Efficient Iterative Solution to M-View Projective Reconstruction Problem

Qian Chen and Gérard Medioni*
University of Southern California, Los Angeles, CA 90089-0273
{chenqian, medioni}@iris.usc.edu

Abstract

We propose an efficient solution to the general M-view projective reconstruction problem, using matrix factorization and iterative least squares. The method can accept input with missing data, meaning that not all points are necessarily visible in all views. It runs much faster than the often-used non-linear minimization method, while preserving the accuracy of the latter. The key idea is to convert the minimization problem into a series of weighted least squares sub-problems with drastically reduced matrix sizes. Additionally, we show that good initial values can always be obtained. Experimental results on both synthetic and real data are presented. Potential applications are also demonstrated.

1. Introduction and Related Work

In this paper, we address the problem of structure recovery given correspondence information. Particularly, we are interested in recovering the non-metric structure from point correspondences in m uncalibrated cameras. This problem, known as *projective reconstruction*, has been approached from three directions previously.

First, it has been established that certain linear forms exist among 2, 3 or 4 views: the *fundamental matrix*, the *trifocal tensor* [7,15] and the *quadrilinearity* [5,18], with the negative result that further views do not add any new linear constraints. Reconstruction algorithms based on the fundamental matrix [4,6] or trifocal tensor [8] have been reported. A comparison of methods for a pair of views is given in [13]. The form of quadrilinearity in [5,18], however, is highly redundant: it has 81 coefficients, whereas a view contributes at most 12 parameters. As a result, its existence is mainly of theoretical significance.

A second, and more generic, form of quadrilinearity appears as the rank-4-ness of the so-called *scaled measurement matrix* which naturally extends itself to factorization type algorithms [16]. Suppose there are m perspective cameras: P_i, i=1,…,m, and and n projective points: X_j, j=1, …, n. Using the symbol "~" to denote equality up to a scale, the following holds

$$[u_{ij}, v_{ij}, 1]^T \sim P_i X_j, \text{ or } \lambda_{ij}[u_{ij}, v_{ij}, 1]^T = P_i X_j \tag{1}$$

where λ_{ij} is a non-zero scale factor (projective depth) which in general depends on both the camera and the point. The equivalent matrix form is:

$$W_s = \begin{bmatrix} \lambda_{11}\begin{bmatrix} u_{11} \\ v_{11} \\ 1 \end{bmatrix} & \cdots & \lambda_{1n}\begin{bmatrix} u_{1n} \\ v_{1n} \\ 1 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ \lambda_{m1}\begin{bmatrix} u_{m1} \\ v_{m1} \\ 1 \end{bmatrix} & \cdots & \lambda_{mn}\begin{bmatrix} u_{mn} \\ v_{mn} \\ 1 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix}_{3m\times 4} \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}_{4\times n} \tag{2}$$

where W_s is the just mentioned scaled measurement matrix. Thus, for points in general positions, a **Rank-4-Factorization** of W_s produces the projective reconstruction. The algorithm described in [16,19] computes the scale factors progressively, by using the fundamental matrix between each pair of successive views. A variation of the algorithm was given in [1]. Their common shortcoming is that it requires W_s be a *full* matrix, meaning that *all* points must be seen in *all* views. Although a method that dealt with missing data was reported in [10] for affine cameras, it was not clear how to extend the method to the perspective case, due to the existence of the scale factors.

From a third direction, adapting the traditional idea of *bundle adjustment* from Photogrammetry, Mohr *et al* [12] proposed a non-linear algorithm by solving

$$\min\left(\sum_i \sum_j \left\{ \left(u_{ij} - \frac{P_{i(1)}^T X_j}{P_{i(3)}^T X_j} \right)^2 + \left(v_{ij} - \frac{P_{i(2)}^T X_j}{P_{i(3)}^T X_j} \right)^2 \right\}\right) \tag{3}$$

where $P_{i(1)}^T, P_{i(2)}^T, P_{i(3)}^T$ are the row vectors of P_i, with the constraints $\|P_i\|$=1 and $\|X_j\|$=1. Since the size of the resultant Jacobian matrix can be large, inverting it may be time-consuming. More severely, our implementation showed that the Jacobian matrix was constantly rank deficient, causing slow convergence (due to the gradient

