Enhanced Continuous Tabu Search for Parameter Estimation in Multiview Geometry

Guoqing Zhou   Qing Wang

School of Computer Science and Engineering

Northwestern Polytechnical University, Xi'an 710072, P. R. China

{zhouguoqing,qwang}@nwpu.edu.cn

Abstract

Optimization using the L_∞ norm has been becoming an effective way to solve parameter estimation problems in multiview geometry. But the computational cost increases rapidly with the size of measurement data. Although some strategies have been presented to improve the efficiency of L_∞ optimization, it is still an open issue. In the paper, we propose a novel approach under the framework of enhanced continuous tabu search (ECTS) for generic parameter estimation in multiview geometry. ECTS is an optimization method in the domain of artificial intelligence, which has an interesting ability of covering a wide solution space by promoting the search far away from current solution and consecutively decreasing the possibility of trapping in the local minima. Taking the triangulation as an example, we propose the corresponding ways in the key steps of ECTS, diversification and intensification. We also present theoretical proof to guarantee the global convergence of search with probability one. Experimental results have validated that the ECTS based approach can obtain global optimum efficiently, especially for large scale dimension of parameter. Potentially, the novel ECTS based algorithm can be applied in many applications of multiview geometry.

1. Introduction

Parameter estimation is one of the most fundamental problems in multiview geometry. The typical measurements of error function include algebraic distance, geometric distance, reprojection error, and Sampson error [9]. Traditional optimization algorithms have been dominated by local optimization techniques based on the L_2 norm, such as Newton or Levenberg-Marquardt iterations [9] or bundle adjustment [20] for finding a local optimum. While some of methods except iterative nonlinear optimization yield closed-form solutions, they are quite efficient and relatively easy to implement. However, solving multiple view geometry prob-

lems in general is difficult due to the inherent non-convexity and the presence of local optima.

To remedy these problems, a number of literatures have shown that many multiview geometric problems are quasi-convex under L_∞ norm [8][13]. A particularly fruitful line of work has been the development of methods that minimize the maximal of reprojection errors (L_∞ norm) across observations, instead of the sum of squared reprojection errors. It has been proven that many multiview problems have a single local optimum under the framework of L_∞ norm. The existence of globally optimal solution enables it effective to use convex optimization in parameter estimation [11]. However, this kind of algorithm is too complicated to solve large-scale geometric problems efficiently.

Recently, researchers proposed a new strategy, that by giving a simple sufficient condition for global optimality that can be used to verify that a solution obtained from any local methods is indeed global [15][17]. This algorithm returns either a certificate of optimality for local solution or global solution. Agarwal *et al*. [1] discovered that Olsson's method[17] is a special case for generalized fractional programming. Dai *et al*. [5] found the sequence of convex problems are highly related and proposed a method to derive a Newton-like step from any given point. The efficiency of L_∞ algorithm has been improved obviously.

All of the above mentioned algorithms are still on the ways of traditional optimization, and fewer modern optimization methods are considered to solve these problems so far. In recent years, Tabu search (TS), a meta-heuristic optimization method originally proposed by Glover [6][7], has extensively attracted attentions of researchers. It enhances the performance of a local search method by using memory structures that describe the visited solutions: once a potential solution has been determined, it is marked as 'tabu' so that the algorithm does not visit that possibility repeatedly. However, the basic TS is proposed for combinatorial optimization problems primitively. Chelouah *et al*. [4] proposed a variant of TS for the global continuous optimization problems (GCOPs), called enhanced continuous tabu search

(ECTS). This scheme divides the optimization process into two sequential phases, namely diversification and intensification. As a common drawback in GCOPs, meta-heuristic approaches cannot guarantee finding the global optimum.

In this paper, we proposed a novel method under the ECTS framework for parameter estimation in multiview geometry. The procedure takes the result of linear method as initial estimation, and utilizes the ECTS to attain the global optimum. In the phase of diversification, we propose a non-iterative way to obtain an initial bounding convex hull that contains the global optimum. At the stage of intensification, we propose a new approach to attain the best neighbor set according to the characteristics of multiview geometric problems. Finally, we prove the convergence of ECTS method in multiview geometry from the viewpoint of probability. The algorithm tends to achieve the global estimation within an arbitrary small tolerance. For the reason, we can prove that the proposed ECTS method converges with **probability one** to the global optimum. Comparing with L_∞ algorithm[11] and its variants or improvements[1][15][5], our method not only obtains accurate estimations, but also decreases computational cost dramatically.

2. Problem formulation

The geometric vision problems we are considering in this paper are the ones where the reprojection error can be written as affine functions composed with a projection, *i.e.,* quotients of affine functions. These problems can be represented as (P0) based on the squared reprojection error,

$$
\begin{aligned}
\min f(\mathbf{x}) &= \sum_{i=1}^{N} f_i(\mathbf{x}) \\
\text{s.t. } f_i(\mathbf{x}) &= \frac{\sum_{j=1}^{2}(\mathbf{a}_{ij}^\top \mathbf{x} + \tilde{a}_{ij})^2}{(\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i)^2}, \mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i > 0
\end{aligned}
\tag{P0}
$$

where $\mathbf{x} \in \mathbb{R}^n$ is the unknowns to be solved for, and $\mathbf{a}_{ij}, \mathbf{b}_i \in \mathbb{R}^n$ and $\tilde{a}_{ij}, \tilde{b}_i \in \mathbb{R}$ are differentiated with geometric problems. The constraint $\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i > 0$ reflects the fact that the reconstructed points should be located in front of the cameras. The dimension of the problem (P0) is n, which is often fixed and intrinsic to particular application. For example, $n = 3$ for multi-view triangulation, $n = 6$ for 2D affinity, $n = 7$ for fundamental matrix, $n = 8$ for planar homography, and $n = 11$ for camera calibration, etc. In order to facilitate the following discussions, here, we take the N-view triangulation as an example.

