3D Reconstruction from Uncalibrated Images Taken from Widely Separated Views

Chunmei Duan, Xiangxu Meng, Lu Wang
School of Computer Science and Technology
Shandong University
Jinan, P. R. China
chunmeiduan@mail.sdu.edu.cn, mxx@sdu.edu.cn, wanglu319@hotmail.com

***Abstract*— In this paper, we present a framework for 3D reconstruction based on uncalibrated Images taken from widely separated views. Our method starts from scale-invariant key points being detected and described, then several schemas to improve the key points matching results being adopted. Consequently, with the fundamental matrix estimated from the key point correspondences, the epipolar geometry constraints between each view are recovered. We refine correspondence result by epipolar line and affine-invariant constraints. As a result, the refined correspondences will improve the fundamental matrix estimation. With the recovered fundamental matrix and epipolar, the sparse projective 3D point cloud of the scene could be recovered. After that, a globally nonlinear optimal procedure combined with Interval Analysis technique is performed to upgrade the projective 3D points to metric structure. The experimental results show our framework is effective for 3D reconstruction task.**

***Keywords*—uncalibrated image, scale-invariant key points, key points matching, self-calibration, 3D reconstruction**

I. INTRODUCTION

3D reconstruction from uncalibrated images is an important issue in computer vision. In the traditional version of the problem, baselines between cameras are assumed to be small for the sake of feature matching problem. However, the applications where two images of a same scene may be taken from very different points of view, namely wide baseline case, have drew great interest of researchers in the last few years [1,2,3]. To perform 3D reconstruction from wide baseline images has several advantages including greater precision, wider application and fewer input images etc. However, the feature correspondence and self-calibration in wide baseline case are still open problems. We need more efficient and appropriate strategies to solve the problems.

In this paper, we present a framework for automatic 3D reconstruction from uncalibrated Images taken from widely separated views. Our method starts from scale-invariant key points being detected and described, then several schemas such as affine-invariant constraint to improve the key points matching results being adopted. Consequently with the fundamental matrix F estimated from the key point correspondences, the epipolar geometry constraints between each view are recovered and then the key point matching result is refined by the constraint. The new correspondences are used again to improve the estimation of F. With the recovered F and epipoles, the sparse projective 3D point cloud of the scene could be recovered. After that, a globally nonlinear optimal procedure combined with Interval Analysis (IA) is performed to upgrade the projective 3D points to metric structure. Fig. 1 shows the flowchart of our reconstruction procedure.

This work was supported by the National Basic Research Program of China (No. 2006CB303102) and the National Nature Science Foundation of China (60703028).

II. KEY POINTS EXTRACTION AND DESCRIPTION

The first step of 3D reconstruction is to extract key points from uncalibrated images and take them as the skeleton points of the final sparse 3D structure. To establish the sparse 3D point cloud is equivalent to recover the depth information of the points and the depth recovery relies on the correspondences of key points of images taken from different point of view. Therefore, we need to detect key points in images firstly and describe them for the following correspondence task.

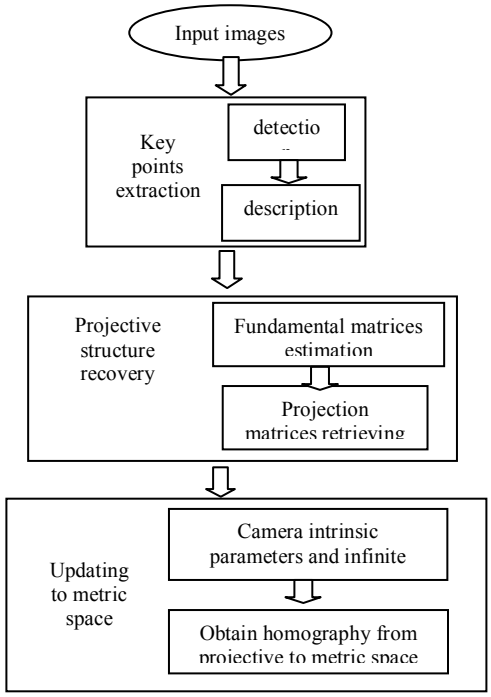

Figure 1. Flowchart of reconstruction procedure

A wide variety of detectors and descriptors have been proposed in the literature [4, 5, 6]. For the wide baseline task, we need the extracted features to be invariant with respect to geometrical variations such as translation, rotation, scaling and affine transformation, and photometric variations such as illumination and intensity change. SIFT [5] has been proven to be the most robust local invariant feature detector and descriptor among the others with respect to geometrical changes [6]. Therefore, we choose SIFT as our key points extraction tool. The work is divided into following three main steps:

1) Build octave pyramids with different Gaussian smooth factors as the scale space.