* This research has been funded [or funded in part] by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152, with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency.

descent part of the Levenberg-Marquardt method). This indicated that some inherent constraints are not utilized. Two such cases have so far been identified. Consider the following matrix formed by accumulating column-wise the homogeneous coordinates of l points:

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_l \\ y_1 & y_2 & \cdots & y_l \\ z_1 & z_2 & \cdots & z_l \\ w_1 & w_2 & \cdots & w_l \end{bmatrix}.$$

If all these points are collinear, only two of them are independent, thus matrix A is of rank 2. Equivalently, the determinant of any order-3 or -4 minor of A is zero. If the points are coplanar, that of any order-4 minor is zero. Although it is easy to check the collinearity condition, to check for coplanarity, one has to compute the fundamental matrix. In addition, the above constraints amount to degree 3 or 4 polynomials, enlarging the non-linearity of the minimization function.

In this paper, we develop a new, efficient solution for multi-view projective reconstruction, which addresses some issues associated with the existing methods, that is, missing data handling and degenerate Jacobian matrix. The basic idea is to treat (3) as a separable problem [3] by alternatively holding P_i's and X_j's constant. It turns out that this treatment drastically reduces the problem size (compared to performing pseudoinverse on the Jacobian of (3)) while preserving the accuracy of the earlier non-linear method.

The paper is organized as follows. Section 2 describes the initialization step using matrix factorization, and how to improve its quality. Section 3 describes the proposed algorithm with a comparison to the factorization algorithm. Experimental results on real data are presented in Section 4. In section 5, two potential applications based upon our algorithm are demonstrated. The paper is summarized in section 6.

2. Initialization

As any iterative algorithm, ours requires initial values. In [12], five non-coplanar points are assigned the standard projective coordinates [1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1] and [1,1,1,1]. A planarity test is conducted to make sure the chosen points are not coplanar. It is claimed that for the remaining data the initial values are not critical. Here, we propose to first extract a full *measurement matrix*

$$W = \begin{bmatrix} \begin{bmatrix} u_{11} \\ v_{11} \\ 1 \end{bmatrix} & \cdots & \begin{bmatrix} u_{1n} \\ v_{1n} \\ 1 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ \begin{bmatrix} u_{m1} \\ v_{m1} \\ 1 \end{bmatrix} & \cdots & \begin{bmatrix} u_{mn} \\ v_{mn} \\ 1 \end{bmatrix} \end{bmatrix} \qquad (4)$$

from the image data and then perform a rank-4-factorization on W, which produces an approximate projective reconstruction. Those points and cameras with missing data are initialized by triangulation using the thus obtained reconstruct.

There are several reasons why we choose such a factorization method for initialization over that of Mohr's. First, to perform the planarity test, the fundamental matrix must be known a priori. Second, it is preferable to treat all data uniformly rather than single out a few of them as "privileged". Third, it will be shown in the next two subsections that either, in some special case, the rank of W is by default 4 or it can be made very close to 4 by repeating the factorization process.

2.1 When is W rank 4?

The answer is given by the following theorem and its corollary.

Theorem. *Let W_s be a scaled measurement matrix with the proper scale factors λ_{ij}, for i=1, …, m and j=1, …, n. Then the corresponding measurement matrix W is rank 4 if and only if there are m+n independent coefficients p_i, $1\leq i\leq m$, x_j, $1\leq j\leq n$, where p_i is related to the i^{th} camera and x_j is related to the j^{th} point, such that $\lambda_{ij} = p_i x_j$.*

The proof is given in the Appendix. ♦

Corollary. *Under the special configuration where all cameras' projection centers are coplanar, and their principal axes are parallel, the resultant measurement matrix is of rank 4.*

Proof.