Consider a set of camera matrixes $\mathbf{P}_i$ and corresponding image points $\mathbf{u}_i = (u_{i1}, u_{i2})^\top$ of $\mathbf{x} = (x_1, x_2, x_3)^\top$. The objective of triangulation is to recover $\mathbf{x}$. The simplest way is based on a linear algebraic method. Though this method may seem attractive, the cost function that it is minimized has no particular meaning and the method is not reliable. Under the framework of L_2 norm, we are led

to minimize the following cost function subject to the constraint of $\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i > 0$,

$$
f(\mathbf{x}) = \sum_{i=1}^{N} \|\mathbf{u}_i - \mathbf{P}_i \mathbf{x}\|^2
\tag{P1}
$$

After a simple expansion, (P1) could be rewritten as,

$$
\begin{aligned}
E(\mathbf{x}) &= \sum_{i=1}^{N} f_i(\mathbf{x}) \\
\text{s.t. } f_i(\mathbf{x}) &= \frac{\sum_{j=1}^{2}(\mathbf{a}_{ij}^\top \mathbf{x} + \tilde{a}_{ij})^2}{(\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i)^2} \\
\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i &> 0
\end{aligned}
\tag{1}
$$

where $\mathbf{x}, \mathbf{a}_{ij}, \mathbf{b}_i \in \mathbb{R}^n$ and $\tilde{a}_{ij}, \tilde{b}_i \in \mathbb{R}$ and

$$
\begin{aligned}
\mathbf{a}_{ij} &= \mathbf{p}_{ij} - u_{ij}\mathbf{p}_{i3}, \tilde{a}_{ij} = \tilde{p}_{ij} - u_{ij}\tilde{p}_{i3}, j = 1, 2, \\
\mathbf{b}_i &= \mathbf{p}_{i3}, \tilde{b}_i = \tilde{p}_{i3}, i = 1, 2, \ldots, N,
\end{aligned}
\tag{2}
$$

where $\mathbf{p}_{ij}$ is the j-th row of $\mathbf{P}_i$ concatenated with the scaler $\tilde{p}_{ij}$.

3. ECTS in multiview geometry

ECTS is a variant of traditional tabu search for the global continuous optimization [6]. It is consisted of five stages, including setting of parameters, diversification, search for the most promising area, intensification, and output of the best point found. The key stages of ECTS are diversification and intensification. At the stage of diversification, the algorithm scans the whole solution space and detects the promising areas, which are likely to contain the global minimum. The centers of these promising areas are stored in a so-called promising list. The aim of diversification is to determine the most promising area from the promising list. When the diversification ends, the step of intensification will start. It searches inside the most promising area for a more optimal result. In this phase, the search is concentrated on the most promising area by making the search domain smaller and gradually reducing the neighborhood structure. This strategy improves the performance of the algorithm and allows exploiting the most promising area with more accuracy.

3.1. The diversification

Now, we present how to carry out the diversification for multiview geometry problems. In order to facilitate discussion, we also take the triangulation as the example. At first, we show the way how to determine the most promising area, which should contain the global optimum $\mathbf{x}_{opt}$. We start with a convex hull and an initial point $\mathbf{x}_{init}$ found by linear algebraic method. If $\mathbf{x}_{opt}$ is the true global optimum of L_2 norm minimization, it follows

$$
E(\mathbf{x}_{opt}) = \min \sum_{i=1}^{N} f_i(\mathbf{x}_{opt}) \leq E(\mathbf{x}_{init}) = \delta^2,
\tag{3}
$$

where δ is a positive value. This means that each $f_i(\mathbf{x}_{opt})$ term is less than δ^2. According to Eq.(1), Eq.(3) can be rewritten in the following form. For each i,

$$\sqrt{\frac{(\mathbf{a}_{i1}^\top \mathbf{x}_{opt} + \tilde{a}_{i1})^2 + (\mathbf{a}_{i2}^\top \mathbf{x}_{opt} + \tilde{a}_{i2})^2}{(\mathbf{b}_i^\top \mathbf{x}_{opt} + \tilde{b}_i)^2}} \leq \delta. \quad (4)$$

This means, for each i, the following two constraints are satisfied,

$$\left| \frac{(\mathbf{a}_{i1}^\top \mathbf{x}_{opt} + \tilde{a}_{i1})}{(\mathbf{b}_i^\top \mathbf{x}_{opt} + \tilde{b}_i)} \right| \leq \delta \text{ and } \left| \frac{(\mathbf{a}_{i2}^\top \mathbf{x}_{opt} + \tilde{a}_{i2})}{(\mathbf{b}_i^\top \mathbf{x}_{opt} + \tilde{b}_i)} \right| \leq \delta. \quad (5)$$

Notice that for N-view triangulation, we have a total number of $4N$ linear constraints on the variable $\mathbf{x}$, formulated by multiplying both sides of the above constraints with the depth term $\mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i > 0$. We wish to obtain a convex hull containing the optimal $\mathbf{x}_{opt}$. That is to find the lower and upper boundaries x_{min}^j, x_{max}^j such that $x_{min}^j \leq x^j \leq x_{max}^j$, $j = 1, \ldots, n$. We can formulate a linear programming (LP) problem by linearizing the constraints for $i = 1, \ldots, N$.

