A linear algorithm for Camera Self-Calibration, Motion and Structure Recovery for Multi-Planar Scenes from Two Perspective Images

Gang Xu*, Jun-ichi Terai and Heung-Yeung Shum
Microsoft Research China
49 Zhichun Road, Haidian, Beijing 100080, China

Abstract

In this paper we show that given two homography matrices for two planes in space, there is a linear algorithm for the rotation and translation between the two cameras, the focal lengths of the two cameras and the plane equations in the space. Using the estimates as an initial guess, we can further optimize the solution by minimizing the difference between observations and reprojections. Experimental results are shown. We also provide a discussion about the relationship between this approach and the Kruppa equation.

1 Introduction

In man-made environments, planes are ubiquitous: buildings, floors, ceilings, streets, walls, furnitures, etc. If we can recover the planar scenes, it can be used in many applications like image-based modeling and rendering, robotics, etc. In this paper we try to solve the problem of using two perspective images of planes to auto-calibrate cameras, recover the motion between the cameras, and reconstruct the planes in space.

It is widely known that the projective transformation between two images of a plane can be described by a homography matrix. Many researchers have proposed various approaches to make use of the homography matrix. In his textbook on vision, Faugeras describes in detail how a homography matrix between normalized images (that is, assuming calibrated cameras) can be decomposed into rotation, translation and plane equation, without describing how to auto-calibrate cameras [Lon 86, Faugeras 93, Fau-Tos 86]. To "auto-calibrate" cameras, Triggs proposed to use an image sequence of a single plane [Triggs 98]. But he needs at least 5 images and has difficulty to initialize. Sturm and Maybank [Str-May 99] and Zhang [Zhang 99] independently proposed to use planar patterns in 3D space to precisely calibrate cameras. While Sturm and Maybank also discuss singularities, Zhang also calibrates radial distortions. They both require a planar pattern known *a priori*. Lowbowitz and Zisserman describe a technique of metric rectification for perspective images of planes using metric information such as a known angle, two equal but unknown angles, or a known length ratio. They also mentioned that calibration is possible without showing results or algorithms. Johansson describes how to synthesize new views given two images of planar scenes. His work does not try to reconstruct scenes in the Euclidean space [Joha 99].

In this paper, we propose a linear algorithm to solve the problem of self-calibrating cameras, and recovering camera motion and plane equations. The solution is generally unique. Using the estimates as an initial guess, we can further optimize the solution by minimizing the difference between observations and reprojections.

In the following, Section 2 presents the configuration, assumptions and homography. Section 3 presents a linear algorithm. Section 4 describes how to do bundle adjustment. Section 5 presents the experimental results, followed by a discussion and a conclusion.

2 Configuration, Assumptions and Homography

In this section, we briefly review how homography is introduced. As shown in Fig. 1, there are two cameras separated by a rotation R and a translation t. They look at a scene that has two or more planes.

Let us denote a point in the first camera's coordinate system $\mathsf{X} = [X, Y, Z]^\mathsf{T}$, and the same point in the second camera's coordinate system $\mathsf{X}' = [X', Y', Z']^\mathsf{T}$. They are related by

$$\mathsf{X} = \mathsf{R}\mathsf{X}' + \mathsf{t} \tag{1}$$

Planes are defined in the second camera's coordi-

*On leave from Ritsumeikan University

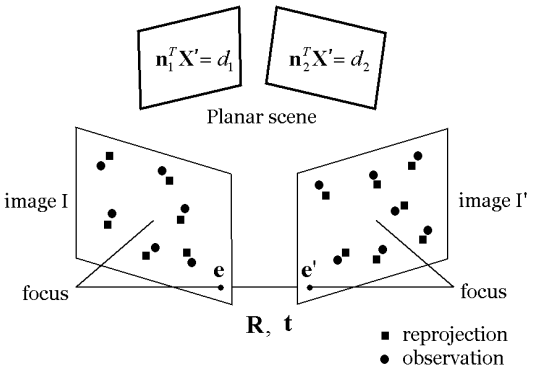

Figure 1: A planar scene with two or more planes

nate system by

$$\mathsf{n}^\mathsf{T}\mathsf{X}'/d = 1 \tag{2}$$

where n is the unit normal vector for the plane and d is the distance from the origin (the second camera's focal point) to the plane.

By multiplying (2) from the right to t in (1), we obtain

$$\mathsf{X} = (\mathsf{R} + \frac{\mathsf{t}\mathsf{n}^\mathsf{T}}{d})\mathsf{X}' \tag{3}$$

This equation means that given the rotation, translation between the cameras and the plane equation, a point can be transformed from one coordinate system to another by a 3×3 matrix.

The projection from the 3D coodinates to the image coordinates can be described by

$$\tilde{\mathsf{m}} \cong \mathsf{A}\mathsf{X}, \qquad \tilde{\mathsf{m}}' \cong \mathsf{A}'\mathsf{X}'. \tag{4}$$

where m and m' are the coodinates of the point in the first and second images, and A and A' are the intrinsic matrices of the first and second cameras, respectively. $\cong$ means equal up to a scale.

Substituting the above equation for (3), we have

$$\tilde{\mathsf{m}} \cong \mathsf{H}\tilde{\mathsf{m}}' \tag{5}$$

where

$$\mathsf{H} \cong \mathsf{A}(\mathsf{R} + \frac{\mathsf{t}\mathsf{n}^\mathsf{T}}{d})\mathsf{A}'^{-1} \tag{6}$$

is the homography matrix. Eq.(5) means that given a point in one image, the corresponding point can be determined in the other image, using the homography matrix given in (6).

Since one pair of points in (5) provides 2 constraints on the homography, given 4 or more feature points that are visible in both images, we can determine the homography matrix up to a scale.

In general, a camera has 5 intrinsic parameters: focal length, aspect ratio, skew angle and the 2 coordinates of the principal point [Faugeras 93]. If very high precision is required, one needs to also consider radial distortion [Zhang 99]. However, for certain purposes like rendering new images that does not require very high precision, one can assume that the cameras have zero skew and unit aspect ratio, and the principal point is at the image center. Zero skew and unit aspect ratio are two assumptions well satisfied by modern cameras. And it is also well known that reconstruction is not sensitive to the position of the principal point [Bougnoux 98]. The only unknown, then, is the focal length. Since focal length changes also with focusing, it is not practical to assume that the focal length is fixed. Therefore, with two images, we have two unknown focal lengths. The intrinsic matrices can then be written as

$$\mathsf{A} = \mathrm{diag}(f, f, 1), \qquad \mathsf{A}' = \mathrm{diag}(f', f', 1) \tag{7}$$

where f, f' are the focal lengths of the first and second cameras, respectively. Using (7) requires the origin of the image coordinate system be moved to the image center first.

3 A Linear Algorithm

Suppose that there are two planes, and we have already obtained two homography matrices H_1 and H_2. We can easily get the epipole e in the first image.

Given two planes, the line linking two optical centers intersect the two planes somewhere. Both intersections are projected onto the first image at the same point, the epipole e and onto the second image at the epipole e'. Given H_1 and H_2, we have a generalized eigen value equation

$$\gamma\mathsf{H}_1^{-1}\mathsf{e} = \mathsf{H}_2^{-1}\mathsf{e}(\cong \mathsf{e}') \tag{8}$$

where γ is an unknown scale. As Johansson shows, (8) is a generalized eigenvalue problem, and has two equal generalized eigenvalues and a third one. e can be obtained as the generalized eigenvector associated with the third eigenvalue [Joha 99]. The sign of e has to be determined later.

Since the scale of H determined from image points is arbitrary, we can rewrite (6) as

$$s\mathsf{H} = \mathsf{A}\mathsf{R}\mathsf{A}'^{-1} + \mathsf{A}\mathsf{t}\frac{\mathsf{n}^\mathsf{T}}{d}\mathsf{A}'^{-1} \tag{9}$$

where s is an unknown scale. Using $\mathsf{A}\mathsf{t} = a\mathsf{e}$ with a being a scale such that $\|\mathsf{e}\| = 1$, and doing some

algebraic manipulations, we obtain

$$s\mathsf{H}\mathsf{A}' - \mathsf{e}\bar{\mathsf{n}}^\mathsf{T} = \mathsf{A}\mathsf{R} \tag{10}$$

where $\bar{\mathsf{n}} = a\mathsf{n}/d = [n_1, n_2, n_3]^\mathsf{T}$. Multiplying the transpose of each side of (10) from the right yields

$$s^2\mathsf{H}\mathsf{A}'\mathsf{A}'^\mathsf{T}\mathsf{H}^\mathsf{T} + n\mathsf{e}\mathsf{e}^\mathsf{T} - s(\mathsf{H}\mathsf{A}'\bar{\mathsf{n}}\mathsf{e}^\mathsf{T} + \mathsf{e}\bar{\mathsf{n}}^\mathsf{T}\mathsf{A}'^\mathsf{T}\mathsf{H}^\mathsf{T}) = \mathsf{A}\mathsf{A}^\mathsf{T} \tag{11}$$

where $n = \bar{\mathsf{n}}^\mathsf{T}\bar{\mathsf{n}}$. The above equation is linear with respect to a 7-dimensional vector
$\mathsf{p} = [s^2 f'^2, s^2, f^2, n, sf'n_1, sf'n_2, sn_3]^\mathsf{T}$.

Since the matrices are symmetric, generally there are only six independent equations. We can rewrite it as

$$\mathsf{L}\mathsf{p} = \mathsf{q}, \tag{12}$$

where L and q are a 6×7 matrix and a 6×1 vector, respectively, uniquely determined from H and e.

We cannot directly obtain a unique solution from the above equation, but we can express the solution in the following form

$$\mathsf{p} = \mathsf{L}^+\mathsf{q} + \lambda\mathsf{g} \tag{13}$$

where $\mathsf{L}^+ = \mathsf{L}^\mathsf{T}(\mathsf{L}\mathsf{L}^\mathsf{T})^{-1}$ and g is the null vector of L which can be obtained as the right singular vector associated with the smallest singular value of L. λ is an unknown scalar to be determined.

We express the i-th component of p as $p_\mathsf{i} = f_\mathsf{i} + \lambda g_\mathsf{i}$. Since we know $\mathsf{n}^\mathsf{T}\mathsf{n} = 1$, we can build the following equation

$$p_5^2 + p_6^2 + \frac{p_1}{p_2}p_7^2 = p_1 p_4 \tag{14}$$

which is cubic with respect to λ. It can be solved in closed form. Experimental results show that the cubic equation usually degenerates to a quadratic equation (see Section 6) and there are usually two solutions. Which one is right can be determined by forcing $\det(\mathsf{R}) = 1$.

Once p is determined, we can express the estimated rotation matrix R' as

$$\mathsf{R}' = s\mathsf{A}^{-1}\mathsf{H}\mathsf{A}' - \mathsf{A}^{-1}\mathsf{e}\bar{\mathsf{n}}^\mathsf{T} \tag{15}$$

The above equation does not guarantee that the matrix is an orthogonal one, nor does it guarantee that its determinant is positive. Only one of two solutions of p gives R' a positive determinant. The other is abandoned. Next, the orthogonal matrix that is closest to R' can be obtained as $\mathsf{U}\mathsf{V}^\mathsf{T}$ where U and V are the left and right singular matrices of R', respectively [Kana 93]. It is the rotation matrix we look for.

There is still one more ambiguity in the sign of e. Since the second term in the righthand side of (6) includes the product of t and n^T, we need only reverse their signs simultaneously to keep the product unchanged.

The sign of e is chosen such that the computed 3D point is in front of both cameras. Given a pair of corresponding points x and x' in the two images, $z\tilde{\mathsf{x}}$ and $z'\tilde{\mathsf{x}}'$ are the 3D coordinates of the point in space where z, z' are unknown. They satisfy

$$z\tilde{\mathsf{x}} = \mathsf{R}z'\tilde{\mathsf{x}}' + \mathsf{t}\,. \tag{16}$$

The depths can be computed as

$$\begin{bmatrix} z \\ z' \end{bmatrix} = (\mathsf{B}^\mathsf{T}\mathsf{B})^{-1}\mathsf{B}^\mathsf{T}\mathsf{t} \tag{17}$$

where

$$\mathsf{B} = [\,\tilde{\mathsf{x}} \quad -\mathsf{R}\tilde{\mathsf{x}}'\,]$$

This computation is repeated for each pair of matched points. When the sign of t is reversed, the signs of the depths are also reversed. If the signs of the depths for the current e are positive, the current e is chosen. If the signs of the depths are negative, we choose $-\mathsf{e}$ and simultaneously $-\mathsf{n}$.

Up to now, we have obtained a unique solution of all the unknown parameters for each plane. The results for those parameters shared by all planes, such as focal lengths, rotation and translation, usually are not identical. we can choose any one set from them. These parameters, together with the normal and distance for each plane, are then used as initial guesses to minimize the difference between the observed image points and their reprojections so that the estimation is optimized in the least-square sense.

4 Bundle Adjustment

Suppose that there are M planes and N matched points in the two images. The difference between the observed image points and the reprojections of the computed 3D points is defined as

$$E = \sum_{\mathsf{j}=1}^{\mathsf{M}}\sum_{\mathsf{i}=1}^{\mathsf{N}} \omega(i,j)\{\|\mathsf{u}_\mathsf{i} - \mathsf{p}((\mathsf{R} + \mathsf{t}\mathsf{n}_\mathsf{j}^\mathsf{T}/d_\mathsf{j})\mathsf{X}'_\mathsf{i})\|^2 + \|\mathsf{u}'_\mathsf{i} - \mathsf{p}'(\mathsf{X}'_\mathsf{i})\|^2\} \tag{18}$$

where

$$\mathsf{p}([X,Y,Z]^\mathsf{T}) = [f\frac{X}{Z}, f\frac{Y}{Z}]^\mathsf{T},$$

$$\mathsf{p}'([X',Y',Z']^\mathsf{T}) = [f'\frac{X'}{Z'}, f'\frac{Y'}{Z'}]^\mathsf{T},$$

and $\omega(i,j) = 1$ when the i-th point belongs to the j-th plane; otherwise it is zero. Adding $\omega(i,j)$ facilitates representing points that belong to two or more planes.

Since the 3 coordinates of a point is not independent given the equation of the plane on which they must lie, we choose the X_i''s, Y_i''s, n_j's and d_j's as the independent structure parameters over which the cost function is to be minimized.

There are 3 more parameters for the rotation and 2 more parameters for the orientation of translation. Note that the scale of translation cannot be recovered, just as the scale of the space cannot be determined. So we set the length of the translation vector to be 1.

The minimization can be executed by the Levenberg-Marquartd algorithm [Press-et-al 88].

5 Experimental Results

We have done experiments for a number of multi-planar scenes. One of the example is shown here. Fig. 2 shows the two input images which are taken of a scene composed of a wall and a floor. The image size is 1280×960 pixels. The homography matrices are determined from point matches using the algorithm proposed by Hartley [Hartley 97]. The points are currently manually matched.

The final focal lengths of the two images are 1246 and 1166 pixels, respectively. The angle between these two planes is estimated to be about 86.3°.

The reconstructed scene is modelled by VRML with texture maps from one of the original images. New images are generated of the scene for new view points. One of them is shown in Fig. 3. The geometry agrees with our perception of scene. The reason for the deformation of the texture is because the texture mapping is done independently for each triangle, so the global consistency is not guaranteed. Also note that the sofa is not treated as a separate object, but as parts of the two planes.

Fig. 4 shows how large the angle between the two planes looks. It looks just like a right angle, though we do not have this knowledge in our program.

Figure 2: Two images of a scene with two planes

6 Relationship to the Kruppa Equation

The experimental results show that in most cases, the first 3 components of g are very close to zero. This means that the two focal lengths can be determined regardless of the unknown λ. In other words, the two cameras can be calibrated separately before computing the structure and motion.

Actually, we find that this is equivalent to the Kruppa equation [Luong-Fau 97, May-Fau 92, Xu-Sugi 99]. Starting from (10), we can derive the Kruppa equation. Multiplying $[\mathsf{e}]_\times$ from the left to both sides of (10), we obtain

$$s[\mathsf{e}]_\times \mathsf{H}\mathsf{A}' = [\mathsf{e}]_\times \mathsf{A}\mathsf{R} \tag{19}$$

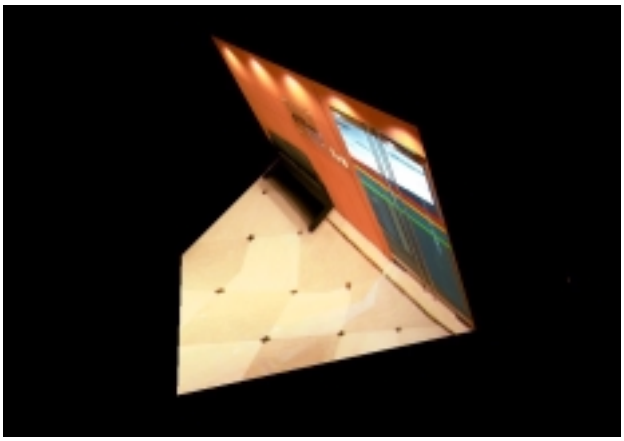

Figure 3: Images are generated for new viewpoints from the reconstructed planes with the original image as texture maps.

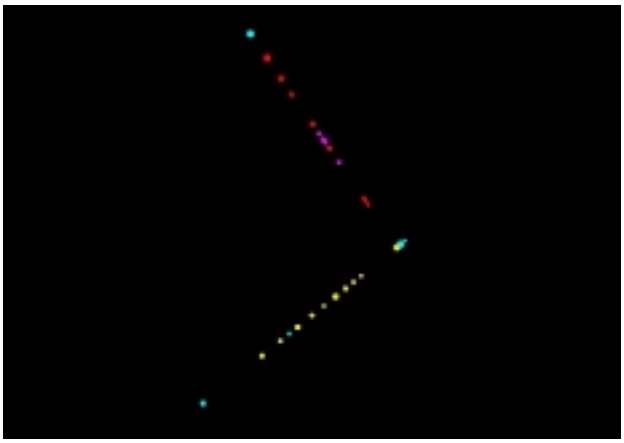

Figure 4: The two reconstructed planes are perpendicular to each other. The small balls show the positions of the feature points.

Again, multiplying the transpose of each side from the right yields

$$s^2[\mathsf{e}]_\times \mathsf{H}\mathsf{A}'\mathsf{A}'^{\mathsf{T}}\mathsf{H}^{\mathsf{T}}[\mathsf{e}]_\times^{\mathsf{T}} = [\mathsf{e}]_\times \mathsf{A}\mathsf{A}^{\mathsf{T}}[\mathsf{e}]_\times^{\mathsf{T}} \qquad (20)$$

This is exactly the Kruppa equation, because $\mathsf{F} = [\mathsf{e}]_\times \mathsf{H}$.

It is widely known and our own experience also shows, that camera calibration using the Kruppa equation is very sensitive to noise [Bougnoux 98]. Because the epipolar geometry provides only two constraints on the intrinsic parameters, we can solve for the two focal lengths if all other intrinsic parameters are already known [Hartley 92, Brooks et al 96, Bougnoux 98]. In our experience, it is often the case that one or both of the focal lengths do not have a real solution in the Kruppa equation, because we have to solve a second order equation in terms of f^2 and f'^2 and they often have negative solutions.

Compared with the solution using Kruppa's equation, our result using this new algorithm is more reliable. The reason is probably that the epipole and the fundamental matrix (though not explicitly expressed in our algorithm) are more reliable in our new algorithm. Recall that we do not compute the fundamental matrix directly from the image points [Xu-Zhang 96], but rather, we compute the homography matrices from image points first, and the epipole and implicit fundamental matrices are determined from the homography matrices. Global constraints of coplanarity are incorporated in this process so that the computation is more robust to noise.

It should be noted that since our algorithm is essentially equivalent to solving the Kruppa equations, it inherits the degeneracies of the Kruppa equations [Brooks et al 96]. One such degeneracy is the case of coplanar optical axes.

7 Conclusion

In this paper we have described a new algorithm to determine camera focal lengths, camera motion and planes in space given two perspective images of multiplanar scenes. The algorithm is based on a linear algorithm solution of the parameters. The solution can be used as an initial guess to minimize the difference between observations and reprojections. Experimental results show that the new algorithm is reliable and robust to noise. We have also given a discussion about the relationship between our new algorithm and those using the Kruppa equation.

We are currently investigating how to match and group/segment points by finding which points satisfy the same homography matrix. The final result will be

an automatic system that can detect and reconstruct planes.

References

[Bougnoux 98] Bougnoux, S., "From projective to Euclidean space under any practical situation, a critism of self-calibration", Proc. of 6th Int. Conference on Computer Vision, pp.790–796, Bombay, India, Jan. 1998, Narosa Publishing House.

[Brooks et al 96] Brooks, M.J., de Agapito, L., Huynh, D.Q. and Baumela, L., "Direct methods for self-calibration of a moving stereo head", Proc of ECCV, Cambridge, UK., 1996

[Faugeras 93] Faugeras, O.D., *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993

[Fau-Tos 86] Faugeras, O.D. and Toscani, G., "The calibration problem for stereo", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.15-20, Miami Beach, FL, June 1986.

[Hartley 92] Hartley, R., "Estimation of relative camera positions for uncalibrated cameras", Proc. of 2nd European Conference on Computer Vision, pp.579-588, Santa Margherita Ligure, Italy, May 1992, Springer Verlag.

[Hartley 97] Hartley, R., "In defense of the eight-point algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.6, pp.580-593,1997.

[Joha 99] Johansson, B., "View synthesis and 3D reconstruction of piecewise planar scences using intersection lines between the planes", Proc. 7th Int. Conference on Computer Vision, pp.54-59, Corfu, Greece, Sept., 1999.

[Kana 93] Kanatani, K., *Geometric Computation for Machine Vision*, Oxford Science Publications, 1993.

[Lie-Zis 98] Liebowitz, D. and Zisserman, A., "Metric rectification for perspective images of planes", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.482-488, Santa Barbara, CA, June 1998.

[Lon 86] Longuet-Higgins, H.C., "The reconstruction of a plane surface from two perspective projections", Proc. R. Soc. London, Series B, Vol.227, pp.399-410, 1986.

[Luong-Fau 97] Luong, Q.-T. and Faugeras, O.D., "self-calibration of a moving camera from point correspondences and fundamental matrices", Int. Journal of Computer Vision, Vol.22, No.3, pp.261-289, 1997.

[May-Fau 92] Maybank, S. and Faugeras, O.D., "A theory of self-calibration of a moving camera", Int. Journal of Computer Vision, Vol.8, No.2, pp.123-152, 1997.

[Press-et-al 88] Press, W.H. and Flannery, B.P. and Teukolsky, S.A. and Vetterling, W.T., *Numerical Recipes in C*, Cambridge University Pres, 1988.

[Str-May 99] Sturm, P. and Maybank, S., "On plane-based camera calibration: A general algorithm, singularities, applications", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.432–437, Fort Collins, Colorado, June 1999.

[Triggs 98] Triggs, B., "Autocalibration from planar scenes", Proc of 5th European Conference on Computer Vision, pp.89-105, Freiburg, Germany, 1998.

[Xu-Sugi 99] Xu, G. and Sugimoto, N., "An algebraic derivation of the Kruppa equation and a new algorithm for self-calibration of cameras", Journal of American Society of Optics A, Vol.16, No.10, pp.2219-2424, 1999.

[Xu-Zhang 96] Xu, G. and Zhang, Z., *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, Kluwer Academic Publishers, 1996

[Zhang 99] Zhang, Z., "Flexible camera calibration by viewing a plane from unknown orientations", Proc. 7th Int. Conference on Computer Vision, pp.666–673, 1999.