Let A_i represent the internal camera parameters, R_i and T_i be the external parameters. The projection process is

$$\begin{bmatrix} u_{ij} \\ v_{ij} \\ 1 \end{bmatrix} \sim P_i \begin{bmatrix} X_j \\ 1 \end{bmatrix} = \begin{bmatrix} A_i & 0 \end{bmatrix} \begin{bmatrix} R_i & T_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_j \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} au_i & \omega_i & u0_i & 0 \\ 0 & av_i & v0_i & 0 \\ 0 & 0 & 1/f_i & 0 \end{bmatrix} \begin{bmatrix} R_{i(1)}^T & t_{i(1)} \\ R_{i(2)}^T & t_{i(2)} \\ R_{i(3)}^T & t_{i(3)} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_j \\ 1 \end{bmatrix}$$

$$= \frac{R_{i(3)}^T X_j + t_{i(3)}}{f_i} \begin{bmatrix} u_{ij} \\ v_{ij} \\ 1 \end{bmatrix} \quad (5)$$

where $R_{i(3)}$ is actually the camera's principal axis in the world, and $t_{i(3)}$ is the camera translation in this direction. If the scale factor λ_{ij} is chosen to be $(R_{i(3)}^T X_j + t_{i(3)})/f_i$, then (5) is an indication that the resultant scaled measurement matrix is of rank 4. In addition, the special configuration implies that, i) all the image planes have the same distance to the world origin, thus $t_{1(3)}=t_{2(3)}=\ldots=t_{m(3)}=t$; ii) $R_{1(3)}=R_{2(3)}=\ldots=R_{m(3)}=V$. Therefore $\lambda_{ij}=(1/f_i)(V^T X_j+s)=(1/f_i)z_j$. Based on the theorem, it is sufficient to conclude that the corresponding measurement matrix is of rank 4.♦

We make two remarks:

1) The parallel condition is somewhat optional–multiple views with coplanar projection centers (for binocular and trinocular stereo, this is naturally true) can always be rectified into such a configuration [17].
2) Under general configurations, the inner product appearing in the scale factor of (5) cannot be separated into two multipliers, thus a rank-4-factorization does not exist.

2.2 Making *W* close to rank 4

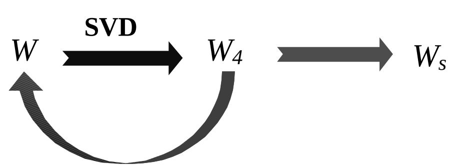


Figure 1. Illustration of IFA

Let W_4 be the matrix obtained from W by keeping the four largest singular values. We argue that W_4 is somewhere "between" W and W_s because it is an approximation of W and is of rank 4. We make a further argument that W can be brought "closer" to W_s by scaling it with the coefficients of W_4. We repeat this "SVD–rescaling" process, trying to bring W closer to W_s until there is no obvious gain on the rank-4-ness of the updated W, which is indicated by the ratio of the fifth and the fourth singular values of the W. This process is depicted in Figure 1. For reference, we call it the **Iterative Factorization Algorithm (IFA)**. Similar ideas were also explored by Berthilsson, *et al*. [1].

IFA was tested on a synthetic data set generated from 8 cameras and 100 points randomly distributed over a sphere. Camera parameters were set to simulate real cameras with an image size of 1000×1000. The cameras were initially arranged to be parallel and coplanar, and are subsequently perturbed. The perturbation was based on a uniform distribution. The distance variation was the percentage of the initial distance between the camera and the centroid of the points. The angular variation was the number of degrees of the Euler angle. Figure 2 shows the result when setting the perturbation level to 20, meaning up to ±20% in distance change and ±20° in orientation change. The fifth singular value drops from the order of 10^0 to 10^{-7} in 63 iterations, while the first four are almost unchanged.