2) *Search in the scale space with DoG (Differential of Gaussian) operator for DoG extrema (maxima and minima), and take extremum pixels as key point locations which will be further refined to a subpixel precision level.*

3) *Represent the neighborhood of every key point as a feature vector which describes the orientation histogram around the neighborhood.*

Fig. 2 shows an image with key points detected by the DoG operator.

Figure 2. Key points detected by DoG operator

III. KEY POINTS MATCHING AND FUNDAMENTAL MATRIX ESTIMATION

With efficient descriptors for the key points extracted from images, the correspondence among key points is replaced with the correspondence among the descriptors. One of the matching strategies is using Euclidean distances between different descriptors as the similarity metric, which means that two key points are matched if their Euclidean distance in descriptor space is smaller than a threshold. According to the strategy, a key point could possibly have more than one match. So, we adopt an alternative matching strategy based on the nearest neighbor of a given key point. Following the strategy, two key points are matched when the ratio between their distance and the distance to the second nearest neighbor is smaller than a threshold. This approach generates only one correct match for a given key point and the precision is higher than the threshold-based strategy.

No matter which type of descriptor and matching approach are adopted, the existence of false matches is inevitable. We use affine-invariant constraint to remove the outliers in RANSAC framework [7]. After obtaining a relatively reliable match list, we estimate fundamental matrix from it and refine the matches further using the recovered epipolar line constraint. The following section shows algorithms for false match removing and fundamental matrix estimation.

Algorithm for false match removing:

Input: key point match list L sorted by similarity metric

Output: believable key point match list BL after removing false matches

Begin

Do RANSAC

1. For i=1 to 4

Random select in the match list L four of the most reliable matches noted by $\left(x_1^i, y_1^i\right)$, $\left(x_2^i, y_2^i\right)$ which are not very near to each other and any three points out of the four will not be collinear.

2. Let $A \in R^{8\times 8}$, and $H,B \in R^{8\times 1}$, where

$$A = \begin{pmatrix} x_1^1 & y_1^1 & 1 & 0 & 0 & 0 & -x_1^1\times x_2^1 & -y_1^1\times x_2^1 \\ 0 & 0 & 1 & x_1^1 & y_1^1 & 1 & -x_1^1\times x_2^1 & -y_1^1\times x_2^1 \\ x_1^2 & y_1^2 & 1 & 0 & 0 & 0 & -y_1^2\times x_2^2 & -y_1^2\times x_2^2 \\ 0 & 0 & 1 & x_1^2 & y_1^2 & 1 & -y_1^2\times x_2^2 & -y_1^2\times x_2^2 \\ x_1^3 & y_1^3 & 1 & 0 & 0 & 0 & -y_1^3\times x_2^3 & -y_1^3\times x_2^3 \\ 0 & 0 & 1 & x_1^3 & y_1^3 & 1 & -y_1^3\times x_2^3 & -y_1^3\times x_2^3 \\ x_1^4 & y_1^4 & 1 & 0 & 0 & 0 & -y_1^4\times x_2^4 & -y_1^4\times x_2^4 \\ 0 & 0 & 1 & x_1^4 & y_1^4 & 1 & -y_1^4\times x_2^4 & -y_1^4\times x_2^4 \end{pmatrix}$$

and

$$B = \left(x_2^1 \quad y_2^1 \quad x_2^2 \quad y_2^2 \quad x_2^3 \quad y_2^3 \quad x_2^4 \quad y_2^4\right)^T.$$

Solve the equation AH=B using SVD algorithm and obtain the H.

Rewrite the H as

$$H = \begin{pmatrix} H[0] & H[1] & H[2] \\ H[3] & H[4] & H[5] \\ H[6] & H[7] & 1 \end{pmatrix}.$$

3. For all $X \in L$

Calculate the distance between X and HX, which is noted by d (X, HX), and remove the X from L if d (X, HX) is above a given threshold T.

4. Take the L which has the largest size as the BL.

Until the RANSAC time is satisfied

End

Fig. 3 shows the epipolar geometry that encoded in fundamental matrix. The corresponding point x' in another view of an image point x lies on a line *l* which is called epipolar line. So, two views of the same rigid scene taken from different view points are related through the epipolar geometry. We can describe the epipolar line of a point x by *l*=Fx where the F obviously encodes the whole epipolar line geometry. More details are available in [8].