$$
\begin{aligned}
-(\mathbf{a}_{i1}^\top + \delta \mathbf{b}_i^\top)\mathbf{x} - \tilde{a}_{i1} - \delta \tilde{b}_i &\leq 0 \\
(\mathbf{a}_{i1}^\top - \delta \mathbf{b}_i^\top)\mathbf{x} + \tilde{a}_{i1} - \delta \tilde{b}_i &\leq 0 \\
-(\mathbf{a}_{i2}^\top + \delta \mathbf{b}_i^\top)\mathbf{x} - \tilde{a}_{i2} - \delta \tilde{b}_i &\leq 0 \\
(\mathbf{a}_{i2}^\top - \delta \mathbf{b}_i^\top)\mathbf{x} + \tilde{a}_{i2} - \delta \tilde{b}_i &\leq 0
\end{aligned}
\quad (6)
$$

This process then provides an initial bounding convex hull that contains the global optimum $\mathbf{x}_{opt}$ based on the L_2 norm cost function. Compared to traditional diversification methods, the above mentioned way does not need iterative process for the most promising area, so it is effective to solve multiview geometry problems.

3.2. The intensification

In the classical ECTS, the intensification carries out the following routines: generation of neighbors, selection of the best neighbor, updating of the various lists and adjustment of the parameters. In other words, if the current solution converges in the local optimum, we must give the feasible direction to enable the solution escaping from the local one.

The generic ECTS generates a specified number of neighbors. But, when the dimensions of the vector or the number of the constraints increase rapidly, the verification of best neighbors is inefficient. In this paper, we propose a new approach to attain the best neighbor set according to the characteristics of multiview geometry problems.

The following cost function $f_i(\mathbf{x})$ is pseudo-convex [3],

$$f_i(\mathbf{x}) = \frac{\sum_{j=1}^{2}(\mathbf{a}_{ij}^{\mathrm{T}}\mathbf{x} + \tilde{a}_{ij})^2}{(\mathbf{b}_i^{\mathrm{T}}\mathbf{x} + \tilde{b}_i)^2} \quad (7)$$

It is well known that pseudo-convex function has some nice properties, such as described in the Lemma 1[16]:

Lemma 1. Let $f(\mathbf{x}) = \max_{i} f_i(\mathbf{x})$, $\mathbf{x}^*$ solves $\mu^* = \min_{\mathbf{x} \in S} f(\mathbf{x})$, $S = \{\mathbf{x} | \mathbf{b}_i^\top \mathbf{x} + \tilde{b}_i > 0, \forall i\}$, if and only if there exists λ^* such that,

$$\sum_{i=1}^{N} \lambda_i^* \nabla f_i(\mathbf{x}^*) = 0,$$

where $\lambda_i^* \geq 0$ if $f_i(\mathbf{x}^*) = \mu^*$ and $\lambda^* = 0$ if $f_i(\mathbf{x}^*) < \mu^*$ for $i = 1, 2, \ldots, N$ and $\sum_i \lambda_i^* = 1$.

The geometric interpretation of Lemma 1 is that, if none of gradients vanish, then in each direction d there is an i such that $\nabla f_i(\mathbf{x})^\top d \geq 0$, that is at least one of $f_i(\mathbf{x})$ does not decrease in each direction. The Lemma 1 roughly states that the gradient does not vanish anywhere except at the optimum. In the proposed method, we take the gradient of $f_i(\mathbf{x}_k)$ (k is the iteration of ECTS) as the descent direction and construct the candidate solution set. We generate the component z^j of $\mathbf{z}$ which satisfies the Gaussian distribution with mean x_k^j and standard deviation σ, $j = 1, \ldots, n$. Initially $\sigma = 1$, when k increases, $\sigma = d \times \sigma$ (d is a factor which is chosen from 0.997 to 0.999). If $\sigma < 10^{-4}$, we set $\sigma = 10^{-4}$ fixedly. The Lemma 1 ensures that the candidate solution set includes the solution trailing off the error, so the better solution could be obtained through the ECTS.

Since pseudo-convex is the sufficient and necessary condition for a global optimum in Lemma 1, each iteration ensures the solution toward to the global optimum. In a word, in our proposed approach, we can construct the most promising area from L_2 based basic method for the diversification and determine the best search direction from L_∞ based optimality conditions for the intensification.

3.3. The convergence analysis

Now we discuss the convergence of the proposed ECTS algorithm. On the basis of the description in section 3.2, the problem (P1) is a global continuous optimization problem. It can be rewritten as,

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}) \quad (\text{P2})$$

where $\Omega = \{\mathbf{x} \in \mathbb{R}^n | x_{min}^j \leq x^j \leq x_{max}^j, j = 1, \ldots, n\}$.

Essentially, the proposed method is an instance of the memory tabu search (MTS) [10]. The MTS has the following pipeline.

Step1: Generate an initial point $\mathbf{x}_0 \in \Omega$. Set $\mathbf{x}_0^* = \mathbf{x}_0$ and $k = 0$.

Step2: If a prescribed termination condition is satisfied, we stop the iteration. Otherwise, we generate a random vector $\mathbf{y}$ by using the generator of probability density function.

Step3: If $f(\mathbf{y}) \leq f(\mathbf{x}_k^*)$ then $\mathbf{x}_{k+1}^* = \mathbf{y}$ and $\mathbf{x}_{k+1} = \mathbf{y}$. Otherwise, if $f(\mathbf{y}) \leq f(\mathbf{x}_k)$, then $\mathbf{x}_{k+1} = \mathbf{y}$, else if $\mathbf{y}$ does not satisfy the tabu conditions, then $\mathbf{x}_{k+1} = \mathbf{y}$, else $\mathbf{x}_{k+1} = \mathbf{x}_k$. Go to step 2.

In order to interpret the convergence of ECTS, we introduce the following definitions [14].

