shown to be a function of the focal length as well as the size of the cubic sub-domain. But because the focal length is known *a priori* for the case of a calibrated camera, the BnB search is shown to be still possible. This is our first contribution.

Secondly, we extend the BnB algorithm one step further in such a way that the focal length is included as an unknown to be computed together with the pose parameters. For this, we show that the problem of estimating the focal length and the translation given the rotation is a quasi-convex problem. That is, given the rotation, we can obtain a globally optimal L_∞ solution of the focal length and the translation. This result provides another quasi-convex function which did not appear in the literature yet. The challenge in the case of unknown focal length lies in the fact that the focal length term appears in the upper bound of the feasibility constraints, because it brings about non-convexity. We solve this problem by deriving a new bound based on a priori information on the possible range of the focal length. That is, given a range of the focal length, another new bound is computed using the maximum and minimum values of it. This makes it possible for the BnB to do global search over the new parameter space. The assumption of known range of the focal length is not impractical because, for example, we know that it usually lies in an interval like [500, 2000] for general web-cams. One strategy is to compute a

☆ This work was supported by the strategic technology development program of MCST/MKE/KEIT [2008-F-030-01, Development of Full 3D Reconstruction Technology for Broadcasting Communication Fusion].

* Corresponding author. Tel.: +82 2 705 8896.
E-mail addresses: kyu@sogang.ac.kr (K. Choi), subin@sogang.ac.kr (S. Lee), yndk@sogang.ac.kr (Y. Seo).

0262-8856/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2010.02.001

new bound based on the entire range of the focal length and use it for the BnB search over the rotation space. The other is to divide the range into several sub-intervals and run the BnB search for each of them. Because each BnB has an additional linear constraint confining the range of the focal length, only the interval having the optimal solution remains through the BnB search and the other sub-intervals will be found to be infeasible.

There are some related works on the pose estimation which is a popular problem due to its utility in robotic navigation, augmented reality, novel view synthesis, etc. Algebraic error minimization is frequently adopted for fast computation [6–8]; object space error in Lu et al. [9] and Schweighofer and Pinz [10]; orthogonal decomposition based on SVD in Fiore [11]. Global optimization has recently been considered in Agarwal et al. [12] using a branch-and-bound based on a convex envelope and fractional programming technique, Hartley and Kahl [4] using a branch-and-bound algorithm based on a coarse-to-fine rotation sub-division search, Olsson et al. [13] with a branch-and-bound algorithm with convex under-estimators, and Schweighofer and Pinz [14] minimizing an object space error with a sum-of-squares global optimization. The problem of estimating the camera pose and focal length is studied for a practical application of augmented reality in Jain and Neumann [15], Park et al. [16] and Matsunaga and Kanatani [17].

Compared to these works, our approach minimizes the L_∞-norm error using the branch-and-bound search technique of Hartley and Kahl [4]. The contribution is that (1) it gives an explicit formulation of a pixel distance instead of the angular distance, (2) the focal length is included as a variable in addition to the pose, (3) the problem of focal length and pose with known rotation is shown to be a quasi-convex problem, and (4) a fast feasibility test algorithm appropriate for the branch-and-bound is provided.

This paper consists of several sections. Section 2 revises the branch-and-bound algorithm of Hartley and Kahl [4] for the camera pose based on the angular error metric. Then, Section 3 shows how the image space error can be used instead of the angular error. In Section 4, we show that the problem of camera pose plus focal length with known rotation is a quasi-convex problem. A new bound is derived and the global solution is computed by the BnB algorithm in Section 5. The computation speed is improved through an efficient feasibility test algorithm presented in Section 6. Experimental results are given in Section 7, and concluding remarks and discussions in Section 8.

2. Branch-and-bound for camera pose

This section introduces the branch-and-bound algorithm of Hartley and Kahl [4] for the camera pose in order for us to make it easy to provide our work. The definition of the pose problem is: *Given a set* $\{(\mathbf{X}_i, \mathbf{v}_i)\}$ *of 3D points with known position and corresponding 2D image points, determine the translation* $\mathbf{t}$ *and rotation* $\mathtt{R}$ *of the camera matrix* $\mathtt{P} = [\mathtt{R}|\mathbf{t}]$.

When the camera is calibrated, the image space may be assumed to be the sphere of unit radius. The image of $\mathbf{X}_i$ on to the spherical surface is then given by the re-projection function:

$$\hat{\mathbf{v}}_i = \frac{\mathtt{R}\mathbf{X}_i + \mathbf{t}}{\|\mathtt{R}\mathbf{X}_i + \mathbf{t}\|} \tag{1}$$

The L_∞ solution for $\mathtt{R}$ and $\mathbf{t}$ is the solution of the following minimization:

$$\min_{\mathtt{R},\mathbf{t}} \max_i \angle(\mathbf{v}_i, \mathtt{R}\mathbf{X}_i + \mathbf{t}) \tag{2}$$

where $\angle(\cdot,\cdot)$ represents the angle between two vectors, and $\mathbf{v}_i \in R^3$ is the measurement vector representing the direction of its image point corresponding to X_i. Unfortunately, this problem is not a convex problem due to the non-linearity of $\mathtt{R}$ – it is not easy to obtain the global optimum.

