

PRECISION ANALYSIS TO 3D RECONSTRUCTION FROM IMAGE SEQUENCES

Sun Min ^{a,*}, He Rixin ^b, Wang Daojun ^a

^aInstitute of RS&GIS, Peking University, 100871 - sunmin@pku.edu.cn

^bThe Information Centre of Beijing Public Security Bureau, 100740

KEY WORDS: 3D Reconstruction, Precision Analysis, Photogrammetry, Multiple View Geometry

ABSTRACT:

The technology of 3D reconstruction based on image sequences is an important technology in computer vision field, its application in close range photogrammetry area brings a new research area. For the application of 3D reconstruction in close range photogrammetry field, precision problems must be fully demonstrated. In this paper, error propagation rules in 3D reconstruction process has deduced on the base of matrix analysis method, and a calculation method of covariance matrix has put forward for evaluate error of elements in 3D reconstruction result, at the same time, the effects from related factors to 3D reconstruction result are analyzed. At last, experiments are introduced to demonstrate the precision that 3D automatic reconstruction could reached in general case.

1. INTRODUCTION

For the limitation from the theory of traditional close range photogrammetry, approximate value of camera position and orientation need to be known to resolve photogrammetry problems. This defect can be overcome with the theory of multiple view geometry which greatly developed at 90's in the last century. As no initial value of camera position and orientation need to be known for the calculation of spatial relationship between images, so image data capture becomes easy and very convenient, hence it could bring revolution to the field of close range photogrammetry.

However, in the research works of multiple view geometry, result precision of the reconstruction is seldom discussed, as the result from 3D reconstruction is a model up to scale. The first problem that applies theory of multiple view geometry to the field of photogrammetry is to estimate precision of 3D result. There are some related works, e.g: Werner et al (2001) discussed the application of automatic reconstruction to close range photogrammetry. Förstner analyzed relations and differences between photogrammetry and computer vision, and also analyzed uncertainty of reconstruction (Förstner, 2002, 2005); Mayer (2003) discussed problems on reconstruction of image triplets. In these works, precision does not discussed.

For the application of computer vision in industrial measurement, quite a few works on precision are presented, however, industrial application need very high requirement to precision, mainly research works on effects of camera intrinsic parameters to industrial measurement have been done in earlier days (Mcvey 1982, Blostein 1987).

In the process of 3D reconstruction, many calculation models are used, it's much complex if compare with the calculation process of photogrammetry, which mainly use one model, i.e. collinear equation. It's difficult to use precision estimation method directly from photogrammetry, but the application of 3D reconstruction in photogrammetry field must resolve the problem of precision estimation. In this paper, we present one method to estimate precision for the result of 3D reconstruction.

2. CALCULATION MODEL OF 3D RECONSTRUCTION

Assume one image sequence is captured with camera in random pose, A and B are two adjacent images in the sequence, and form one image pair, i.e. A and B satisfy the condition: for certain area S in one scene, a random point X in S can be projected on image A and B, and get two image point x and x', then calculation models in 3D reconstruction process can be described with following matrices.

2.1 Calibration Matrix and Projective Matrix

Calibration matrix K is composed of camera intrinsic parameters, it contains five parameters: horizontal focal length f_x , vertical focal length f_y , principal point (u v) and skew s , s generally has value 0.

K can be obtained by interactive camera calibration or automatic calibration, Kruppa method (Faugeras 1992) and ADQ method are two typical automatic calibration methods (Triggs 1997, Pollefeys 2004). Interactive calibration generally could obtain very high precision close to 1 pixel, while automatic calibration can not obtain high precision even with the help of calibration board, calibration precision can only reach to several pixels (Zhang 1999).

Projective matrix describe the map relationship between a random point in one scene and a image point, assume rotation matrix between A and B is R, translation vector is t, and camera calibration matrix is K, projective matrix can be written as:

$$x = P[X \ 1]^T \quad P = K[R \ t] \quad (1)$$

Projective matrix essentially keeps coincident with collinear equation, when K, R and t are all unknown variables, it's hard to calculate P matrix, actually P is calculated by decomposing fundamental matrix F or essential matrix E when K is known.