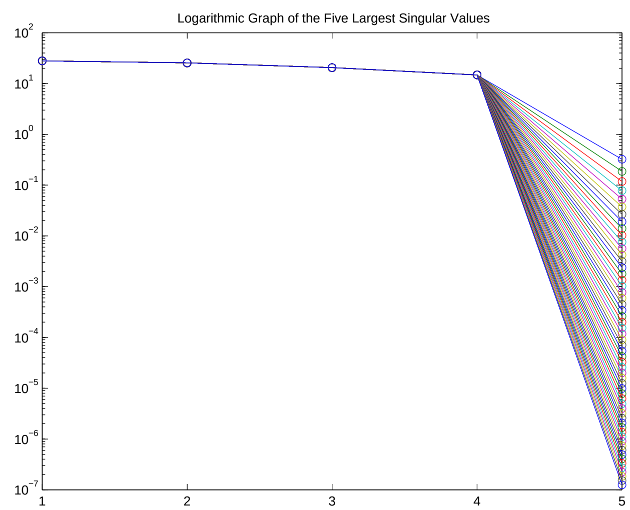


Figure 2. Change of the five largest five singular values

3. The Iterative Algorithm

While addressing the missing data issue, projective bundle adjustment would be an ill-posed problem if it were formulated as in (3). Our experiments show that the source of the degeneracy lies in the attempt to minimize the projections and the points at the same time. In fact, one can treat (3) as a separable problem [3] by alternatively holding P_i's and X_j's constant. Since the terms involving points are independent to each other (the corresponding Jacobian matrix is block-diagonal), each term (thus each point) can be minimized separately. This is true for projections as well. Hence, we have a novel way of minimizing (3): starting from some initial guess, we repeatedly perform **intersection** for each point followed by **resection** for each camera. Furthermore, both sub-problems have least squares solutions.

3.1 Algorithm description

Eliminating the scale factors in (1), we have the following basic equations

$$u_{ij} - \frac{P_{i(1)}^T X_j}{P_{i(3)}^T X_j} = 0 \text{ and } v_{ij} - \frac{P_{i(2)}^T X_j}{P_{i(3)}^T X_j} = 0$$

for i=1, …, m, and j=1, …, n, which can also be written in linear forms as

$$\begin{bmatrix} u_{1j} P_{1(3)}^T - P_{1(1)}^T \\ v_{1j} P_{1(3)}^T - P_{1(2)}^T \\ \dots \\ \dots \\ u_{mj} P_{m(3)}^T - P_{m(1)}^T \\ v_{mj} P_{m(3)}^T - P_{m(2)}^T \end{bmatrix}_{2m\times 4} X_j = AX_j = 0 \qquad (7a)$$

for all points X_1, …, X_n, and

$$\begin{bmatrix} X_1 & 0 & -u_{i1}X_1 \\ 0 & X_1 & -v_{i1}X_1 \\ \dots & & \\ \dots & & \\ X_n & 0 & -u_{in}X_n \\ 0 & X_n & -v_{in}X_n \end{bmatrix}_{2n\times 12} \begin{bmatrix} P_{i(1)} \\ P_{i(2)} \\ P_{i(3)} \end{bmatrix} = B \begin{bmatrix} P_{i(1)} \\ P_{i(2)} \\ P_{i(3)} \end{bmatrix} = 0$$

(7b)

for projections P_1, …, P_m. The solution is the null vector of the corresponding coefficient matrix, which in return can be thought of minimizing

$$fx' = \sum_i \left\{ \left(u_{ij} P_{i(3)}^T X_j - P_{i(1)}^T X_j \right)^2 + \left(v_{ij} P_{i(3)}^T X_j - P_{i(2)}^T X_j \right)^2 \right\}$$

and

$$fp' = \sum_j \left\{ \left(u_{ij} P_{i(3)}^T X_j - P_{i(1)}^T X_j \right)^2 + \left(v_{ij} P_{i(3)}^T X_j - P_{i(2)}^T X_j \right)^2 \right\} \qquad (8)$$

The difference between (3) and (8) is that algebraic errors are minimized in the latter. Therefore the solution from (7) may be biased. Hartley [9] suggested a correction for this type of problems by repeatedly reweighting each equation in (7) by $1/P_{i(3)}^T X_j$ until the change of weights is negligible (we observed that 5 loops were sufficient). This scheme is implemented in the **Weighted Iterative Eigen** (**WIE**) algorithm (Figure 3). It so named because of the way (7) is solved: the solution is the eigenvector of $A^T A$ (or $B^T B$) corresponding to the smallest eigenvalue. Besides addressing the rank deficient problem mentioned earlier, this algorithm has another advantage−the sizes of the involved matrices are drastically reduced. For example, in (2), the size of W_s is 3mn. In (3), the size of the Jacobian matrix is $2mn \times (12m + 4n)$. But in WIE, as shown in (7), A is of $2m \times 4$ and B is of $2n \times 12$.