Definition 1. Let $\{\xi_m\}$ be a sequence of random variables defined on a probability space. We say that $\{\xi_m\}$ converges in probability towards a random variable ξ, if $\forall \epsilon > 0$, $\lim_{m \to \infty} Pr\{|\xi_m - \xi| \geq \epsilon\} = 1$, denoted as $\xi_m \xrightarrow{p} \xi$.

Definition 2. Let ξ_m be a sequence of random variables defined on a probability space. We say that $\{\xi_m\}$ converges with probability one (or almost surely) towards a random variable ξ (denoted as $\xi_m \xrightarrow{a.s.} \xi$), we have

$$P\{\lim_{m \to \infty} \xi_m = \xi\} = 1,$$

or, when for any $\epsilon > 0$,

$$P\{\cap_{m=1}^{\infty} \cup_{k \geq m} [|\xi_k - \xi| \geq \epsilon]\} = 1.$$

Without doubt, $\xi_m \xrightarrow{a.s.} \xi$ is stronger than $\xi_m \xrightarrow{p} \xi$.

Theorem 1 (Borel-Cantelli theorem). Let $\{A_n\}$ be a sequence of events in a probability space, and $P_k = Pr\{A_k\}$. Then $\sum_{n=1}^{\infty} P_n < \infty$ implies $Pr(\limsup_{n \to \infty} A_n) = Pr\{\cap_{n=1}^{\infty} \cup_{k \geq n} A_k\} = 0$. If $\sum_{n=1}^{\infty} P_n = \infty$ and A_n are independent, then $Pr\{\cap_{n=1}^{\infty} \cup_{k \geq n} A_k\} = 1$.

The Lemma 2 and Theorem 2 give the global convergence property of the objective optimal value sequence induced by MTS, as described in solving problem (P2). f is supposed to have a global minimum $f^* = \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$, for any $\epsilon > 0$. Let $D_0 = \{\mathbf{x} \in \Omega \mid |f(\mathbf{x}) - f^*| < \epsilon\}$, $D_1 = \Omega \setminus D_0$.

Lemma 2. Solving (P2) by using MTS, we set $\mathbf{x}_k^* \in D_1$. Let the probability of $\mathbf{x}_{k+1}^* \in D_1$ be q_{k+1} and the probability of $\mathbf{x}_{k+1}^* \in D_o$ be p_{k+1}. If $y^j, j = 1, 2, \ldots, n$ satisfies the Gaussian distribution, then $q_{k+1} \leq c \in (0, 1)$.

Theorem 2. Solving (P2) by using MTS, if y^j, $j = 1, 2, \ldots, n$ satisfies the Gaussian distribution, then $Pr\{\lim_{k \to \infty} f(\mathbf{x}_k^*) = f^*\} = 1$. Namely $\mathbf{x}_k^*$ converges with **probability one** to the global optimal solution of (P2).

The proofs of Lemma 2 and Theorem 2 are given in the Appendix of the paper. In Theorem 2, f^* is the a global optimum of f and $\mathbf{y}$ is the candidate solution in each step of tabu search. Therefore, we could start from an initial estimate $\mathbf{x}_{init}$. The algorithm tends to achieve the global estimation within an arbitrary small tolerance. For the reason, we can predicate that the proposed ECTS converges in **probability one** to the global optimum.

4. The algorithm

Now we summarize our framework of ECTS for estimating parameters in multiview geometry.

Input: 2D point $\mathbf{u}_i$ in the image and camera matrix $\mathbf{P}_i$, the step size of tabu search is $t = 10^{-4}$, the tolerance is $\epsilon > 0$ and K is a predefined maximal number of iteration.

Output: The global optimal solution $\mathbf{x}_{opt}$.

Step 1. Take the result of non-linear method as the initial solution $\mathbf{x}_0^* = \mathbf{x}_0$. Set $k = 0$ and tabu list as empty.

Step 2. Construct the convex hull Ω, which contains the global optimization $\mathbf{x}_{opt}$, as mentioned in the section 3.1.

Step 3. If $|\sum_i f_i(\mathbf{x}_{k+1}^*) - \sum_i f_i(\mathbf{x}_k^*)| < \epsilon$ or $k > K$, the algorithm terminates, else continues.

Step 4. Generate the candidate set. If $\nabla f_i(\mathbf{x}_k) > 0$, generate the candidate element $\mathbf{z}$ along the gradient direction $\nabla f_i(\mathbf{x}_k)$. z^j, the element of $\mathbf{z}$, satisfies the Gaussian distribution with mean x_k^j and standard deviation σ. The details of σ and d can be seen in section 3.2. If $\mathbf{z} \in \Omega$ and it is not in the tabu list, we put $\mathbf{z}$ in the candidate set S.

Step 5. For each $\mathbf{z}_s, s = 1, \ldots, |S|$, we obtain $\mathbf{y} = \arg \min_s (\max_i f_i(\mathbf{z}_s))$ (based on L_∞ norm).

Step 6. If $\max_i f_i(\mathbf{y}) \leq \max_i f_i(\mathbf{x}_k^*)$ then $\mathbf{x}_{k+1}^* = \mathbf{y}$ and $\mathbf{x}_{k+1} = \mathbf{y}$. Otherwise, if $\max_i f_i(\mathbf{y}) \leq \max_i f_i(\mathbf{x}_k)$ or $\mathbf{y}$ is not in the tabu list, then $\mathbf{x}_{k+1} = \mathbf{y}$, else $\mathbf{x}_{k+1}^* = \mathbf{x}_k$. Put $\mathbf{x}_{k+1}$ into the tabu list.

Step 7. $k = k + 1$. Goto step 3.