2.2 Fundamental Matrix and Essential Matrix

Both fundamental matrix and essential matrix describe the map relationships from one point to a line between two images, hence describe the spatial relationship between two images. In fig.1, point x on image A and line l on image B satisfy equation: $l = Fx$. If matched image point pair (x, x') is known, following formula holds:

$$x'^T F x = 0 \quad (2)$$

F is a 3x3 matrix, as one free scale is unknown, so only 8 elements need to be calculated, i.e. 8 points are required for F calculation. Fundamental matrix F contains information of calibration matrix K, while essential matrix E has no relation with calibration matrix K, E and F has relation: $E = K^T F K$. The calculation of elements in F is linear, so the precision estimation of fundamental matrix can use linear method. In order to simplify the problem, we assume K keeps unchanged in the whole image sequence.

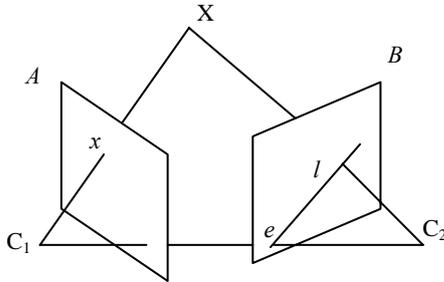


Fig.1. Principal of fundamental matrix

2.3 The Difference between 3D Reconstruction and Photogrammetry

The main process of 3D reconstruction can be summarized as following:

- 1) Calculate fundamental matrix using matched image pairs;
- 2) Estimate projective matrix with epipolar point and fundamental matrix;
- 3) Using matched image points and projective matrix calculate 3D point set X;
- 4) Calibrate X with calibration matrix K from projective space to Euclidean space;
- 5) Optimize the result with bundle adjustment method;

For the traditional theory of photogrammetry, this calculation process just need step 5, so approximate value for unknown variables should be known, while in this process of 3D reconstruction, only matched image points are required to be known.

3. ERROR PROPAGATION ANALYSIS

In order to simplify the problem, we take calibration matrix as known data, therefore, unknown variables in 3D reconstruction process are: F, E, t, R and X. The known data are image point x and calibration matrix K.

3.1 Error of Fundamental Matrix F

3D reconstruction process begins from the fundamental matrix calculation of first image pair, as state in section 2.2, fundamental matrix can be obtained by formula (2). Formula (2) is linear equation for elements in F, it can be written as $Af + L = 0$. Where A is a coefficient matrix composed of image point coordinates. L is a vector with elements equal to 1. Formula $Af + L = 0$ can be rewritten as: $f = -A^{-1}L$, calculate differential on both sides of the equation, one get:

$$\frac{dF}{df} = A^{-1} \frac{dA}{dx} A^{-1} L$$

For the convenience, in the following text, the differential matrix is denoted with a subscript d , so the above equation can be written as:

$$F_d = A^{-1} A_d A^{-1} L \quad (3)$$

Where F_d is the differential matrix of F, and A_d is differential matrix of A. As two inverse matrices of A exist, so we can draw a conclusion for error propagation from image point to elements in F matrix:

For a given image error, more large image coordinates corresponding to smaller elements errors in F matrix.

Therefore, in the calculation of F matrix, image points close to the image boundary should be selected as much as possible, otherwise, if too many points close to the principal point are selected, then the result of F would be unstable. As in matrix A, quadric items of the image point coordinates exist, while in A_d , only simple items exist. Therefore, the error of elements in F matrix is very sensitive to the image point coordinates. According to this conclusion, wide-angle lens should be used instead of common lens to capture data, so the result of F matrix would be more robust.

3.2 Error of 3D Point

As we assume K is known, so F can be replaced with E, E can be decomposed by SVD method to get rotation matrix R and translation vector t. SVD decomposition of E can be expressed as: $E = USV^T$.