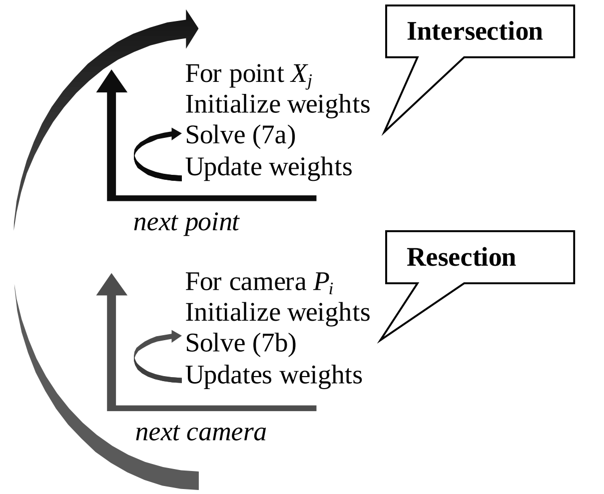


Figure 3. Illustration of the WIE algorithm

3.2 Experiment on synthetic data

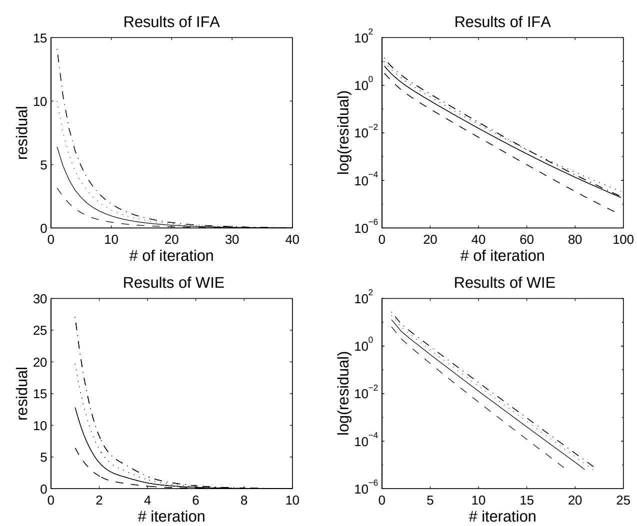


Figure 4. Comparison of IFA and WIE on synthetic data Perturbation levels: dash-dot-20, dot-15, solid-10, dash-5

To test the algorithm, the same synthetic data used in 2.2 were applied here, but with different perturbation levels. Three steps of SVD were run to generate the initial values. The final results are shown in the second row of Figure 4. For comparison, results from IFA are shown in the first row. We observe that i) both algorithms converge linearly since the logarithmic curves are almost lines; and ii) WIE converges faster than IFA roughly by a factor of 4.

4. Results on Real Data

4.1 Qualitative results

All of our experiments were conducted on an image sequence with eight-frames of a computer terminal (three of them shown below).