In the step 6, we introduce a variable $\mathbf{x}_{k+1}^*$ to record the optimal one of $\{\mathbf{x}_i | i = 1, \ldots, k+1\}$. This is the main distinction between our method and traditional TS. It is worthy noted that, in the step 1, if the initial point is an unreliable algebraic result or a random initial value, the iteration of Tabu search will increase accordingly. But, the accuracy can still be guaranteed.

5. Experimental results

We have tested the proposed method on both synthetic and real scene data. At the first stage of evaluation, we compared our method with bisection algorithm (Bisect-I) [11] to verify the effectiveness and efficiency for moderate scale problems, taking the triangulation and resection as examples. Then, in order to evaluate the efficiency for large scale problems, we compare our method with some state-of-art methods discussed in [1][5][15], taking the SfM with known camera orientation as an example.

The synthetic data comes from the linfinity-1.0[1]. Most of real scene data are from VGG group[2] and Notre Dame data courtesy is from [19]. Our algorithm is coded in MATLAB. The experimental environment is a standard PC (P8600, 6GB 64bits) and Matlab 2010a. We use Matlab profiler to report the timings and performance comparisons. In L_∞ algorithm based methods, the adopted SOCP solver is SE-DUMI and linear programming is MOSEK. The reported runtimes are the total time spent in optimization routines and the time of setting up the problem is omitted.

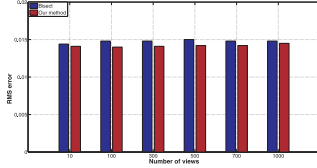[1] See http://www.maths.lth.se/matematiklth/personal/fredrik/download.html
[2] See http://www.robots.ox.ac.uk/~vgg/data.html

Figure 1. RMS error of the triangulation on varying view amount.

Figure 2. RMS error of the resection on varying point amount.

Figure 3. RMS error of triangulation with different noise levels.

Figure 4. RMS error of the resection with different noise levels.

Table 1. Runtimes of the triangulation on synthetic data.

Problems	Timing (s)		Speedup
	Bisect-I[11]	Our ECTS	
10-views	117.5779	0.4191	280.5
100-views	150.2874	3.0974	48.5
300-views	263.2977	16.2152	16.2
500-views	341.4276	25.6921	13.3
700-views	421.1711	39.8433	10.6
1000-views	482.4539	51.6829	9.3

Table 2. Runtimes of the resection on synthetic data.

Problems	Timing (s)		Speedup
	Bisect-I[11]	Our ECTS	
10-points	4.1865	0.0583	71.8
100-points	5.4075	0.2102	25.7
300-points	9.3340	0.8087	11.5
500-points	9.6668	0.8856	10.9
700-points	14.4366	1.3857	10.4
1000-points	17.7806	1.7039	10.4

5.1. Test on synthetic data

In this section we compare the ECTS algorithm with Bisect-I algorithm [11]. For the moderate scale problems we tested the algorithms on randomly generated instances of triangulation and resection problems with different sizes.

We simulated a 3D scene with 1,000 points within a cube and set N-views in front of the scene. The corresponding points of synthetic image are normalized into $[-1, +1]$ and Gaussian noises up to 0.01 are added in randomly. In Figures 1 and 2, the RMS (Root Mean Squares) errors of triangulation and resection problems with different sizes are reported. One can see that the errors by the ECTS are smaller than those of Bisect-I algorithm. Kahl *et al*. have pointed out that the improved bisection algorithm based on the L_∞ norm can obtain the global optimum [12]. Apart from the expected global optimum being achieved by the ECTS algorithm, Table 1 and 2 clearly show that the ECTS algorithm is more efficient than Bisect-I algorithm. The time cost of the triangulation on synthetic data shows that the speedup of our ECTS method decreases when the number of views increases. The main reason is that the number of iterations in the ECTS increases with the larger number of views. Fortunately, the view amount is rarely more than 100 for a 3D point in practical applications.

We have also validated algorithms on different Gaussian noise level for the triangulation and resection. Figure 3 and 4 show the RMS errors of two methods with different noise levels. The RMS errors of our ECTS method are lesser than Bisect-I algorithm.
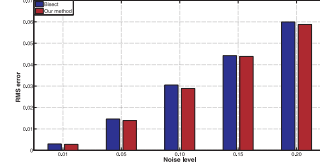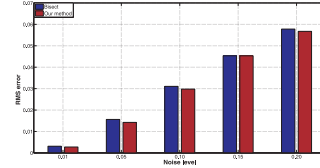
5.2. Test on real scene data

A. Triangulation

In the triangulation, we validated our method on two real scene data sets. In the VGG, the Model house contains 10 views and 672 tracks, and the Wadham contains 5 views and 1331 tracks. The Dinosaur data contains 36 views of 4,983 tracks and 16,432 feature points. The Notre Dame sequence contains 595 views of 277,877 tracks and about one million feature points. We only tested 212 images and randomly selected 27000 (from 160147) tracks which is probably sufficient enough to give a performance indication. The obtained reconstructions are shown in Figures 5 and 6. In order to illustrate more intuitive comparison of reconstruction results, Figure 7 shows the main part of the reconstructed dino's head, where the red crosses and blue circles are reconstruction results of Bisect-I method and the ECTS algorithm respectively. One can see that most of points in 3D space are coincident. At the edge of point clouds, reconstruction results appear slightly inconsistent.

In paper [15], the author reported that the speedups of Dinosaur and Notre Dame are 10.1 (3676s *vs*. 365s) and 7.2 (35815s *vs*. 4968s) respectively. From Table 3, we find it only took 12.28s for our ECTS method to obtain the result for Dinosaur data, while Bisect-I method spent 3199s. For

Table 3. Performance evaluation and speedup of the triangulation.