Let T express skew symmetric 3×3 matrix which composed of vector t, then:

$$E = TR \quad \text{where } T = UWSU^T, \quad R = UWV^T \quad (4)$$

While W is a skew matrix, and has two cases of value, W and S are all constant matrices, the specific value of W and S is given by Hartley and Zisserman (2000).

Obviously, for the existence of two constant matrices, T actually is one matrix which diagonal element equal to 0, and t can be expressed with T elements as: $t = [-T_{23} \ T_{13} \ -T_{12}]^T$. Considering formula (1), projective matrix P can be expressed as:

$$P = K[UWV^T | t] \quad (5)$$

As $x = P[X \ 1]^T$, one get: $X = VW^T U^T (K^{-1}x - t)$

Calculate differential coefficient for variables in this equation:

$$dX = V_d V^T X + VW^T U_d (K^{-1}x - t) - VW^T U^T K^{-1} K_d K^{-1} x + VW^T U^T (K^{-1} dx - dt) \quad (6)$$

Simplify this equation with above expressions:

$$dX = V_d V^T X + R^T U U_d R X - R^T K^{-1} K_d (R X + t) + R (K^{-1} dx - dt) \quad (7)$$

Where U_d , V_d and dt are indirect variables, as $E = USV^T$, $E = TR$ and $S = \text{diag}[1 \ 1 \ 0]$, so one can get:

$$U_d S = E_d V \quad V_d S = E_d^T U^T \quad \text{and} \quad T_d = [dt]_x = E_d R^T \quad (8)$$

The existence of S matrix shows that the third column of U and V are zero vectors, so the complete expression of U_d and V_d can be written as:

$$U_d = E_d V + [0 \ 0 \ U_{d3}] \quad V_d = E_d^T U^T + [0 \ 0 \ V_{d3}] \quad (9)$$

Where U_{d3} and V_{d3} are respectively differential part of epipolar points on image A and B, so it can be calculated from Kt , as $E = K^T F K$, so:

$$E_d = K_d^T F K + K^T F_d K + K^T F K_d \quad (10)$$

In the case that both A_d and K_d are known, from above equation (7) (8) (9) and (10), one can summarize dX expression as:

$$dX = M_1 A_d M_2 + M_3 K_d M_4 + M_5 K_d^T M_6 + M_7$$

This expression shows error propagation rule from error of image point x and calibration matrix K to the last result, i.e. 3D point cloud X. As this expression is really complex, here we only give analysis to influence from single variable error to the last 3D point cloud.

1) Only error of image point exists:

When $K_d = 0$, one get:

$$E_d = K^T F_d K \\ dX = V_d V^T X + R^T U U_d R X + R (K^{-1} dx - dt)$$

From these two equations, we can draw one conclusion: firstly, 3D point error is relative to spatial position, large error corresponding to large distances, this viewpoint keeps

coincident with common knowledge. Secondly, 3D point error is very sensitive to camera intrinsic parameters, this is determined by quadric items of K element in equation $E_d = K^T F_d K$.

One important conclusion is: the longer focal length corresponding to bigger reconstruction error. While general viewpoint is longer focal length would obtain higher image resolution, and hence the precision should be higher. But this conclusion negates the general viewpoint. Actually, this conclusion keeps coincident with the conclusion about F matrix in section 3.1. Therefore, in order to obtain high precision for 3D reconstruction, in the process of image data capture, one should try to capture image data with short distance and short focal length.

2) Only errors of calibration matrix exist

In this case $F_d = 0$, $dx = 0$, $dt = 0$, so:

$$dX = -R K^{-1} K_d K^{-1} x$$

In this formula, as two items of inverse matrix of K exist, calibration matrix error is not sensitive to reconstruction result, this conclusion keep coincident with the practical calculation, as in this formula, one free unknown scale contains, i.e. result of 3D point cloud X is up to scale. If scale is 1:100, than error from camera parameters would be enlarged 100 times. Therefore, in traditional photogrammetry, when survey scale is smaller, final result is very sensitive to camera parameters, so in the camera calibration process, the practice requirement should take into account, smaller scale map survey need higher precision of camera calibration.