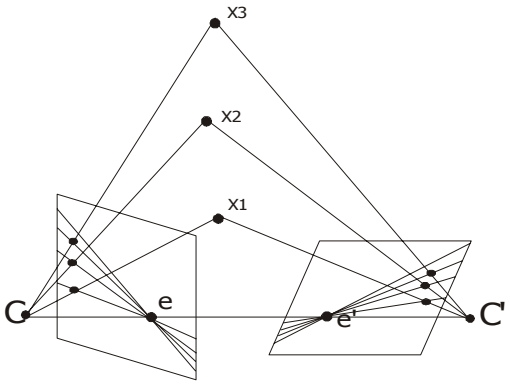

Figure 3. Multiple 3D points create a pencil of epipolar lines centered in the epipoles. The intersections between the different epipolar lines determine the epipoles e and e'.

Algorithm for Fundamental Matrix estimation:

Input: believable key point match list BL

Output: Fundamental Matrix F, Epipolar e_2 and refined BL

Begin

Do RANSAC

1. Random select 8 points from the BL.
2. Perform the Eight-point algorithm [8] on the random-selected 8 points and obtain the initial F.
3. Remove the outliers and add more inliers using the recovered epipolar line constraint.
4. Update the BL using previous result.
5. Estimate F again using updated BL and select the F minimizing the energy function $E = \sum_{(x,x')\in BL}(d^2(x', Fx) + d^2(x, F^T x'))$, where d(*) denotes the Euclidian distance.

Until the RANSAC time is satisfied

End

IV. Depth Recovery

A. Retrieving the projection matrices s

It has been shown by [8] that given a set of fundamental matrices satisfying the constraints, the corresponding projection matrices can be determined up to an unknown projective ambiguity. Based on the projection matrices, the coordinates of the 3D points in the projective frame can be computed easily.

We deduce one of the approaches to retrieve the projection matrices (as shown in the following). Algorithm for retrieving projection matrices:

Algorithm for retrieving projection matrices:

Input: Fundamental matrix F

Output: projection matrices P_1, P_2 of the first and the second view respectively

Begin

1. Calculate the epipolar e_2 of the second view by $F^T e_2 = 0$, where e_2 can be solved linearly.
2. Assume that the first projection matrix $P_1 = [\, A_1 \mid a_1 \,]$ and the second one P_2 has the form $[\, A_2 \mid a_2 \,]$, we have [8]:

$$\lambda e_i = a_i - A_i A_j a_j \quad (1)$$

$$\mu F_{ij} = [e_j]_\times A_j A_i^{-1} \quad (2).$$

Here λ, μ are arbitrary positive factors and i=1, j=2 in our case.

3. Suppose $P_1 = [\, I \mid 0 \,]$, and substitute it into (1), then we obtain

$$\lambda e_2 = a_2 .$$

From (2), we deduce that

$$\mu F = [e_2]_\times A_2 \Rightarrow \mu F^i = e_2 \times A_2^i ,$$

where $i \in \{1,2,.3\}$ represents the i-th column of the matrix;

$\Rightarrow \mu F^i \times e_2$, e_2 and A_2^i are coplanar, and $\left|F^i \times e_2\right| = \left|e_2\right|^2 \left|A_2^i\right| \sin\left\langle e_2, A_2^i \right\rangle$

$\Rightarrow A_2 = \dfrac{F^i \times e_2}{\left|e_2\right|^2} + r e_2$. Here r is an arbitrary positive factor.

4. We take $P_1 = [\, I \mid 0 \,]$ and $P_2 = \left[[e_2]_\times F \mid e_2 \right]$ as the final results for further purpose.

End

This approach for retrieving projection matrices totally depends on the fundamental matrix that calculated in previous phases. So it is more robust and stable than other methods. The projective reconstruction is equivalent to recover the metric one up to an unknown homography of the projective space. To compute this homography from assumptions on the cameras and projective results is called a self-calibration problem.

B. Self-calibration

Self-calibration problem is equivalent to recovering the unknown intrinsic parameters of the cameras. A great deal of work has been devoted to the problem in the last few decades. The solution to self-calibration can be mainly divided into three groups:

(1) The Kruppa equations [9].

(2) The stratified approach [10, 11].

(3) To compute the upgrading homography of projective-to-metric directly [12].