Dataset	Images	3D points	RMS error (pixels)		Timing (s)		Speedup
			Bisect-I[11]	Our ECTS	Bisect-I[11]	Our ECTS	
Model house	10	672	0.3930	**0.3741**	455.8962	**2.2511**	202.5
Wadham	5	1331	0.1553	**0.1533**	751.4879	**3.1245**	240.5
Dinosaur	36	4983	0.4283	**0.4060**	3199.8564	**12.2819**	261.5
Notre Dame	212	27000	0.5739	**0.5488**	19721.1946	**150.6106**	130.9

Table 4. Performance evaluation and speedup of the resection.

Dataset	Images	3D points	RMS error (pixels)		Timing (s)		Speedup
			Bisect-I[11]	Our ECTS	Bisect-I[11]	Our ECTS	
Model house	10	672	0.1000	**0.0368**	18.0754	**0.5111**	35.4
Wadham	5	1331	0.0091	**0.0084**	13.6696	**1.9669**	6.9
Library	3	667	0.0173	**0.0124**	7.8832	**0.3587**	22.0
Oxford	11	737	0.0615	**0.0305**	21.4552	**0.8945**	24.0

the Notre Dame data set, the improvement of efficiency is also significant, and the speedup is 130.9.

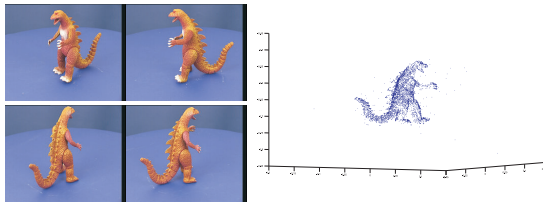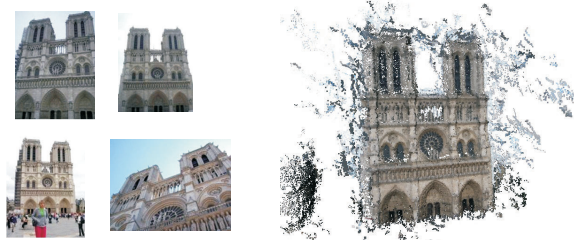Figure 5. The reconstructed 3D points of Dinosaur.

Figure 7. The enlarged details of the reconstructed Dino's head.

Figure 6. The reconstructed 3D points of Notre Dame Cathedral.

B. Resection

As far as to the resection issue, we chose four public available benchmarks from VGG data sets. The details of average reprojection error and computational cost can be found in Table 4.

Since it is hard to illustrate camera pose comparing with the ground truth for University library and Oxford, we only show comparisons of camera pose estimation by Bisect-I algorithm and our ECTS method with the ground truth for Model house and Wadham in Figure 8 and 9 respectively. The black rectangular pyramids illustrate the ground truth. Figure 8(b) and 9(b) show comparisons of Bisect-I algorithm and the ground truth. The green rectangular pyramids represent the position and orientation of cameras by Bisect-I method. In Figure8(c) and 9(c), the red rectangular pyra-
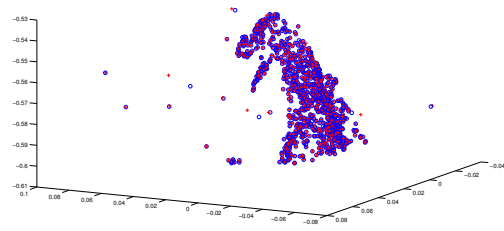
mids represent the position and principle axis of cameras by our ECTS method. From these results we can find that the ECTS method can achieve more accurate estimation of camera pose than Bisect-I method, nearly approaching the ground truth.

C. SfM with known camera orientation

Now, we present experiments on two benchmark data sets, Dinosaur and Oxford, which are publicly available from the Oxford VGG. The Dinosaur data set contains 36 cameras and 328 3D points [1]. The Oxford data set contains 11 cameras and 737 3D-points. Since our ECTS algorithm focuses on L_2 norm reprojection error, we compare it with Bisect-II, Dinkel-II, and Gugat algorithms discussed in [1] (the source code is provided by the author).

In [1], the author pointed out that Gugat's algorithm is the best one comparing to others. In [5], the author reported the experimental result of Dinosaur with 127 3D points, which took 1.07s on L_2 norm. Comparing to Gugat's algorithm (11.84s), its speedup is nearly 11 times. But for Oxford data, the author reported that Gugat's algorithm failed. In our experiments, we obtain about 3.7 and 4.8 times speedup comparing to Gugat's algorithm.

More importantly, we carry out the ECTS algorithm and all baseline algorithms for structure and motion recovery on the whole data set of Dinosaur (4983 3D points, 15054 un-
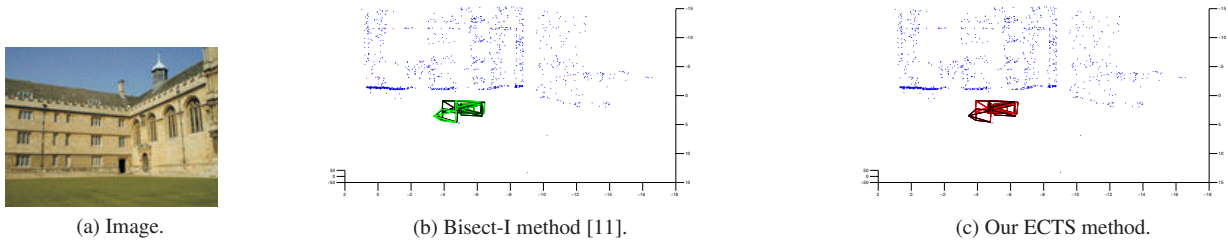
| (a) Image. | (b) Bisect-I method [11]. | (c) Our ECTS method. |

Figure 8. Comparisons of resection results of Wadham college.