4. CALCULATION ERROR ANALYSIS

Above error analysis is based on the model of 3D reconstruction, but this analysis just provides a rough qualitative result from the theory, for the quantitative result of error estimation, one need compute covariance matrix of unknown variables. For one image pair, one simple method of error estimation is to use analyzed result from section 3. Under the assumption that errors of feature points and calibration matrix are known, and the error of feature point has unit weight, one can get covariance matrix for unknown variables R, t, X by following steps:

- 1) Use $F_d = A^{-1} A_d A^{-1} L$, one can get covariance of F relate to covariance of image point;
- 2) Use follow formula, one get covariance of E:

$$E_d = K_d^T F K + K^T F_d K + K^T F K_d$$
- 3) From $E = USV^T$, one get covariance of decomposed elements of E:

$$U_d = E_d V + [0 \ 0 \ U_{d3}] \\ V_d = E_d^T U^T + [0 \ 0 \ V_{d3}] \\ U_{d3} = -V_{d3} = K_d t + K t_d$$
- 4) From $T_d = E_d R^T$, $R = UWV^T$, one can get covariance of R and t;
- 5) Using follow formula, and combine with above covariance, one get covariance of X relate to covariance of image point x and calibration matrix K:

$$dX = V_d V^T X + R^T U U_d R X - R^T K^{-1} K_d (R X + t) + R (K^{-1} dx - dt)$$

Rotation matrix R actually has three independent variables, while above calculation result in 9 nondependent covariance, so one can calculate independent covariance for three angle variables.

For one image sequence, traditional bundle adjustment method could be used for precision estimation (Koch 2003). However, weights for known variables should be determined. We use condition adjustment method with unknown variables on the base of bundle adjustment, and take feature points, calibration matrix as known data, and take R , t and X as unknown data.

Rewrite equation $x = K[R | t]X$ into linear format, and establish normal equation as:

$$\begin{aligned} J_a Q_p J_a^T K + J_b \delta_x + W &= 0 \\ J_b^T G &= 0 \end{aligned} \quad (11)$$

Where J_a J_b are respectively Jacobin parts of known variables and unknown variables, G and δ_x are unknown variables, and δ_x contains unknown errors. Q_p is covariance matrix of feature points and camera intrinsic parameters. As errors of feature points and calibration matrix are known, so it's easy to determine Q_p , from equation (11), one get:

$$\begin{pmatrix} G \\ \delta_x \end{pmatrix} = - \begin{pmatrix} J_a Q_p J_a^T & J_b \\ J_b^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} W \\ 0 \end{pmatrix} = - \begin{pmatrix} Q_{rr} & Q_{rt} \\ Q_{tr} & Q_{tt} \end{pmatrix} \begin{pmatrix} W \\ 0 \end{pmatrix}$$

Where Q_{tt} is covariance matrix of unknown variables.



Fig.2. one image pair of experiment scene

5. EXPERIMENTS

In order to give a demonstration to result precision of 3D reconstruction, we did some experiments. During the data capture process, we follow conclusion in section 3, i.e. we use shortest focal length and shortest distance to take the image sequence. Fig.2 shows one image pair of one building in Peking University. The digital camera we used is Sony DSC-H1, image size is 2592x1944, and the camera is calibrated using Toolbox, which was developed by Klaus and Wolfgang (http://www.vision.caltech.edu/bouguetj/calib_doc/), and the calibration result is:

$$K = \begin{bmatrix} 2792.05 & 0 & 36.59 \\ & 2796.72 & 33.82 \\ & & 1 \end{bmatrix} \quad K_d = \begin{bmatrix} \pm 5.86 & 0 & \pm 7.41 \\ & \pm 5.59 & \pm 6.13 \\ & & 1 \end{bmatrix}$$

Fig.3 shows matched result of feature points on the image pair, assume feature point error is 0.5 pixel, using error propagation rules that stated in above sections, we get an average value for errors of 3D points:

$$X_d = [-0.228 \quad -0.098 \quad -0.038]$$

As different 3D point has different error, so this is an approximate value, and this error is deduced from original error of feature points and calibration matrix, so its unit is pixel, but it contains on unknown scale, so it need to multiply one geography scale before it's used to estimate real error.