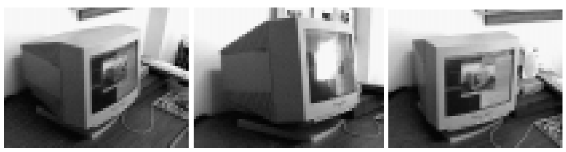


Figure 5. Examples of the terminal sequence

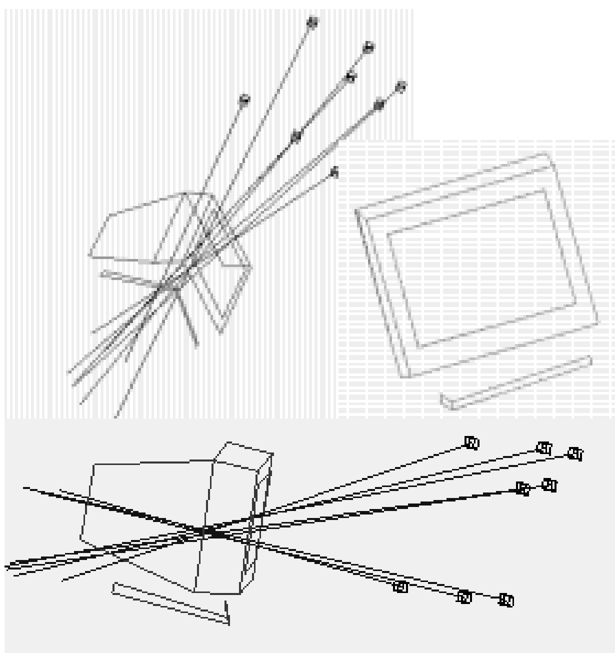


Figure 6. Reconstruction results of the terminal sequence

Corners were detected using the Harris corner detector. Among them, about 20 or so (not all appear in all views) were interactively selected and the correspondences established accordingly. The reconstructed points (in projective space) were then aligned with the Euclidean coordinates measured with a 3-D digitizer. Figure 6 shows three synthetic views of the resulting Euclidean reconstruction. The boxes indicate the locations of the camera projection centers. The lines indicate the principal axes of the recovered cameras.

4.2 Quantitative results

Table1 and Table2 compare quantitatively the results from different projective reconstruction methods. Four algorithms were implemented: the simple SVD method, the iterative factorization method (IFA), the proposed method (WIE) and the non-linear minimization method (MIN) using the Levenberg-Marquardt algorithm. The comparison was made on three types of measurements: the rank-4-ness of the final scaled measurement matrix (when it applied), the 3-D alignment error (in inch) and the 2-D reprojection error (in pixel). Results over four views are summarized in Table 1, while results over all eight views are summarized in Table 2. In both tables, the execution time (in seconds) was recorded from running the examples on a SGI O2 workstation.

Table 1. Comparison over 4 views

Items	**SVD**	**IFA**	**WIE**	**MIN**
σ_5/σ_4	0.14	0.026		
3D Error	0.3	0.119	0.117	0.116
2D Error	2.6	0.844	0.788	0.787
# iterations	1	80	30	20
Run time	0.014s	3.79s	4.103s	97.274s

Table 2. Comparison over 8 views

Items	**SVD**	**IFA**	**WIE**	**MIN**
σ_5/σ_4	0.138	0.028		
3D Error	0.288	0.114	0.097	0.098
2D Error	2.662	0.951	0.898	0.899
# iterations	1	40	30	20
Run time	0.036s	5.19s	8.516s	353.93s

Some remarks about the tables are:

- The rank-4-ness of W_s is much weaker for real images, as expected.
- The accuracy of WIE and MIN are indistinguishable, but the former is much faster. The speed difference increases as the number of views go up.
- Both WIE and MIN make marginal but necessary improvement over IFA.

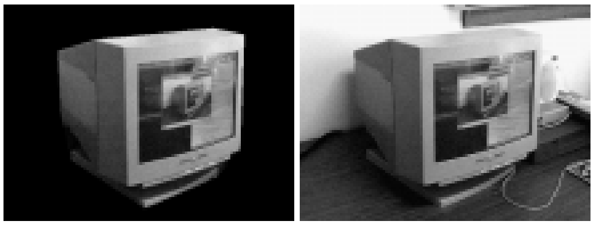
WIE provides the optimal trade-off in terms of accuracy and efficiency.

5. Applications

5.1 Image-Based Rendering

Recently, Image-Based Rendering [21] has emerged at the crossroad between computer graphics and computer vision. The goal is to create new images with photorealistic quality from existing images. The original idea of synthesizing new views without constructing the underlying 3-D geometry [1,14] is fully captured by the concept of projective reconstruction. Warping can be implemented as reprojection. The additional advantage of the projective reconstruction approach is that it is able to accept an arbitrary number of views arranged arbitrarily.

To support the claim, five views from the example of section 4 were chosen, projective reconstruction was performed and reprojected to a sixth view. To fill the pixels, the source images were tessellated into triangles, sorted, and warped into the synthesized view using projective texture mapping. The result is shown in Figure 7(a). The original view from the same viewpoint is shown in (b).

