# Semi-automatic metric reconstruction of buildings from self-calibration: Preliminary results on the evaluation of a linear camera self-calibration method*

D. Q. Huynh† Y. S. Chou‡ H. T. Tsui‡

† School of Information Technology
Murdoch University
Perth WA 6150 Australia
d.huynh@murdoch.edu.au

‡ Department of Electronic Engineering
Chinese University of Hong Kong
Shatin, Hong Kong
{yschou,httsui}@ee.cuhk.edu.hk

## Abstract

*In this paper, we investigate the linear self-calibration method proposed by Newsam et al [7] for our project on 3D reconstruction of architectural buildings. This self-calibration method assumes that the principal point is known, the camera has square pixels and has no skew. It allows 3D shape to be reconstructed from two images while giving the camera the freedom to vary its focal length. Since the paper by Newsam et al reports only the theoretical work on camera self-calibration, in this paper, we evaluate the focal lengths obtained from their method with both synthetic data and real data. In real data where known 3D data are available, Tsai's calibration method is used for comparison. Our experimental results show that the focal lengths from the two methods differed by less than 5% and the reconstructed 3D shape was very good in that angles were well preserved. Future research will focus on improvement of 3D reconstruction in the presence of small image noise and further develop this method into a package for 3D reconstruction of buildings to be used by a layperson.*

## 1. Introduction

It is now widely known that given a sufficient number of corresponding points the fundamental matrix F can be recovered from corresponding points alone. The 7 degrees of freedom property of F allows only 5 extrinsic parameters and 2 intrinsic parameters to be retrieved. To estimate more intrinsic camera parameters one must approach the problem by considering more images (of a static scene) while keeping the camera setting invariant [1] or assume that certain properties of the cameras are known (e.g. no skew, the pixels are square). In the latter approach, an additional assumption that the principal point is known has been incorporated [2, 7, 8]. This latter approach can be taken to be camera self-calibration for a partially calibrated camera whose focal length[1] is variable or camera self-calibration for two distinct partially-calibrated cameras.

The primary aim of our project is to reconstruct architectural buildings from partially calibrated images. The system to be built will be semi-automatic in that prominent image features will be automatically detected by a feature detector but a human operator will be involved to do some manual editing to the image correspondences, if necessary. More image feature correspondences will be automatically established, after the epipolar geometry is recovered, to achieve a dense reconstruction.

From a pair of images taken by a partially calibrated camera (We will not distinguish the case where two cameras are involved with the case where one camera which undergoes motion and whose focal length is variable. For a scene that contains an architectural building, the scene is static and the two cases above are identical) to the final metric reconstruction, a number of steps are involved: *(i)* partially calibrate the camera to estimate the principal point, *(ii)* estimate the epipolar geometry by optimally computing the fundamental matrix, *(iii)* retrieve the two unknown focal lengths of the images involved from the fundamental matrix, *(iv)* compute the extrinsic parameters or relative orientation between the two images for triangulation, *(v)* recover the 3D information of each pair of image corresponding points. To ensure that the final reconstruction in step *(v)* is optimal, the computation in all the precedent steps must be optimal. In this paper, we will present our preliminary results on the study of some of the aforementioned steps. In particular, we will use the linear self-calibration method proposed by Newsam et al [7] and will focus more

*This research was supported by the RGC grant CUHK4310/98E and in part supported by the Murdoch Special Research Grant MU.AMH.D.413.

[1]The term *focal length* here means the *effective focal length* for a pinhole camera model. This is different from the focal length of the lens in optics. Photogrammetrists use the term *principal distance* which may cause less confusion.

on step *(iii)* above. We chose to work on this method because the original paper [7] is a theoretical paper without experimental evaluation. More importantly, their method has a number of advantages as described below. First, it allows general camera motion which makes it possible for using a hand-held camera for 3D reconstruction; second, as it is an essentially linear algorithm, it is computationally efficient; third, it allows the focal length to vary so the camera can freely zoom in and out of the scene and has no restriction on its viewing distance and angle to the object(s) of interest. We hope to further develop this method into a package for 3D reconstruction of buildings used by a layperson. We will present our initial 3D reconstruction in the form of sparse 3D points at this stage. Development of a hybrid intensity-based and partial model-based stereo matching system is currently underway for dense 3D reconstructions.

## 2. Metric reconstruction from partially calibrated images

### 2.1. Estimating the camera principal point

The term *principal point* used in this paper describes the intersection point of the optical axis with the image plane. For images captured by a digital video camera, the assumption that the principal point is at the centre of the image frame has been used by many researchers, although for non-imaging process (e.g. digitised photographs) [11] the image centre is often unrelated to the principal point of the camera.

