How useful is projective geometry?

P. Gros[1] R. Mohr[1] L. Quan[1] R. Hartley[2]
[1]: LIFIA - INRIA Rhône-Alpes
46, avenue F. Viallet 38031 Grenoble Cedex 1 - France
[2]: G.E. CRD, Schenectady, NY, 12301, USA

In this response, we will weigh two different approaches to shape recognition. On the one hand we have the use of restricted camera models as advocated in the paper of Pizlo et. al. to give a closer approximation to real calibrated cameras. The alternative approach is to use a full projective camera model and take advantage of the machinery of projective geometry.

1 About the perspective projection

Before discussing the usefulness of projective geometry, we revisit perspective projection.

A pin hole camera can be modeled as a linear mapping in homogeneous coordinates from the 3D space onto a plane. Usually this mapping is represented by the product of a rigid motion in space with matrix D, followed by a standard perspective projection expressed as a 3×4 matrix P_0, and finally a rescaling in the image due to the camera parameters represented by a 3×3 matrix K. So the final projection is represented by a 3×4 matrix P.

$$\begin{aligned} P &= KP_0D && (1) \\ &= \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} & R & & t \\ 0 & 0 & 0 & 1 \end{pmatrix} && (2) \end{aligned}$$

Such a transformation has 5 d.o.f. (degrees of freedom) for K plus 6 d.o.f. for D (3 for the rotation R and 3 for the translation t). This makes a total of 11.

1.1 How many degrees of freedom has a plane to plane projection?

Let us first consider the number of degrees of freedom of a projection between two planes. This question is also discussed in the discussion paper of Pizlo et. al.

Without loss of generality we can assume that the points are selected from the $\vec{x}, \vec{y}$ plane. The camera mapping induces a plane-to-plane projective mapping and therefore can at most have 8 degrees of freedom, so all the 11 d.o.f. are not independent.

The 3×3 projection matrix between two planes can be related to the physical parameters of the imaging system and the coordinate system of the object plane as follows:

$$A_{3\times3} = KT_{\sigma\tau}SD'$$

$$= \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\sigma\cos\tau & -\sin\tau & 0 \\ \cos\sigma\sin\tau & \cos\tau & 0 \\ \sin\sigma & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & C \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha & t_x \\ \sin\alpha & \cos\alpha & t_y \\ 0 & 0 & 1 \end{pmatrix}$$

In this relation, K is as before the matrix of intrinsic parameters, accounting for 5 d.o.f. [7], D' is a Euclidean displacement in the object plane (3 d.o.f.), S is a scaling matrix (1 d.o.f.) and $T_{\sigma\tau}$ represents the orientation of the object plane with respect to the camera (2 d.o.f). Note that S can be absorbed either by $T_{\sigma\tau}$ or by D. When it is absorbed by $T_{\sigma\tau}$, it can be interpreted as the third parameter of the object plane. Otherwise it can be interpreted as a global scaling factor in the object plane.

1.1.1 Calibrated camera

If the camera is calibrated, the calibration matrix K is known, and so there remain only $11-5=6$ d.o.f.

Are they independent? There exists an elegant proof in this calibrated case. Consider the image of a circle in space. It is well known that a circle can be projected as a general conic (see Apollonios de Perga 200BC, or more recent work on determining the orientation of a plane containing a circle [5]). This makes already five d.o.f., since a conic is specified by five parameters.

Now, if a point is distinguished on this conic, by rotating the circle on itself we can bring the projection of this point anywhere on the conic. This additional degree of freedom is the missing one which now leads us to a total number of 6 degrees of freedom.

A more systematic algebraic proof can be provided as follows. Consider three points in a plane, and, after rotating and translating this configuration in space (through D), project them onto the image. The coordinates of these three points define a vector of dimension 6. It only remains to check if the manifold spanned by the parameters of the rigid motion has dimension 6. Computing the determinant of the Jacobian shows easily that, except for singular points, the Jacobian has full rank and therefore the manifold is of dimension 6. This computation can be verified using Maple or Mathematica.