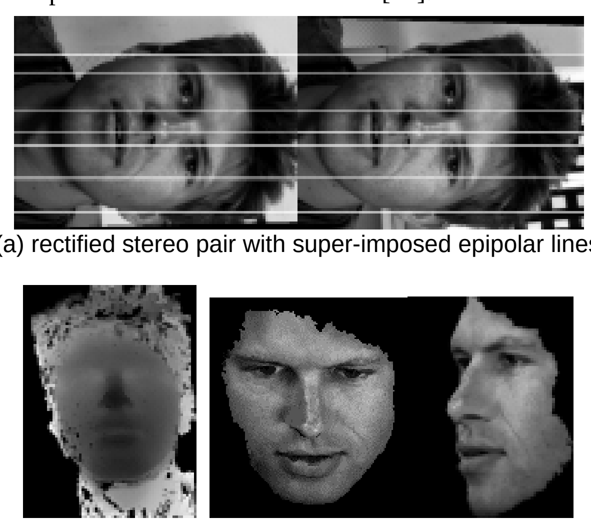


(a) synthesized view (b) original view

Figure 7. Demonstration of Image-Based Rendering

5.2 Shape-From-Stereo

Since we have a direct method of computing the projection matrices for two views, we in fact have an alternative way of computing the fundamental matrix [11]. Experiments confirm that our results are comparable to that of [22] on many examples. Once the fundamental matrix is known, we can rectify the image pair, compute the disparity map, and infer 3D shapes. The process is depicted in Figure 8. The data including the initial correspondences were obtained from [23].



(a) rectified stereo pair with super-imposed epipolar lines

(b) disparity (c) inferred face shape

Figure 8. Demonstration of Shape-From-Stereo

6. Summary

This paper presents an efficient iterative solution to the general M-view projective reconstruction problem by converting the non-linear bundle adjustment into a series of weighted least squares sub-problems. This solution addresses two existing issues, namely, data missing and rank deficiency. Experiments indicate that the proposed algorithm provides an excellent trade-off in terms of accuracy and efficiency. Additionally, it has been demonstrated that good initial values can always be obtained. With these results, we can conclude the following:

1) For two or three views, one can resort to computing the fundamental matrix or the trifocal tensor.
2) For coplanar views, one can first rectify them and then perform a single step of SVD.
3) If there are no missing data, one can use iterative factorization.
4) In the most general case, one should consider using the algorithm described in this paper.

References

1. R. Berthilsson, A. Heyden and G. Sparr, “Recursive Structure and Motion from Image Sequences Using Shape and Depth Spaces,” CVPR’97, 444-449, 1997
2. S. Eric Chen and Lance Williams, “View Interpolation for Image Synthesis”, SIGGRAPH’93, 279-288, 1993
3. G. Dahlquist and Åke Björck, Translated by N. Anderson, “Numerical methods”, Englewood Cliffs, N.J., Prentice-Hall, 1974
4. O. Faugeras, "What Can be Seen in Three Dimensions with an Uncalibrated Stereo Rig", ECCV’92, 563-578, 1992
5. O. D. Faugeras and B. Mourrain, “On the geometry and algebra of the point and line correspondences between N images”, ICCV’95, 951-956, 1995
6. R. Hartley, R. Gupta and T. Chang, "Stereo from uncalibrated cameras", CVPR’92, 761-764, 1992
7. R. Hartley, “Lines and Points in Three Views: A Unified Approach”, IUW’94, 1009-1016, 1994
8. R. Hartley, “A linear method for reconstruction from lines and points”, ICCV’95, 882-887, 1995
9. R. Hartley and P. Sturm, "Triangulation", CVIU, 68(2):146-157, 1997
10. D. Jacobs, "Linear Fitting with Missing Data: Applications to Structure-from-Motion and to Characterizing Intensity Images", CVPR’97, 206-212, 1997
11. Q. Luong and T. Viéville, "Canonic Representations for the Geometries of Multiple Projective Views", ECCV’94, 589-596, 1994
12. R. Mohr, F. Veillon and L. Quan, "Relative 3D Reconstruction Using Multiple Uncalibrated Images", CVPR’93, 543-548, 1993

13. C. Rothwell, G. Csurka, and O. Faugeras, "A Comparison of Projective Reconstruction Methods for Pairs of Views", ICCV'95, 932-937, 1995

14. S. M. Seitz and C. R. Dyer, "View Morphing", SIGGRAPH'96, 21-30, 1996

15. A. Shashua, "Trilinearity of Three Perspective Views and Its Associated Tensor", ICCV'95, 920-925, 1995

16. P. Sturm and B. Triggs, "A Factorization Based Algorithm for Multi-Image Projective Structure and Motion", ECCV'96, 709-720, 1996