Fig.3. Matched result of feature points



Fig.4. Marked feature points

In order to verify practical precision, we measured the width and height of one monument in front of the building, and distance between pillars on ground floor are also measured. In

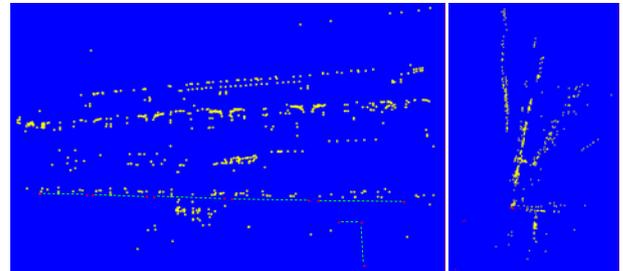


Fig. 5. 3D point cloud from reconstruction (right is lateral view)

fig.4, distances between pillars marked 0-1, 2-3, 4-5, 6-7 and 8-9 are respectively: 343.0cm, 343.0cm, 373.5cm, 343.0cm, 343.0cm. The measurement error is under 1cm. Mounment has



Fig.6. one image sequence of experiment scene

width marked 10-11 and height marked 11-12 are respectively 83.85cm, 168.80cm, the measurement error is under 0.2cm.

Fig.5 shows 3D point cloud from reconstruction, as the monument is measured with high precision, so we calculate the unknown scale using its reconstruction data and measured data, the result scale is 1:435.61. Multiply this scale, we get reconstruction distances between pillars respectively are: 330.26cm, 348.40cm, 383.81cm, 349.88cm, 343.12cm (in fig.5, dash lines shows the reconstructed distances). Compare the reconstructed result and measured result, we find that the monument size and distance between pillars in the back of the monument have very high precision, which no bigger than 1cm. While the error of distance between pillars that have biggest distance from viewpoint is the largest, which reach to 13cm, the average error is 7.1cm. Fig.3 shows two image matched result, one can find that the baseline is small, that means result is sensitive to the distances.

Multiply X_d with known scale 1:435.61, we get:

$$X_d = [-99.319 \ -42.69 \ -16.55]$$

Considering image resolution is 72dpi, convert it into geographic size: $X_d = [-34.5 \ -14.82 \ -5.75]$ mm, i.e. X position error is about 3.8cm. As it's difficult for us to measure large amount data for verify the difference between deduced errors and measured errors, but 5 measured points has error 7.1cm is on the same degree as deduced error 3.8cm. This result is very well if it's to be used in visualization process. In fact, in the system, such as 3D city models, one model with such as precision is really accurate, but in survey industrial, the majority of application can be satisfied.

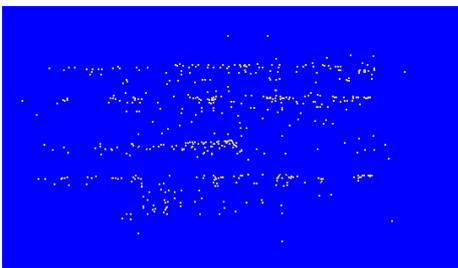


Fig.7. 3D point cloud reconstructed from the image sequence

In order to promote precision of 3D reconstruction, one can use same method in traditional photogrammetry, and could promote precision by increase observe times and length of baselines. Therefore, one image sequence is required as shows in fig.6. By using this image sequence, we get reconstruction result shows in fig.7, under the assumption that all unknown variables have equal weights, we use bundle adjustment method to optimize the result, finally we found that errors of all distance between pillars are no large than 5cm.