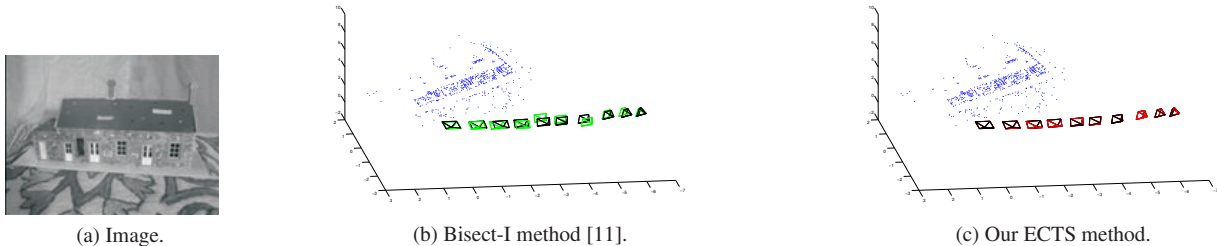| (a) Image. | (b) Bisect-I method [11]. | (c) Our ECTS method. |

Figure 9. Comparisons of resection results of Model house.

Table 5. Runtimes for L_2 norm reprojection error. All times are in seconds. f denotes numerical failure. Parameter settings, $\epsilon_1 = 0.01$, $\epsilon_2 = 0.001$, $\sigma = 1e6$.

Dataset	Images	Points	Unknowns	Observations	Bisect-II	Dinkel-II	Gugat	Our ECTS
Dinosaur (Partial data)	36	328	1089	2663	12.6306	13.9046	17.4233	**4.7787**
Oxford	11	737	2241	4035	46.2654	45.0122	35.7344	**7.4617**
Dinosaur (Full data)	36	4983	15054	16432	f	f	f	**1040.6616**

knowns and 16432 observations). The RMS error of our algorithm based on L_2 norm reprojection is 0.2247 and the runtime is 1040.66s. However, Bisect-II, Dinkel-II and Gugat fail to obtain the result, encountering the problems of out of memory. We think the key reason is that all these improved algorithms solved by SeDumi can not hold the robustness when the dimension of problem is becoming large. This has proven that the proposed ECTS algorithm is suitable to large scale multiview geometric problems.

We have also accomplished comparisons with the classical BA on this problem. The BA code is Vincent's SfM Toolbox [3]. Agarwal *et al.* have pointed out the BA has space complexity $O(N^2)$ and time complexity $O(N^3)$ [2]. But traditional Tabu search applicable to SfM is only $O(N^2)$ in each iteration [18]. In this paper, we propose to generate the candidate set along guided directions rather than randomized hyper-cube around the current solution. So time complexity is less than $O(k*N^2)$, where k is the number of iteration. In experiments, k did not exceed 15. For Oxford and whole Dinosaur data, the runtime of BA is 31.3211s and 9900s. But our algorithm only costs 7.4617s and 1041s, which show our method is more efficient when the scale of problem becomes larger. At the same time, the RMS errors of BA are 0.5217 and 0.4761 pixels and ours are 0.2267

---

[3]See http://vision.ucsd.edu/~vrabaud/toolbox/doc/

and 0.2247 pixels for Oxford and Dinosaur respectively. It is obvious that our method outperforms BA.

6. Discussions and conclusions

In this paper, we have demonstrated that many problems of parameter estimation in multiple view geometry can be formulated within a unified framework of enhanced continuous tabu search (ECTS), which is guaranteed to converge in **probability one** to the global optimum theoretically. We have validated the ECTS method for multiview geometric problems on both synthetic and real scene data, including triangulation, resection and N-view based structure and motion recovery. Experimental results have shown that the proposed ECTS algorithm can obtain accurate results as same as the traditional Bisect-I method. More importantly, the ECTS algorithm significantly speeds up parameter estimation many times than Bisect-I method and some state-of-art algorithms. Another encouraging result is that the RMS error of our method is lesser than the baseline algorithm regardless the number of images or noises. Therefore, the proposed method can be extended to many problems of parameter estimation in multiview geometry.

References

[1] S. Agarwal, N. Snavely, and S. M. Seitz. Fast algorithms for l_∞ problems in multiview geometry. In *Proc. CVPR*, 2008.

[2] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski. Bundle adjustment in the large. In *Proc. ECCV*, 2010.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.

[4] R. Chelouaha and P. Siarry. Tabu search applied to global optimization. *European Journal of Operational Research*, 123(2):256–270, 2000.

[5] Z. Dai, Y. Wu, F. Zhang, and H. Wang. A novel fast method for l_∞ problems in multiview geometry. In *Proc. ECCV*, 2012.

[6] F. Glover. Tabu search-part i. *ORSA Journal on Computing 1*, 1:190–206, 1989.

[7] F. Glover. Tabu search-part ii. *ORSA Journal on Computing 2*, 1:4–32, 1990.

[8] R. Hartley and F. Schaffalitzky. l_∞ minimization in geometric reconstruction problems. In *Proc. CVPR*, 2004.

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision, 2nd edn*. Cambridge University Press, Cambridge, 2004.

[10] M. Ji and H. Tang. Global optimizations and tabu search based on memory. *Applied Mathematics and Computation*, 159(2):449–457, 2004.

[11] F. Kahl. Multiple view geometry and the l_∞-norm. In *Proc. ICCV*, 2005.

[12] F. Kahl and R. Hartley. Multiple view geometry under the l_∞-norm. *IEEE TPAMI*, 30(9):1603–1617, 2008.

[13] Q. Ke and T. Kanade. Quasiconvex optimization for robust geometric reconstruction. *IEEE TPAMI*, 29(10):1834–1847, 2007.