17. R. Szeliski and P. Golland, "Stereo matching With Transparency and matting", ICCV'98, 517-524, 1998

18. B. Triggs, "Matching Constraints and the Joint Image", ICCV'95, 338-343, 1995

19. B. Triggs, "Factorization Methods for Projective Structure and Motion", CVPR'96, 845-851, 1996

20. S. M. Seitz and C. R. Dyer, "View Morphing", SIGGRAPH'96, 21-30, 1996

21. Website, Workshop on Image-Based Modeling and Rendering, http://www.graphics.standord.edu/workshops/ibr98

22. Z. Zhang, "Determine the Epipolar Geometry and Its Uncertainty: A Review", Research Report 2927, INRIA Sophia-Antipolis, 1996

23. Z. Zhang, O. Faugeras and R. Deriche, "An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment", Journal of Computer Vision Research, 1(1):58-68, 1997. On-line version at http://mitpress.mit.edu/e-journals/Videre

Appendix – Proof of the Theorem

Sufficient Condition.

Pre-multiply W_s by the diagonal matrix

$$D_1 = diag\left[\frac{1}{p_1}, \frac{1}{p_1}, \frac{1}{p_1}, \cdots, \frac{1}{p_m}, \frac{1}{p_m}, \frac{1}{p_m}\right],$$

and post-multiply it by

$$D_2 = diag\left[\frac{1}{x_1}, \frac{1}{x_2}, \cdots, \frac{1}{x_n}\right],$$

we have

$$W = D_1 W_s D_2 = \begin{bmatrix} \frac{P_1}{p_1} \\ \vdots \\ \frac{P_m}{p_m} \end{bmatrix} \begin{bmatrix} \frac{X_1}{x_1} & \cdots & \frac{X_n}{x_n} \end{bmatrix}.$$

Necessary Condition.

Let the Rank-4-Factorization of W be

$$W = \begin{bmatrix} P_1' \\ \vdots \\ P_m' \end{bmatrix} \begin{bmatrix} X_1' & \cdots & X_n' \end{bmatrix}.$$

Since X_j's and X_j''s are two projective reconstruction of the same set of points, they are equivalent up to a 4x4 non-singular collineation H. This means that there exist n non-zero factors: x_1, …, x_n, such that

$$[x_1 X_1' \quad \cdots \quad x_n X_n'] = H[X_1 \quad \cdots \quad X_n].$$

Therefore

$$W = \begin{bmatrix} P_1' \\ \vdots \\ P_m' \end{bmatrix} H H^{-1} [X_1' \quad \cdots \quad X_n']$$

$$= \begin{bmatrix} P_1' \\ \vdots \\ P_m' \end{bmatrix} H \begin{bmatrix} \frac{X_1}{x_1} & \cdots & \frac{X_n}{x_n} \end{bmatrix}$$

$$= \begin{bmatrix} P_1'' \\ \vdots \\ P_m'' \end{bmatrix} \begin{bmatrix} \frac{X_1}{x_1} & \cdots & \frac{X_n}{x_n} \end{bmatrix}.$$

Consider the two projection matrices of the i^{th} camera. The original definition says $\lambda_{ij}[u_{ij}, v_{ij}, 1]^T = P_i X_j$. Hence P_i and P_i'' both project X_1, ..., X_n to (u_{i1}, v_{i1}), ..., (u_{in}, v_{in}). As a result, they should be different only up to a scale factor. Denote the factor by p_i and write $P_i = p_i P_i''$. Knowing this is true for all $1 \le i \le m$, we have $\lambda_{ij}[u_{ij}, v_{ij}, 1]^T = p_i P_i'' X_j = p_i x_j [u_{ij}, v_{ij}, 1]^T$, for $1 \le i \le m$, $1 \le j \le n$, which implies $\lambda_{ij} = p_i x_j$, for $1 \le i \le m$, $1 \le j \le n$.

The above derivation also states that p_i is related to the i^{th} camera, and x_j is related to the j^{th} point. ♦