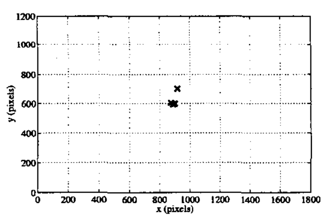


**Figure 1. The estimated principal point coordinates from Tsai's calibration method.**

We use a Fujitsu digital video camera with a zoom lens for our building reconstruction task. To estimate the location of the principal point of our camera, we conducted a number of experiments using Tsai's method [9] to calibrate the principal point that is required by Newsam et al's method [7]. Fig. 1 shows the estimated principal point coordinates for our digital camera whose image buffer is $1800 \times 1200$ pixels. Discarding the two principal points $(872.58, 604.06)$ and $(916.33, 698.64)$ that are slightly off the center of image frame, the average principal point was computed to be $(896.33, 598.64)$, which is very close to $(900, 600)$, the centre of the image buffer. In these experiments, the focal length varies from 995.72 to 4681.52 pixels due to zooming. Lavest et al [5] reported that the principal point was very stable under the change of the camera zoom. We would like to note that the principal point can move slightly when the camera changes its focus setting but for a good digital video camera this movement is fairly small. At the time of writing, we are conducting similar experiments with a Sony DCR-PC100 digital video camera. We hope to report further findings on the principal point of this camera in the future.

### 2.2. Optimal computation of F

The essential element of a good 3D reconstruction is an optimally computed fundamental matrix for the recovery of the epipolar geometry. Hartley [3] reports estimating the fundamental matrix using SVD with the image coordinates normalised. Since the linear method only minimises the algebraic error which has no meaningful geometric interpretation, nonlinear minimisation with a proper objective function must be sought. Luong and Faugeras [6] examine two minimisation criteria for the nonlinear method. These criteria, together with a few others, have been implemented by Zhang et al [12] in their `FMatrix` program. Thanks to Zhang for making the program available on the web. We were able to use it to conduct our experiments (see next section) on focal length recovery from synthetic and real data.

### 2.3. Linearly recovering two focal lengths from F

By assuming that the principal point is known (so the origin of the image coordinate system can be set at the principal point) and the camera contains square pixels, the camera matrices $A$ and $A'$ for the two viewing positions can be simplified to diagonal matrices. This allows the extrinsic parameters to be eliminated nicely from the $3 \times 3$ matrix $FF^{\top}$ and leads to a linear self-calibration method for recovering two focal lengths. The full algorithm of this linear method and the proof of two classes of degenerate stereo configurations for self-calibration are reported in [7].

Experiments reported in this paper focus on focal length recovery using the linear self-calibration method described above. We carefully set up the experiments such that the degenerate stereo configurations (especially for class 1) mentioned in [7] were avoided (e.g. by enforcing a (small) tilt angle between the two camera orientation).

### 2.4. Essential matrix and triangulation

Having recovered the focal lengths, the essential matrix E can be estimated easily using the formula $E = A'^{-\top} F A^{-1}$ Our current version of triangulation for 3D reconstruction still has room for optimisation and is part of the on-going work of our project. Triangulation in the presence of image noise has been discussed by Weng et al [10] and recently by Hartley and Sturm [4]. We will conduct further investigation on this issue.

## 3. Experiments and discussion

For experiments with synthetic images where the true focal lengths were known, different levels of Gaussian noise of zero mean (the noise standard deviation, $\sigma$, varied from 0.1 to 1.0) were added to the true coordinates of the corresponding image points in both the left and right images. The perturbed image coordinates were passed to Zhang's `FMatrix` program for fundamental matrix computation and then focal length computation using Newsam et al's method. We had experimented the various minimisation criteria provided by `FMatrix`. We found that the criterion (`-ng`) that normalises (as mentioned in [3]) the corresponding point coordinates and minimises the gradient-weighted epipolar errors gave the best fundamental matrix, which in turn yielded the smallest focal length errors (Fig. 2(a)). The criterion (`-nnl`) that performs the normalisation and minimises the distances of corresponding points to the epipolar lines gave slightly larger focal length errors (Fig. 2(b)). The number of corresponding points used was 18. They were well distributed in the two simulated images of $1800 \times 1200$ pixels. The principal point was assumed to be known. The true focal lengths were 1720.40 and 1598.57 pixels. Each point in the plot is the average error of 50 simulations. Our experiments on synthetic data reveal that the fundamental matrix is susceptible to error in the presence of image noise. Furthermore, any perturbations to the image coordinates that are larger than 3 pixels may give an outlier effect to the fundamental matrix estimation. An outlier detection and elimination procedure is therefore essential if a fully automated building reconstruction system is desired.