However, the CMS (Critical Motion Sequences) are camera motions which will defeat any self-calibration algorithm. Fig. 4 shows the generic critical motions for more than 2 views. Fig. 4 (a) shows the situation that optical axes are parallel, namely, the motions of cameras are translating. Fig. 4 (b) shows that all camera optical axes coincide except at two locations (at most) where the cameras can be arbitrary oriented. Fig. 4 (c) shows that camera centers move on two conics, one is an ellipse and

the other is a hyperbola. The two conics lie on orthogonal

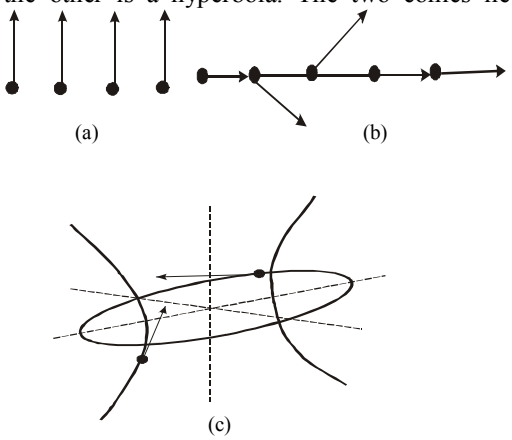

Figure 4. (a) Camera optical axes are parallel; (b) Camera centers are aligned; (c) Camera centers move on two conics.

planes and camera optical axes lying on the corresponding support planes are tangent to the conics.

Pollefeys [11] presented a self-calibration method which starts from solving a linearized version of the problem and then refine it by iterative non-linear optimization procedure. Unfortunately, the linearization will introduce more artificial CMS and fail both in case of small and wide baseline. Apart from it, the iterative nonlinear optimization is prone to falling in local minima. So we need a deterministic method for solving the nonlinear self-calibration problem which does not require an initial solution.

Our formulation of the problem is:

Let $K_i = \begin{pmatrix} foc & 0 & u \\ 0 & foc & v \\ 0 & 0 & 1 \end{pmatrix}$ be the intrinsic parameter matrix of the i-th camera, where foc is the focus length of the camera, (u, v) is the principal point of the image plane. According to the theory in computer vision (more details are available in [8]), we obtain that:

$\omega_i^* \sim K_i K_i^T \sim P_i \Omega^* P_i^T$,where ω_i^* is the image of dual absolute conic; $\Omega^* = \begin{pmatrix} KK^T & -KK^T\pi_\infty \\ -\pi_\infty^T KK^T & \pi_\infty^T KK^T \pi_\infty \end{pmatrix}$ is the dual absolute quadric, where the infinite plane $\Pi_\infty = (\pi_\infty, 1)$. Then, our goal is to find the minimal optimization of the following cost function:

$$f(K_i, \Omega^*) = f(foc, u, v, \pi_\infty) = \sum_{i=1}^{N} (\|F(K_i K_i^T)\| - \|F(P_i \Omega^* P_i^T)\|)^2 \quad (3),$$

where || F(*)|| is the Frobenius norm.

We adopt Interval analysis (IA) technique to directly solve the non-linear problem. Interval analysis is a technique for global optimization. An interval is denoted by $x = [\underline{x}, \bar{x}]$, where $\underline{x}$ and $\bar{x}$ are the lower bound and the upper bound of $x \in IR$ respectively. An interval vector is denoted by $X = [\underline{x}_i, \bar{x}_i, i = 1, \cdots, n]$, $X \in IR^n$. Internal vectors are called boxes. More details about IA can be found in [13, 15]. With a cost function f(X), $X \in IR^n$ and a list of boxes denoted by L for X , the basic schema for IA along with a branch and bound algorithm to find the minimal of f is:

1) Initialize L using the initial search box of X.

2) Do

Remove a interval box X from L.

Processing (doing cut-off test, monotonicity test and concavity test).

Bisect X and insert the derived box into L.

Until L is empty

We use GlobSol [14] to implement the IA technique. Combined with IA, we can solve equation (3) and obtain the focal length foc and the infinite plane Π_∞. Therefore, the final homography $T = \begin{pmatrix} K^{-1} & 0_3 \\ \pi_\infty^T & 1 \end{pmatrix}$, which will upgrade the structure from projective space to metric space, is retrieved.

V. EXPERIMENTAL RESULTS

In our experiment, we assume that the intrinsic parameters of cameras are constant. We tested our framework on some real images as shown in Fig. 5. Those 768×576 images are taken from very different point of views. If the principal point (u, v) of camera is supposed to be known, the number of images used for reconstruction can be reduced to two, otherwise, at least three of them are needed.