5cm distance in 3D space equals 2.89cm coordinates error, considering errors of image point, camera intrinsic parameters and marked points, this result is an ideal result.

For the limitation of our experiment condition, it's hard for us to verify more high precision, in order to give an estimation from the theory for the result precision of 3D reconstruction, and under the assumption that camera precision is promote 10 times, and image point error limit to 0.1 pixel, which corresponding to unit weight, then we deduce covariance of unknown variables R, t, and X, and estimate the precision of 3D result, still multiply with scale 1:435.61, we get the final result is 0.83mm. this result completely could satisfy industrial requirement. But this is just one estimate result. In the practice, more factors should be taken into account for very high precision.

6. CONCLUSION

In this paper, we give an clear answer to the precision that 3D reconstruction could reached, i.e. without any control points and any manual operation, by using one common digital camera, one can get result of 3D reconstruction from one image sequence, its error would limit to 10cm. If control point still not be used, and want to promote reconstruction precision, one need to resolve: 1) calibrate camera with very high precision, e.g. control error of focal length and principal point no larger than 1 pixel, and distortion coefficient should also be calibrated; 2) robust feature extraction algorithm should be used, which could ensure error of image point no large than 0.1 pixel; 3) obtain enough images if data capture condition permit, and bundle adjustment method should be used to optimize the result. In addition, select one digital camera with good capabilities is also very important. As a summary, it's feasible to apply automatic 3D reconstruction method in close range photogrammetry field, but very high precision application (e.g error is around 0.5mm) is still need to be verify and do further research.

ACKNOWLEDGEMENT

The authors hereby deliver their thanks to National Science Foundation of China for its financial supports to this research work of project No: 40471104.

REFERENCES

- Blostein S. D., Huang T S., 1987, *Error analysis in stereo determination of 3D point position*. IEEE Transactionson Pattern Analysis and Machine Intelligence, 9(6): 752-765
- Faugeras O., et al., 1992, *Camera Self-Calibration: theory and experiments*. In Proc. European Conf. On Computer Vision. Lecture Notes in Computer Science, Vol.588, p321-334, Springer

- Förstner W. , 2002, *Computer Vision and Photogrammetry -- Mutual Questions: Geometry, Statistics and Cognition*, in: Bildtechnik/Image Science, Swedish Society for Photogrammetry and Remote Sensing, p151-164
- Förstner W., 2005, *Uncertainty and Projective Geometry*, in: Handbook of Geometric Computing, Bayro Corrochano, Eduardo (Ed.), Springer 2005, ISBN: 3-540-20595-0, pp.493-535.
- Mcvey E. S., 1982, *Some accuracy and resolution aspects of computer vision distance measurement*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 4(6): 646-649
- Pollefeys M., et al, 2004, *Visual modeling with a hand-held camera*, International Journal of Computer Vision, 59(3), 207-232
- Triggs B. 1997, *Auto-calibration and the absolute quadric*. In Proc. Conf. on Computer Vision and Pattern Recognition, Puerto Rico, 609-614
- Hartley R., Zisserman A. 2000, *Multiple View Geometry in Computer Vision*, Cambridge University press.
- Koch R., 2003, *3D-Scene modeling from image sequences*, ISPRS Archives, Vol. XXXIV, Part 3/W8, Munich, Sept. 2003, p17-19.
- Mayer H., 2003, *Robust Orientation, Calibration, and Disparity Estimation of Image Triplets*. DAGM-Symposium, p281-288
- Werner T., et al., 2001, *Automated Architecture Reconstruction from Close-Range Photogrammetry*, In: Proc. on CIPA 2001 International Symposium: Surveying and Documentation of Historic Buildings, Monuments, Sites, Traditional and Modern Methods
- ZHANG Z., 1999. *Flexible camera calibration by viewing a plane from unknown orientations*. Proceedings of the 7th IEEE International Conference on Computer Vision. Corfu, Greece: IEEE Computer Society Press, p666-673.