(a) The `-ng` criterion (b) The `-nnl` criterion

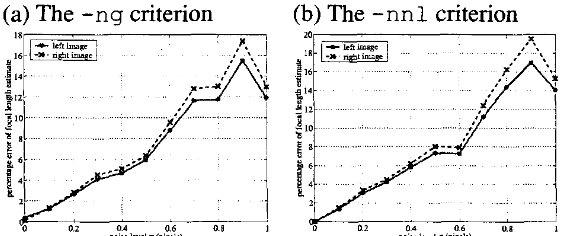


**Figure 2. Percentage error of focal lengths against image noise.**

Experiments on real images of indoor and outdoor scenes were also conducted. Images of indoor scenes were fully calibrated with a calibration target and the application of Tsai's method [9]. The idea was to compare the estimated focal lengths from Newsam et al [7] with those from Tsai [9] where true 3D data were available.

Fig. 3 shows a pair of images of a calibration target, with a number of corresponding points superimposed, in one of our indoor experiments. The calibration target has two orthogonal surfaces. The image on the left is frame 1 and the image on the right is frame 30 from an image sequence.

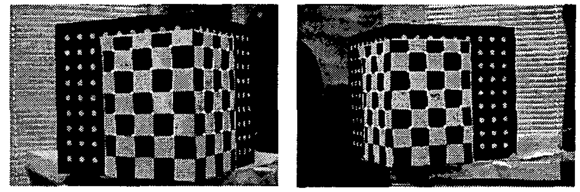

**Figure 3. An image pair of a calibration target.**

Feature points were detected and tracked by a corner detector with some manual editing as a post-process. Using the mean value of the estimated principal points reported in Section 2.1 for the linear algorithm [7], the estimated focal lengths for the two methods for nine different experiments are plotted in Fig. 4. The best fitted line to the computed

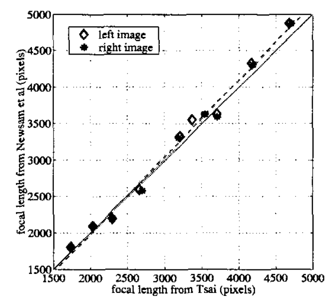


**Figure 4. The estimated focal lengths from the two methods.**

focal lengths is shown as a dashed line. Its slope was computed to be 0.95, which corresponds to an angle of inclination of 43.55°. The percentage error of the angle of inclination from the 45° line (solid diagonal line) is 3.23%. The vertical intercept of the fitted line is $-113.33$ pixels. One may argue that as we move outside the focal length interval [1500, 5000] the two diagonal lines will be further apart (see Fig. 4). However, it is unlikely that the camera will have its focal length significantly below 1500 or above 5000 pixels as neither can the focal length of a camera vanish nor can it, for a perspective camera model, be infinite. Moreover, it is simply meaningless to extrapolate the errors in an error analysis this way. The results shown in Fig. 4 demonstrate that the linear algorithm [7] performs well in comparison with the calibration results from Tsai's method for a wide range of focal lengths.

The focal lengths estimated from the two methods for the nine experiments are tabulated in Table 4, in which each entry for the percentage error was computed as $\frac{f_N - f_T}{f_T}$, where $f_T$ and $f_N$ are the focal lengths from Tsai and Newsam et al respectively. The reconstruction of the sparse 3D points on the calibration target is shown in Fig 5. The angle between the two surfaces of the calibration target was estimated to be 88.10°, corresponding to an error of 2.11%. Since the true values of the focal lengths were not known in these real experiments and since Tsai's focal length estimates may contain small errors, the comparison above is only relative.

| Left image | | | Right image | | |
|---|---|---|---|---|---|
| Tsai | Newsam et al | %error | Tsai | Newsam et al | %error |
| 4679.38 | 4877.61 | 4.24 | 4708.14 | 4876.17 | 3.57 |
| 4168.55 | 4325.65 | 3.77 | 4188.58 | 4310.44 | 2.91 |
| 3706.14 | 3632.60 | −1.98 | 3712.18 | 3593.14 | −3.21 |
| 3375.03 | 3549.70 | 5.18 | 3544.41 | 3628.50 | 2.37 |
| 3201.23 | 3323.73 | 3.83 | 3184.26 | 3307.75 | 3.88 |
| 2657.89 | 2598.74 | −2.23 | 2697.37 | 2581.35 | −4.30 |
| 2290.41 | 2201.54 | −3.88 | 2287.37 | 2192.26 | −4.16 |
| 2025.87 | 2097.49 | 3.54 | 2038.49 | 2093.15 | 2.68 |
| 1730.10 | 1803.67 | 4.25 | 1732.42 | 1802.12 | 4.02 |