Therefore a perspective projection of a plane onto the image plane of a calibrated camera has 6 degrees of freedom.

1.1.2 Uncalibrated camera

If now the camera is no longer assumed to be calibrated, obviously the number of d.o.f. cannot be higher than 8. So if we consider that two of the intrinsic parameters might change (for instance the focal length and the aspect ratio, or the position of the camera principal point), then, using the manifold technique, one can easily check that the number of d.o.f. is 8 in each of these cases.

It might be noticed however that real parameters are really limited in range: this reduces the possible parameter space to a small region of the whole space, but it should noticed that this region is still of dimension 8. This point will be discussed later on.

1.2 The perspective projection from the full 3D space

It is also interesting to compare the number of d.o.f with that of the general projection of 3D (non planar) points. The general model has 11 d.o.f. as the general projective mapping from the projective space $\mathcal{P}^{\ni}$ on the projective plane $\mathcal{P}^{\in}$.

However it may be noted that, for standard cameras, the parameter s in K has value 0; thus only 10 d.o.f. remain. So, for the uncalibrated case, the number of d.o.f. for a perspective projection is lower than that of the full projective projection.

Now considering the case of a calibrated camera, the number of d.o.f. comes to 6 as in the case for a plane to plane projection, notably fewer than the 11 d.o.f. of the full projective projection.

1.3 What can be extracted from these numbers?

First, notice that in the general case of uncalibrated cameras, the projective model has a number of d.o.f. equal or close to the number of d.o.f. of the perspective projection. Consequently, the projective model is a very good approximation to the calibrated camera model. This fact, added to simplicity and to the possibilities described in the next section, means that the projective model is a very useful model for analysis of the imaging process.

However, it is true that the manifold of shapes obtained via projective projection is larger than the one obtained by perspective projection. This is mainly due to the fact that camera parameters are limited in range (for instance the ratio α_u/α_v ranges usually between 0.7 to 1.4, s is close to 0). Similarly the positions of shape in space are also constrained: for instance, they do not cross the viewer eye and extreme configurations are not probable [1].

Such considerations have led to the introduction of quasi-invariants. These quasi-invariants may just be considered as approximate values of invariants. In practice, they appear to be very useful for qualitative tasks like identification [3, 9]. On the other hand, if accuracy is needed, then an exact model is necessary and exact invariants (i.e. the one provided by the projective group) are to be used [18, 20].

1.4 Comments on the approach of Pizlo et al.

In their paper, Pizlo et al. work with calibrated cameras, thus K is known. The *shape* defined in [23] is implicitly constituted of similarity invariants, modulo a scaled planar Euclidean transformation. (The concept of shape is clear in [23] but is quite confusing in the discussion paper). When similarity invariants are concerned, SD is absorbed and only $T_{\sigma\tau}$ remains. The determination of $T_{\sigma\tau}$ is equivalent to that of the vanishing line of the object plane. If σ and τ were known, we could compute the exact similarity invariants of the shape. But, for general recognition purpose, we do not have a priori knowledge on the plane in which the object lies, so $T_{\sigma\tau}$ remains unknown. At this stage, the computed invariants cannot be independent of the unknown parameters of $T_{\sigma\tau}$. With some assumptions on σ and τ, an approximation of the similarity invariants might be expected. This is discussed in the previous paper of Pizlo et al. [23].

In this discussion paper, $FCDP$ is defined by Pizlo et. al. to be composed of TS which has 3 d.o.f., together with the constraint that the object is in front of the camera. This constraint has been previously exploited by [13, 15] and also by people working on computer graphics. $E - FCDP$ is defined to be composed of T, S, and D, and not of the calibration matrix K. Thus $E - FCDP$ has 6 d.o.f. and is nothing but the transformation between the calibrated camera and the object plane.

It can be easily checked that T is a two-parameter transformation which fails to define a transformation group, since the product of two such transformations is no longer of the desired form. Therefore neither $FCDP$ nor $E - FCDP$ can be a group, as clearly indicated by the authors.