We used our method described in section 2 to detect and describe key points in images automatically and used strategies reported in section 3 to obtain key point correspondences and in turn retrieve fundamental matrix. Subsequently, to obtain the projection matrices, we used algorithm presented in section IV.A. Next, taking the retrieved projection as the input, a global optimization procedure was carried out minimizing cost function proposed in (3) to obtain the final homography upgrading projective to metric structure. The starting interval of interval analysis was chosen as follows: initial interval for focal length is [300, 3000] and principal point (u, v) is [cx-W*20%,cy+H*20%], where (cx,cy) is the center of the image, W and H are image width and height respectively.

Table I shows the retrieved intrinsic parameters and Fig. 6 shows the final recovered 3D metric point cloud in a perspective view.

Figure 5. Some test images taken from wide separated views.

TABLE I. THE RETRIEVED INTRINSIC PARAMETERS

focal-length (pixels)	**u (pixels)**	**v (pixels)**
996	*362*	*304*

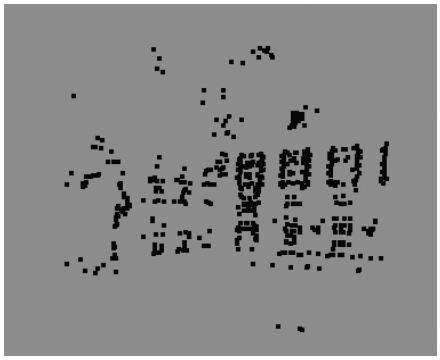

Figure 6. Recovered 3D metric point cloud in a perspective view

VI. CONCLUSIONS

We have presented a framework for 3D reconstruction from images taken from widely separated views. The wide baseline case introduces many new challenges such as feature correspondence and self-calibration. First, we detect the scale invariant key points from the images using DoG operator in scale space and match them with the affine-invariant and epipolar line constraints imposed on. The matching result is precise enough to estimate fundamental matrix and then retrieve the projection matrices. Finally, we adopt interval analysis to globally optimize the non-linear cost function which encodes parameters including camera focal length and infinite plane, and obtain the upgrading homograph from projective space to metric space. Experimental results show that our framework is feasible, flexible and efficient.

REFERENCES

[1] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Wide-baseline multiple-view correspondences, " in CVPR, vol. I, 2003, pp. 718-728.

[2] Christoph Strecha, Rik Fransen, and Luc J. Van Gool, "Wide-baseline stereo from multiple views: A probabilistic account, " in CVPR, 2004, pp. 552 – 559 .

[3] T. Tuytelaars and L. Van Gool, "Wide baseline stereo based on local, affinely invariant regions, " in BMVC, 2000 , pp. 412 – 422.

[4] C.Harris and M. Stephens, "A combined corner and edge detector," in proc. of the Alvey Vision Conference, 1988, pp. 147 – 151.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints, cascade filtering approach, " IJCV, vol. 2, Jan. 2004, pp. 91 – 110.

[6] K. Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors, " IEEE Tans. On PAMI, vol. 27, 2005, pp. 1615 – 1630.

[7] Fischer MA, Bolles RC, " Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, " Communication of ACM J., vol. 24, 1981, pp. 381−395.

[8] R. I. Hartley and A. Zisserman, Multiple view geometry in computer vision, 2nd ed., Cambridge University Press, 2004.

[9] O. Faugeras, Q.-T. Luong, and S. J. Maybank, "Camera self-calibration: theory and experiments, " in ECCV, 1992, pp. 321–334.

[10] R. Hartley, E. Hayman, L. de Agapito, and I. Reid, "Camera calibration and the search for infinity, " in ICCV, 1999, pp. 510–517.

[11] M. Pollefeys and L. Van Gool, "Stratified self-calibration with the modulus constraint, " IEEE Trans. on PAMI, vol. 21, 1999, pp. 707–724.

[12] A. Heyden and K. Astrom, "Euclidean reconstruction from constant intrinsic parameters, " in ICPR, vol. 1, 1996, pp. 339–343.

[13] E. R. Hansen, G. W. Walster, Global optimization using interval analysis, 2nd ed., Marcel Dekker, 2003.

[14] http://interval.louisiana.edu/GlobSol.

[15] A. Fusiello, A. Benedetti, M. Farenzena,, A. Busti, "Globally convergent autocalibration using iInterval analysis, " IEEE Trans. on PAMI, vol. 26, 2004, pp. 1633–1638.