**Table 1. The computed focal lengths (in pixels) for the 9 image pairs.**

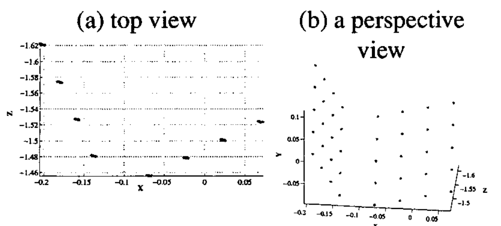


**Figure 5. Metric reconstruction of the calibration target.**

Fig. 6 shows a pair of images and the metric reconstruction of a building which has a large curved surface whose shape is a section of a cylinder. Using the self-calibration method, the focal lengths of the left and right images were computed to be 1804.30 and 1841.90 pixels. A good conic fitting program will be required to assess the reconstructed 3D shape in this experiment.

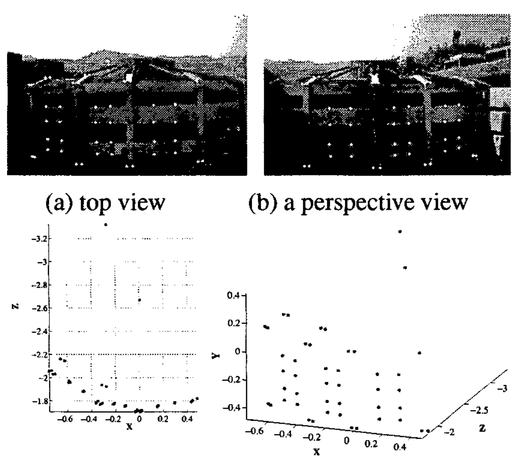


**Figure 6. An image pair and the metric reconstruction of a building.**

## 4. Conclusion

The linear method of Newsam et al [7] for recovering focal lengths in self-calibration has a number of advantages as discussed in the Introduction section of this paper. Our preliminary results show that the method gives good estimates of the focal lengths if the fundamental matrix has been accurately computed. We have also shown that using the computed focal lengths the 3D structure of a building can be achieved with a sufficiently good accuracy for visualisation. We believe that the reconstruction can be further improved if an optimal triangulation procedure is adopted. At the time of preparing this manuscript, we have not been able to achieve focal length estimation using other self-calibration methods, such as those of Hartley [2] and Pollefeys et al [8]. This will be part of our future research, in addition to the requirement to develop and enhance our method into an easy-to-use package for 3D reconstruction of buildings.

## References

[1] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera Self-Calibration: Theory and Experiments. In *Proc. ECCV*, pages 321–334, May 1992.

[2] R. I. Hartley. Estimation of Relative Camera Positions for Uncalibrated Cameras. In *Proc. ECCV*, pages 579–587, May 1992.

[3] R. I. Hartley. In Defence of the 8-point Algorithm. In *Proc. ICCV*, pages 1064–1070, Jun 1995.

[4] R. I. Hartley and P. Sturm. Triangulation. In *ARPA*, 1994.

[5] J.-M. Lavest, G. Rives, and M. Dhome. Three-Dimensional Reconstruction by Zooming. *IEEE Trans. on Robotics and Automation*, 9(2):196–207, April 1993.

[6] Q. T. Luong and O. D. Faugeras. The Fundamental matrix: Theory, Algorithms, and Stability Analysis. *IJCV*, 17:43–75, 1996.

[7] G. N. Newsam, D. Q. Huynh, M. J. Brooks, and H.-P. Pan. Recovering Unknown Focal Lengths in Self-Calibration: An Essentially Linear Algorithm and Degenerate Configurations. In *International Archives of Photogrammetry and Remote Sensing*, volume XXXI, part B3, commission III, pages 575–580, Jul 1996.

[8] M. Pollefeys, R. Koch, and L. V. Gool. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. In *Proc. ICCV*, pages 90–95, Jan 1998.

[9] R. Y. Tsai. A Versatile Camera Calibration Technique For High-Accuracy 3D Machine Vision Metrology Using Off-The-Shelf TV Cameras And Lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, Aug 1987.

[10] J. Weng, T. S. Huang, and N. Ahuja. *Motion and Structure from Image Sequences*. Springer-Verlag, 1993.

[11] R. G. Willson and S. A. Shafer. What is the Center of the Image? In *Proc. CVPR*, pages 670–671, 1993.

[12] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A Robust Technique for Matching Two Uncalibrated Images through the Recovery of the Unknown Epipolar Geometry. *Artificial Intelligence*, 75(1-2):87–120, 1995.