It means that neither $FCDP$ nor $E - FCDP$ can be a geometry in the modern view of the geometries. Even worse, it means that no $FCDP$ or $E - FCDP$ invariant can be defined, other than full projective invariants. To distinguish $FCDP$-inequivalent shapes that are nevertheless projectively equivalent it is necessary to use quasi-invariants. For this to work, it is necessary to assume that the images are taken from non-extreme view-points. Approximate invariants might be estimated with a priori knowledge as suggested in [23].

Thus, the authors have raised an interesting point: how can be characterized a shape observed by a calibrated camera? $FCDP$ and $E - FCDP$ might be more restricted transformations than others, but unfortunately, the invariants related to $FCDP$ and $E - FCDP$ are not computable from images.

2 What is projective geometry useful for?

2.1 Understanding the projection

As we have seen in the previous section, projective geometry nicely models perspective projection. Understanding what projective geometry is avoids some errors encountered in the past.

For instance, conics are equivalent within projective geometry. Understanding this point avoids the common mistaken assertion that circle projects only onto ellipses. In fact, if the observed circle is large enough, and if a part of it is positioned behind our visual system, it projects as a hyperbola. This hyperbola is however restricted to what is visible in front of us: only one branch exists in the image. This agrees with the known fact that convexity is preserved for a set of seen points.

From this well-known property that convexity is preserved for points that are seen, one may be led to a false assumption that convexity is preserved by perspective projection for all sets of points. At least this might be inferred from Pizlo's discussion paper. Let us consider a simple counter-example which nicely illustrates the fact that convexity is not preserved if the points under consideration are not observable ones.

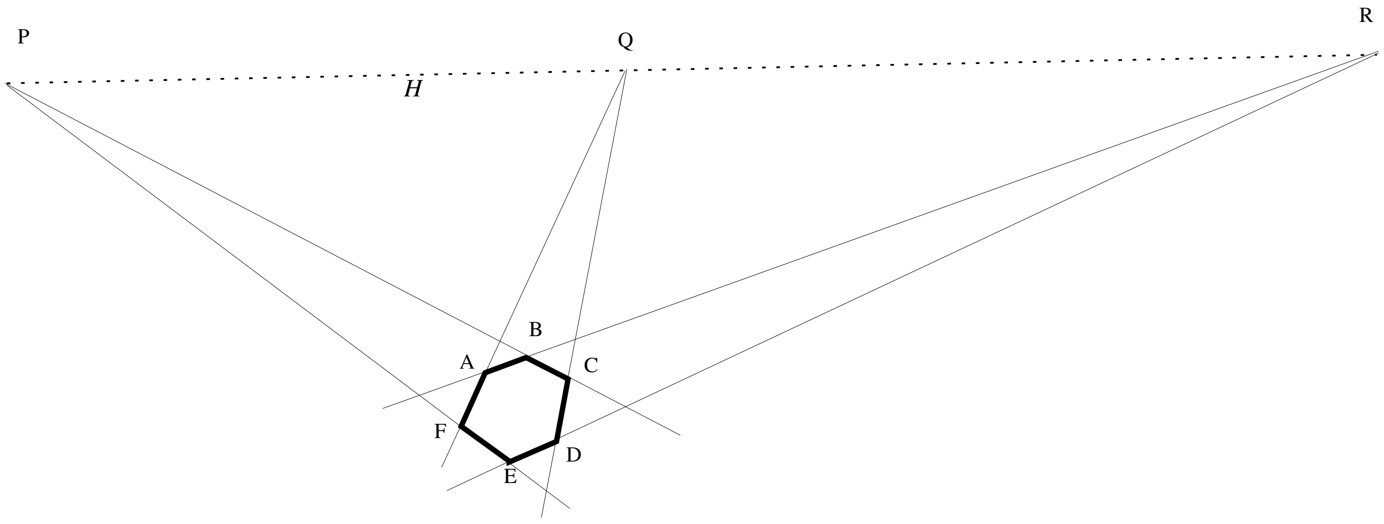

Figure 1: A flat hexagon with parallel sides.

Figure 1 show a planar projection of a hexagon with parallel sides. H is the projection of the vanishing line (horizon) of the hexagon plane, and P, Q, and R are projections of the vanishing points corresponding to intersections of the opposite sides. When the camera rotates to the left, points move to the right in the image. When AB becomes parallel to H, R is the point at infinity on H. After rotating a little bit more, R appears on the left of H, and at this point the order of P, Q, R is changed. In the context of projective geometry, this fact is not surprising since in this geometry, the projective line is closed on itself.

2.2 Providing generic geometric tools

Projective geometry is a well known scientific discipline [28], and it provides a set of wonderful tools for reasoning and computing.

Using projective geometry, several authors have been able to prove that 3D vision is possible without calibrated cameras [6, 12, 19, 25]. Much more, the fundamental matrix described in [6, 12] allows one to compute epipolar geometry in a reliable and efficient way just using point matches in the images (see [30], and their software available by ftp). Everybody who has calibrated a stereo head knows the burden that is thereby avoided. This has led to a set of different programs allowing structure to be computed either from multiple views, or from a set of completely uncalibrated cameras [4] or from a set of images taken by an unknown but single camera. Faugeras and Maybank [8, 17] pioneered the field, and their work is nicely integrated in general framework described in [29].

If calibration is needed, these kinds of tools offer ways to calibrate without the need of a calibration objet. A simple a robust approach can be found in [14].

2.3 Computing invariants

We have seen that if we want to restrict ourselves to real cases of perspective projection, or to calibrated cameras, no true invariants may be defined, other than invariants of the full projective group. On the other hand, the projective framework provides invariants. The basic invariant is the cross ratio and other useful invariants have been developed using geometric or algebraic tools. For the basic cross-ratio invariant, Maybank [16] presents a detailed study of the probability of false matches when using this invariant for indexing a set of shapes. Indexing using projective invariants have been demonstrated for planar figures by the Oxford team [27, 26], and for shapes consisting of sets of planar points in [18, 20].

Invariant methods have been extended to the 3D case where objects are observed by a pair of uncalibrated cameras [2, 11, 10]. Calibration is no longer needed in this case. One has just to know the correspondence between features (points, lines or conics) and the epipolar geometry. The latter may be computed from the correspondences. Such invariants may be computed for configurations of 6 points, 2 conics, 3 points and 2 lines, along with others.

This work was recently extended to the case where 6 points in 3D are observed in three images [24]. Of course, the correspondences have to be known but calibration is not needed. As the invariants associated with these points are the projective coordinates of the sixth point with respect to the five first ones, this method allow one to reconstruct the scene up to a projective transformation.

For a general overview, the reader is referred to [21, 22].

3 Conclusion

Projective geometry has led to considerable understanding of perspective projection. The large body of work that was developed around these tools in the last few years allows robust 3D image perception without camera calibration. Without doubt, theses tools are going to provide robust methods for the structure-from-multiple-image problem in the near future.In particular, projective geometry allows one to compute invariant values under perspective distorsion. These invariants have already been used for shape indexing and their performance for discriminating shapes has already been studied.

Unfortunately, there are fundamental difficulties in attempting to extend such methods to more restricted projection models, such as those defined by calibrated or partially calibrated cameras. In the case of projections of planar objects, this difficulty stems from the fact that the set of transformations that are induced by the image projection do not form a group. In fact, the smallest group containing the set of transformations is the full projective group. In restricting the camera model, one is throwing away a large and still growing body of effective and elegant techniques for scene reconstruction, or model indexing.

On the other hand, the restricted camera models discussed by Pizlo et. al. exclude extreme configuration which are not admissible for human perception. It is possible that there are some instances in which the added discriminatory power of a restricted model may be important, however, the problem of distinguishing shape becomes much harder. Projective geometry does not provide tools for these kinds of limitations. The only answers that might be found in the computer vision literature are quasi invariants or the cheirality invariants ([13]).