In the method of Hartley and Kahl [4], the domain of rotation space, that is, the 3D sphere of radius π, is divided into small sub-sets $\{D_j | \cup_j D_j = \text{domain}(\mathtt{R})\}$, where each D_j is a cube of half-side length σ. Given an estimation $\{\hat{\mathtt{R}}, \hat{\mathbf{t}}\}$ with the L_∞ angular error ϵ, the method tests whether a rotation sub-domain D_j may have a rotation that yields a smaller residual than ϵ. This test problem, called the feasibility problem, is given as:

$$\begin{aligned} &\text{find } \mathbf{t} \\ &\text{s.t.} \quad \angle(\mathbf{v}_i, \bar{\mathtt{R}}_j\mathbf{X}_i + \mathbf{t}) < \epsilon + \sqrt{3}\sigma, \quad \forall i \end{aligned} \tag{3}$$

where $\bar{\mathtt{R}}_j$ is the rotation matrix corresponding to the center of D_j. Because this problem is a convex problem, it can be easily solved via any LP or SOCP solver such as GLPK or SeDuMi [18,19]. The value $\sqrt{3}\sigma$ means the maximum possible variation of the maximum residual (right-hand side of the inequality) caused by the variation of the rotation $\mathtt{R}$ in D_j whose half-side length is σ.

Algorithm 1. Branch-and-bound for camera pose

Input: Find an estimate $\{\hat{\mathtt{R}}^{(0)}, \hat{\mathbf{t}}^{(0)}, \epsilon^{(0)}\}; \sigma \leftarrow \sigma^{(0)}$.
1: **repeat**
2: $\sigma \leftarrow \sigma/2$.
3: Divide the domain into cubes $\{D_j(\sigma)\}$.
4: **for** each of $\{D_j\}$ **do**
5: Solve the feasibility problem (F);
6: **if** *infeasible* **then**
7: discard the domain D_j.
8: **else**
9: find a new estimate $\{\hat{\mathtt{R}}^{(k)}, \hat{\mathbf{t}}^{(k)}, \epsilon^{(k)}\}$;
10: **if** $\epsilon^{(k)}$ is smaller **then** update the estimate.
11: **end if**
12: **end for**
13: **until** $\sigma < \sigma_{\min}$

Now the BnB algorithm is shown in Algorithm 1. After dividing the domain of the rotation into small sub-domains (line 2), this algorithm repeatedly checks whether the domain D_j is a feasible candidate or not (line 5). If D_j is found to be feasible, there is a possibility of the existence of a better solution inside the cube D_j. So a new estimate is computed via the bisection algorithm (line 9). We give an explanation of the bisection algorithm later. The resolution of the domain is then refined by dividing again the feasible D_j into eight cubes (lines 2 and 3). The stopping condition in line 1 means that the error of the best solution is in the interval: $\epsilon^* \leqslant e_{\min} \leqslant \epsilon^* + \sqrt{3}\sigma$, where ϵ^* is the maximum residual obtained in line 9 during the loop. Specifically, the line 9 solves the following quasi-convex problem with the bisection algorithm:

$$\begin{aligned} &\min \ \epsilon \\ &\text{s.t.} \quad \angle(\mathbf{v}_i, \bar{\mathtt{R}}_j\mathbf{X}_i + \mathbf{t}) \leqslant \epsilon, \quad \forall i \end{aligned} \tag{4}$$

Any algorithm is applicable in finding the initial input estimate because it will eventually be upgraded during the branch-and-bound iteration. Note that the BnB algorithm is not dependent on the initial solution. Indeed, not all the parameters are required as input but only a feasible upper bound for $\epsilon^{(0)}$ is sufficient to start the algorithm. So, the simplest way is to choose an $\epsilon^{(0)}$ adequately large to allow for the constraints to yield a non-empty solution space.

The angular error in the constraints is replaced by the tangent of it in solving the problems of (3) and (4):

$$\tan(\mathbf{v}_i, \bar{\mathtt{R}}\mathbf{X}_i + \mathbf{t}) = \frac{\|\mathbf{v}_i \times (\bar{\mathtt{R}}\mathbf{X}_i + \mathbf{t})\|}{\mathbf{v}_i^\top (\bar{\mathtt{R}}\mathbf{X}_i + \mathbf{t})} \tag{5}$$

The problem definition of (3) is then re-written as follows: