Affine Stereo Calibration for Relative Affine Shape Reconstruction

Long QUAN

LIFIA — INRIA

46, avenue Felix Viallet,

38031 Grenoble, France

Abstract

It has been shown that relative projective shape, determined up to an unknown projective transformation, with respect to 5 reference points can be obtained from point-to-point correspondences of a pair of images; Affine shape up to an unknown affine transformation with respect to 4 points can be obtained from parallel projection. We show in this paper that affine shape with respect to 4 reference points can be obtained from two perspective images provided that the pair of images is affinely calibrated. By affine calibration, it means the establishment of a special plane collineation between two image planes, this collineation is the product of two plane collineations each of which establishes a $(1, 1)$ correspondence between an image plane and the plane at infinity. Experimental results are also presented.

1 Introduction

Recent works [1, 4, 9] show that it is possible to get invariant projective shape representation from a pair of non calibrated images, with the assumption that a sufficient number of points are previously matched between images. These works are originated from the pioneer work of Koenderink and Van Doorn [5] on affine shape representation from restricted camera projection, that is parallel projections and other related works [13, 14, 8].

Affine shape reconstruction has been extensively studied for parallel projections or so-called weak perspective projections [11, 13], there have been many attemps to get affine shape from the full perspective images using point matches without calibration, it is now clear that it is mathematically impossible. Especially in [11], it has been shown that additional special reference points are needed to get affine shape from perspective images. Sparr (cf. [13, 14]) reconstructs the affine shape using available affine information of objects such as the rectangular patches. Faugeras [1] dealt with the family of affine shapes. In this paper we will argue and show that the affine shape is obtainable from a pair of images using point-matches only provided that an *affine calibration* is furnished. Later in Section 2, we will show that by affine calibration, we mean the establishment of a special plane collineation between two image planes. Intuitively, in this affine calibration step, naturally pure projective incidence property of point-to-point correspondence is not enough, a kind of affine knowledge should be introduced, for example, parallelism of lines. However Euclidean knowledge such as the exact coordinates necessary for classical stereo calibration is no more necessary. So the basic idea developed in this paper can be regarded as exploring the capabilities

of partially calibrated stereo system, as without any calibration (however with on-line point-to-point correspondences) of a stereo system, only invariant projective shape, that is the shape defined up to an unknown collineation, is obtainable. In particular, we will show what is necessary for getting the affine representation of the shape, since affine shape often presents a good compromise bwteen projective shape which is metrically too poor and Euclidean shape which is metrically rich enough but difficult to obtain.

2   Affine calibration

We first show how different calibrations are related to different 3D reconstruction, which allows us to introduce what it means for us by affine calibration.

Projective shape reconstruction is possible from a pair of images provided that the epipolar geometry is established. This epipolar geometry is entirely determined by *fundamental matrix* [2, 7] which can be considered as the projective version of the familiar *essential matrix* [6] for motion analysis. Recall that fundamental matrix is projectively a correlation between two image planes. A plane correlation is a linear transformation which transforms points into lines and lines into points, that is what the epipolar geometry does. The correlation is represented by a matrix of 3×3 whose rank is of 2. Therefore the degrees of freedom are of $7 = 9 - 1 - 1$. Since the determination of fundamental matrix needs no more than the projective incidence properties, the determination of fundamental matrix can be considered as projective calibration, it is also called *weak calibration* in [12]. Naturally, from projective calibration, only relative projective shape which is defined up to a projective transformation in space (a matrix of 4×4) can be obtained [1, 4, 9, 3]. Needless to say that there is no hope to recover any other more metric shape representation, since we have no this kind of metric information.

Naturally the classical calibration process needed to stereo vision can be considered as Euclidean calibration, since explicit Euclidean metric is required during calibration step. Euclidean calibration of a stereo pair has $22 = 2 \times (3 \times 4 - 1)$ degrees of freedom.

By affine calibration is meant that besides of projective calibration, some affine information should somehow introduced. Therefore this so called affine calibration for a stereo should turn out the corresponding affine shape representation. That is, the shape is defined up to an affine transformation in space. And an affine transformation is a linear transformation which leaves the plane at infinity invariant. Affine transformations constitute a subgroup of the projective group, the general linear group. Obviously, we should have something to do with the plane at infinity, more precisely, the plane at infinity should be somehow observable in image planes. The points on the plane at infinity represent the directions of family of parallel lines in the affine space. As we known that these points at infinity are perspectively projected on to image plane as normal points known as vanishing points associated to parallel lines in space. Three points at infinity define the plane at infinity, therefore intuitively three vanishing points should be enough.

More concretely, our affine calibration needs the knowledge of the parallelism, however we do not need any further Euclidean information such as Euclidean coordinates of points or the distance between two parallel lines, only the pure affine information, parallelism is taken into account. The detection of vanishing points can be implemented as that described in [10]. Thus three vanishing points should be detected in the affine calibration step. This coincides with what it is pointed out in [1] that with the epipolar geometry, affine shape reconstruction has still

$3 = 22 - 12 - 7$ degrees of freedom, since each vanishing point fixes one remaining degree of freedom, three fix entirely the unique affine shape representation.

Some algebraic consideration can be studied as follows. Each camera's projection matrix is of form $(P_i \ p_i)$, where P_i is its non singular 3×3 submatrix for perspective projection, p_i is a 3×1 vector. The point at infinity of each direction in space d_∞ (a 3D vector) has homogenious coordinates $(d_\infty \ 0)^T$, its projection,

vanishing point, in the two images $v_i = (P_i \ p_i) \begin{pmatrix} d_\infty \\ 0 \end{pmatrix} = P_i d_\infty$,

therefore P_i can be considered as a plane collineation between image plane and the plane at infinity. Then $v_2 = P_2 P_1^{-1} v_1 \overset{def}{=} A v_1$, which is still a plane collineation between two image planes. Three vanishing points are not yet enough to determine A. However, as the epipolar geometry of the pair is known, i.e. th eipoles are known and it is important to note that obviously $e_2 = A e_1$. That totalizes 4 points to determine a unique A up to a scaling factor.

It leads to that *Affine calibration is, in addition to the determination of fundamental matrix, the establishment of a special plane collineation between two image planes, this collineation is the product of two plane collineations each of which establishes a $(1,1)$ correspondence between an image plane and the plane at infinity. Algebraically it is equivalent to have determined globally $P_2 P_1^{-1}$, P_i is the 3×3 non sigular submatrix of projection matrix.*

$P_2 P_1^{-1}$ can be determined with at least 3 vanishing points correspondence provided the fundamental matrix.

3 Affine shape reconstruction

Relative affine shape representation

From any four distinct points, say O, X, Y and Z which are neither coplanar nor three of them collinear, we can construct a unique affine frame and assign the coordinates representations $(0,0,0)^T, (1,0,1)^T, (0,1,1)^T$ and $(1,1,1)^T$ to these reference points, then any fifth, say P and all other points can be assigned the unique affine coordinates $(\alpha, \beta, \gamma)^T$, these affine coordinates constitute the affine shape representation of these points with respect to the first 4 reference points, Obviously, the affine shape is only defined up to an unknown affine transformation.

The geometric way to define (α, β, γ) is as follows. Any point P is projected onto the planes OXY, OYZ and OXZ along the directions of OZ, OX and OY axes (see Figure 1). The projection points are respectively denoted by P_{xy}, P_{yz} and P_{xz}. Then in each plane, say the plane OXY, project P_{xy} onto the lines OX and OY along the direction of OY and OX. These projection points are denoted by P_x, P_y. In the same way, we can get P_z. Thus (α, β, γ) are defined by position ratios

$$\alpha = \frac{OP_x}{OX}, \quad \beta = \frac{OP_y}{OY}, \quad \text{and} \quad \gamma = \frac{OP_z}{OZ}$$

However, in our context, these spatial measures are not available. We have access only to perspective image measures. Thus the shape should be reconstructed by means of invariants. As we are considering the perspective projections, the basic
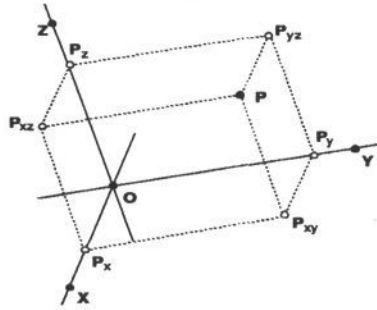
Figure 1: Affine coordinates in space.

invariant is the cross ratio[1]. However the basic affine invariant is the position ratio. In order to get affine representation from projective projections, the following property of cross ratio establishes the transition between cross ratios and position ratios via the point at infinity.

If one of the points of the projective line is perspectively mapped to the point at infinity on an affine line, the projective coordinate defined by cross ratio equals to the affine coordinate defined by position ratio, that is,

$$\alpha = \{P, X; O, \infty\} = \frac{OP}{OX}.$$

In the following, uppercase letters denote the points in space, image points are denoted by lowercase letters subscripted by the number corresponding to the image number. For instance, a point in space P is projected onto the first image plane as p_1 and onto the second one as p_2.

Viewing line reference plane intersection

Given 4 points, take 3 of them as a reference plane, examine the relative position of the fourth corresponding point with respect to this reference plane. It has been firstly proved in [9] that a necessary and sufficient condition for 4 points to be coplanar in space can be established from the epipolar geometry. We present a variant of this condition in a more constructive way, that is we determine in one image the intersection point of the viewing line of another image with the reference plane. Algebraically, this operation can be simply determined like follows. With the similar reasoning as in affine calibration step, a plane collineation between two image planes can be specified with 3 points which define the reference plane provided the epipolar geometry. This collineation is the product of collineations between image plane and the unknown reference plane in space. If B denotes this plane collineation, given a point p_1 in the first image, the intersection point of the viewing line of p_1 with the reference plane is located simply at $p'_2 = Bp_1$ in the second image. It is evident that if p'_2 is superimposed with p_2, then it means that the fourth point is coplanar with the reference plane.

---

[1] The cross ratio of the 4 numbers is defined as

$$\{a, b; c, d\} = \frac{(a - c)/(a - d)}{(b - c)/(b - d)}$$

Line reference plane intersection

Another essential operation for geometric reconstruction is to be able to realize a general line and reference plane intersection. It has been proposed in [11], called the piercing point, in which it is limited to the reference plane, referenced by 4 coplanar points. Thanks to the epipolar geometry, the same operation goes for any reference plane referenced by any three points, since the epipoles provide always the necessary fourth point. Algebraically, $l_2' = (B^{-1})^T l_1$, where $(B^{-1})^T$ is the dual transformation of B, then the intersection point is $p = l_2 \times l_2'$.

Intersection with the plane at infinity

First, we should determine the projection of the points at infinity along three affine axes specified by reference points. Obviously these points are just the intersection points of the plane at infinity with the affine frame axis. These intersection points can be easily located by using the previous line plane intersection operation. If A still denotes the plane collineation of affine calibration,

$$x_{\infty 2} = (o_2 \times x_2) \times ((A^{-1})^T (o_1 \times x_1))$$

Similar expressions hold for $y_{\infty 2}$ and $z_{\infty 2}$. Inverse the image number, we can get $x_{\infty 1}$, $y_{\infty 1}$ and $z_{\infty 1}$.

Reconstruction on the reference planes

Then, given a pair of corresponding image points p_1 and p_2, to parallelly project a point along a reference axis in space is equivalent to drawing a line going through the corresponding vanishing point of the axis in the image. So the projection of a given point along the affine axis can be realized in the image plane. For example for the point P_{xy}, we obtain

$$p_{xy2} = (p_2 \times x_{\infty 2}) \times ((A^{-1})^T (p_1 \times x_{\infty 1}))$$

Therefore all these reconstructed points in the second image plane are the perspective view of the points on the reference plane OXY, i.e. a non singular plane collineation exists between the reference plane OXY and the image plane used for reconstruction. In fact, we are considering the subordinate projective geometry of dimension 2 between the reference plane OXY and the image plane. The similar operations hold for other reference planes such as OXZ and OYZ.

Affine rectification from projective plane

Now considering the subordinate plane geometry on the reference plane, this projectively deformed planar shape can be rectified into its real affine shape by a planar collineation defined by O, X, Y and E. E is the intersection point of XY_∞ and YX_∞. That is to find the 3×3 plane collineation A_{xy} which transforms o_2, x_2, y_2 and $e_2 = (x_2 \times y_{\infty 2}) \times (y_2 \times x_{\infty 2})$ into the unique canonical affine coordinates representation

$$A_{xy} : o_2 \rightarrow (0,0,0)^T, x_2 \rightarrow (1,0,1)^T, y_2 \rightarrow (0,1,1)^T, e_2 \rightarrow (1,1,1)^T$$

Thus, if $(x_1, x_2, x_3)^T = A_{xy} p_{xy}$, then the affine subcoordinates $(\alpha, \beta)^T = (\frac{x_1}{x_3}, \frac{x_2}{x_3})^T$. While considering another subordinate plane geometry on the reference plane OXZ, we get $(\alpha, \gamma)^T$. These lead to $(\alpha, \beta, \gamma)^T$.

4   Experimental results

We have firstly experimented on the simulated glass data to validate the method. The fundamental matrix between the pair of images is determined based on a nonlinear optimization algorithm. Then, a simulated cube is put in the scene to perform affine calibration.
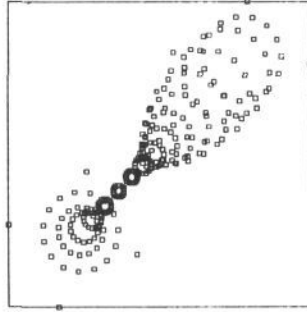
Figure 2 shows one of the original simulated image.

Figure 2: One of the original simulated image

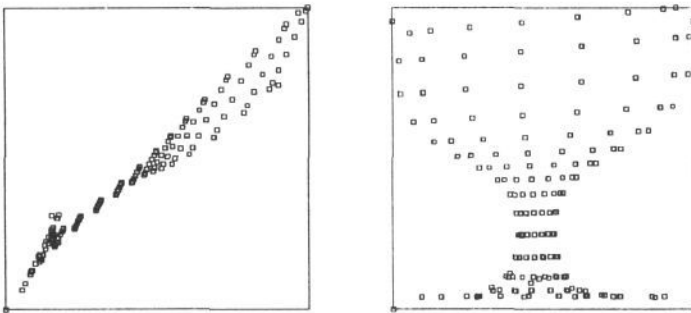Different reconstruction steps of a subordinate shape are illustrated by Figure 3.

Figure 3: The projective and affine subcordinate shape on the reference plane OXZ.

Figure 4 shows the different views of the 3D reconstructed affine shape.

The real image data set is obtained from a regular pattern. Thanks to the regularity of the pattern, it makes possible the matching of the points. Since only one planar pattern is available, we create a kind of *"transparent"* spatial pattern, that is, once the camera is fixed in a position, the pattern plane is then translated. This is equivalent to have several transparent regular pattern. In this experimentation, we used 3 transparent planes and 2 positions to simulate a stereo pair. The affine calibration is based on the location of the vanishing points of the bounding box of the spatial pattern. Figure 5 shows one of the original images.
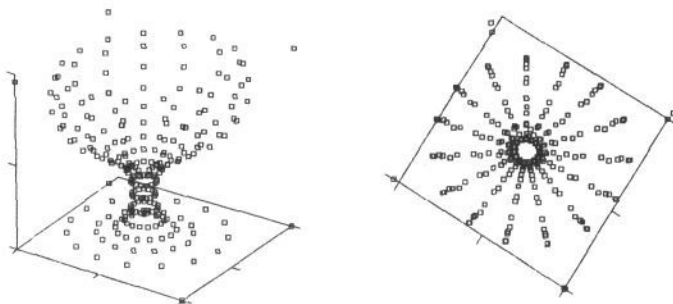
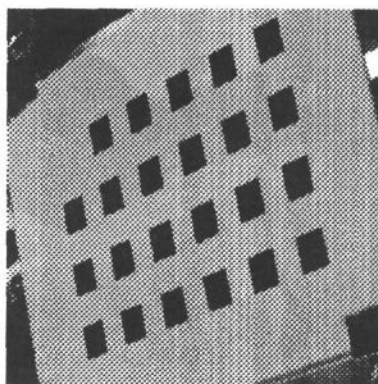Figure 4: Two views of the 3D affine reconstructed shape.

Figure 5: One of the pattern images.

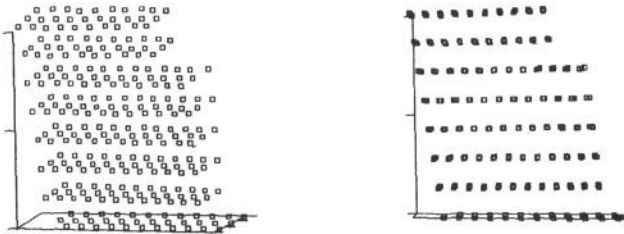Figure 6 shows two views of the reconstructed 3 planes pattern.

Figure 6: Two views of the 3D affine reconstructed 3 planes pattern.

While selecting different reference points, the different affine reconstructions are obtained and are superimposed in Figure 7, which illustrates that the affine shape reconstruction is defined up to an unknown affine transformation.
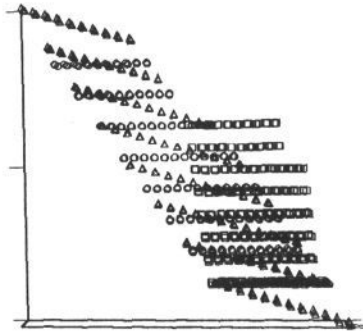

Figure 7: Three superimposed affine shape with different reference points.

5   Conclusion

There have been many attempts to obtain affine shape representation from a pair of projective images, it is clear that it is not possible without an affine calibration

step. So this paper just provides what is needed to get the real affine shape. The results presented in this paper is not contradictory with the affine reconstruction from the epipolar geometry described in [1], since the affine shape obtained in [1] is not uniquely determined, it is a three-parameter family affine reconstruction, that is up to 3 independent parameters apart from an unknown affine transformation. The introduction of affine calibration step makes the solutions down to be a unique one. Therefore, we clearly indicate one of the possible ways to fix the three independent parameters which define the family of affine solutions.

The method presented here covers also the affine reconstruction from parallel projections, in this case the assumption of parallel projection provides what is needed for affine calibration.

Another practical point is that the relative shape reconstruction is directly located in the visible range of image plane, makes the solution more stable in some sens more relative than directly going through the canonical basis and singular cases are even more easily detected.

Acknowledgements

References

[1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[2] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera Self-Calibration: Theory and Experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.

[3] P. Gros and L. Quan. 3D projective invariants from two images. In *Geometric Methods in Computer Vision II, SPIE's 1993 International Symposium on Optical Instrumentation and Applied Science*, page to appear, July 1993.

[4] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[5] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

[6] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. In *Nature*, volume 293, pages 133–135. XX, September 1981.

[7] Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur*. Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.

[8] R. Mohr, L. Morin, C. Inglebert, and L. Quan. Geometric solutions to some 3D vision problems. In R. Storer J.L. Crowley, E. Granum, editor, *Integration and Control in Real Time Active Vision*, ESPRIT BRA Series. Springer-Verlag, 1991.

[9] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiples uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, LIFIA–IRIMAG, 1992.

[10] L. Quan and R. Mohr. Determining perspective structures using hierarchial Hough transform. *Pattern Recognition Letters*, 9(4):279–286, 1989.

[11] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992. Also in IEEE Workshop on Visual Motion, New Jersey, pages 249–254, 1991.

[12] L. Robert and O. Faugeras. Relative 3D positionning and 3D convex hull computation from a weakly calibrated stereo pair. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, 1993.

[13] G. Sparr. An algebraic/analytic method for reconstruction from image correspondance. In *Proceedings of the 7th Scandinavian Conference on Image Analysis, Aalborg, Denmark*, pages 274–281, 1991.

[14] G. Sparr. Depth computations from polyhedral images. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 378–386. Springer-Verlag, May 1992.