From Patrick.Gros@imag.fr Wed Oct 26 12:55:04 1994 Date: Wed, 26 Oct 94 17:49:22 +0100 From: Patrick.Gros@imag.fr (Patrick Gros) To: hartley@crd.GE.COM Subject: texte.bbl Content-Length: 10477 X-Lines: 216 Status: RO

References

[1] K. Åström. Fundamental difficulties with projective normalization of planar curves. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 377–389, October 1993.

[2] E.B. Barrett, M.H. Brill, N.N. Haag, and P.M. Payton. Invariant linear methods in photogrammetry and model-matching. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 122–128, 1992.

[3] T.O. Binford and T.S. Levitt. Quasi-invariants: theory ans exploitation. In *ARPA Image Understanding Workshop*, pages 819–829, 1993.

[4] B. Boufama, R. Mohr, and F. Veillon. Euclidian constraints for uncalibrated reconstruction. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 466–470, May 1993.

[5] M. Dhome, J.T. Lapreste, G. Rives, and M. Richetin. Spatial localization of Modelled Objects of Revolution in Monocular Prespective Vision. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 475–488, 1990.

[6] 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.

[7] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint.* Artificial intelligence. M.I.T. Press, Cambridge, MA, 1993.

[8] 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.

[9] P. Gros. Matching and clustering: two steps towards automatic model generation in computer vision. In *Proceedings of the AAAI Fall Symposium Series: Machine Learning in Computer Vision: What, Why, and How?, Raleigh, North Carolina, USA*, pages 40–44, October 1993.

[10] 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*, pages 75–86, July 1993.

[11] R. Hartley. Invariants of points seen in multiple images. Technical report, G.E. CRD, Schenectady, 1992.

[12] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 579–587. Springer-Verlag, 1992.

[13] R. I. Hartley. Cheirality invariants. In *DARPA Image Understanding Workshop*, pages 745–753, 1993.

[14] R.I. Hartley. An algorithm for self calibration from several views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 908–912, 1994.

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

[16] S.J. Maybank and P.A. Beardsley. Applications of invariant to model based vision. *Journal of Applied Statistics*, page to appear, 1994.

[17] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[18] P. Meer, S. Ramakrishna, and R. Lenz. Correspondence of coplanar features through P^2-invariant representations. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, page sumitted, 1994.

[19] R. Mohr, F. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 543–548, June 1993.

[20] L. Morin, P. Brand, and R. Mohr. Indexing with projective invariants. In *Proceedings of the Syntactical and Structural Pattern Recognition workshop, Nahariya, Israel*, page to appear. World Scientific Pub., 1994.

[21] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, Massachusetts, USA, 1992.

[22] J.L. Mundy and A. Zisserman, editors. *Proceedings of the second* ESPRIT - ARPA *workshop on Applications of Invariance on Computer Vision, Ponta Delagada, Azores, Portugal*, October 1993.

[23] Z. Pizlo and A. Rosenfeld. Recognition of planar shapes from perspective images using contour based-invariants. *Computer Vision, Graphics and Image Processing: Image Understanding*, 56(3):330–350, 1992.

[24] L. Quan. Invariants of 6 points from 3 uncalibrated images. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume II, pages 459–470. Springer-Verlag, 1994. also to appear in IEEE-Pami.

[25] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. IEEE *Transactions on PAMI*, 17(1), January 1995.

[26] C.A. Rothwell. Hierarchical object descriptions using invariants. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 287–303, October 1993.

[27] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. Canonical frames for planar object recognition. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 757–772. Springer-Verlag, May 1992.

[28] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry.* Oxford Science Publication, 1952.

[29] T. Vieville and Q.T. Luong. Computing motion and structure in image sequences without calibration. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages I 420–425, Jerusalem, 1994.

[30] J.Y. Zhang. Acquiring 3-D models from sequences of contours. IEEE *Transactions on PAMI*, 16(2):163–178, February 1994.