[14] R. Laha and V. Rohatgi. *Probability Theory*. John Wiley and Sons, 1979.

[15] H. Li. Efficient reduction of l-infinity geometry problems. In *Proc. CVPR*, 2009.

[16] C. Olsson and F. Kahl. Generalized convexity in multiple view geometry. *J. Math. Imaging Vis*, 38(1):35–51, 2010.

[17] C. Olsson, F. Kahl, and R. Hartley. Projective least-squares: global solutions with local optimization. In *Proc. CVPR*, 2009.

[18] G. Paul. An efficient implementation of the robust tabu search heuristic for sparse quadratic assignment problems. *European Journal of Operational Research*, 209(3):215–218, 2011.

[19] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *SIGGRAPH, pages 835-846*, 2006.

[20] W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment for structure from motion. *In Vision Algorithms: Theory and Practice*, pages 504–509, 2000.

A. Appendix: proofs of Lemma 2 and Theorem 2 in section 3.3

A.1. Proof of Lemma 2

Proof: Let $\mathbf{x}_{min}$ be a global optimal solution of (P2). Since f is a continuous function, there exists a $r > 0$, such that $|f(\mathbf{x}) - f(\mathbf{x}_{min})| < \epsilon/2$. Let $Q_{\mathbf{x}_{min},r} = \{\mathbf{x} \in \Omega \|\|\mathbf{x} - \mathbf{x}_{min}\| \le r\}$. Obviously, $Q_{\mathbf{x}_{min},r} \subset D_0$. By assumption $\mathbf{x}_k^* \in D_1$, we have $f(\mathbf{x}_{k+1}^*) \le f(\mathbf{x}_k^*) \le f(\mathbf{x}_k)$. $y^j \sim N(x_k^j, \sigma^2), j = 1, \ldots, n$ leads to the generation of probability density function $g = \frac{1}{\sqrt{2\pi}\sigma} \exp^{\frac{(y^j - x_k^j)^2}{2\sigma^2}}$. Thus, the acceptance probability is

$$A = \begin{cases} 1 & f(\mathbf{y}) \le f(\mathbf{x}_k), \\ \mu\{\Omega - \cup_{k-L}^k (S_1 \cap S_2 \cup S_3)\}/\mu\{\Omega\} & f(\mathbf{y}) > f(\mathbf{x}_k), \end{cases}$$

where S_1, S_2 and S_3 are three criteria used to determine whether the candidate solution is tabu or not,

$$S_1 = \{\mathbf{y} \in \Omega \|\|\mathbf{x}_k - \mathbf{y}\| < \delta_1\},$$
$$S_2 = \{\mathbf{y} \in \Omega \||f(\mathbf{x}_k) - f(\mathbf{y})| < \delta_2\},$$
$$S_3 = \{\mathbf{y} \in \Omega \||f(\mathbf{x}_k) - f(\mathbf{y})|/f(\mathbf{y}) < \delta_3\}.$$

Obviously $A \le 1$. The probability of $\mathbf{x}_{k+1}^* \in Q_{\mathbf{x}_{min},r}$ is

$$\begin{aligned} Pr\{\mathbf{x}_{k+1}^* \in Q_{\mathbf{x}_{min},r}\} \quad &= Pr\{\mathbf{y} \in Q_{\mathbf{x}_{min},r}\} \\ &= \int_{Q_{\mathbf{x}_{min},r}} (g \times A)\, d\Omega \\ &\le \int_{Q_{\mathbf{x}_{min},r}} g\, d\Omega \end{aligned}$$

Since $\varnothing \ne Q_{\mathbf{x}_{min},r} \subset D_0$, we know $0 < Pr\{\mathbf{x}_{k+1}^* \in Q_{\mathbf{x}_{min},r}\} < 1$. $\mathbf{y}$ is a continuously random variable produced by the Gaussian distribution and $Q_{\mathbf{x}_{min},r}$ is the closed boundary set, so that there exists P, such that $P = \min_{\mathbf{y} \in Q_{\mathbf{x}_{min},r}} Pr\{\mathbf{y} \in Q_{\mathbf{x}_{min},r}\}$. As a result of $Q_{\mathbf{x}_{min},r} \subset D_0$, we have $p_{k+1} \ge Pr\{\mathbf{x}_{k+1}^* \in Q_{\mathbf{x}_{min},r}\} \ge P$. Let $c = 1 - P$, obviously $c \in (0, 1)$. Since $q_{k+1} + p_{k+1} = 1$, we have $q_{k+1} = 1 - p_{k+1} \le 1 - P = c < 1$. Thus, $q_{k+1} \le c \in (0, 1)$. $\square$

A.2. Proof of Theorem 2

Proof. $\forall \epsilon > 0$, let $q_k = Pr\{|f(\mathbf{x}_k^*) - f^*| \ge \epsilon\}$. If $\exists j \in \{0, 1, \ldots, k\}$ such that $\mathbf{x}_j^* \in D_0$, then $q_k = 0$. If $\forall j \in \{0, 1, \ldots, k\}$ such that $\mathbf{x}_j^* \notin D_0$, we set $q_k = \bar{P}$. By Lemma 2, we have

$$\bar{P}_k = P\{\mathbf{x}_0^* \in D_1, \mathbf{x}_1^* \in D_1, \ldots, \mathbf{x}_k^* \in D_1\}$$

So

$$\sum_{k=1}^\infty P_k \le \sum_{k=1}^\infty c^k = \frac{c}{1-c} < \infty$$

Then by Theorem 1, we get

$$Pr\{\cap_{n=1}^\infty \cup_{k \ge n} [|f(\mathbf{x}_k^*) - f^*| \ge \epsilon]\} = 0$$

According to the Definition 2, we gain the proof. $\square$