Normalized Shape

Richard I. Hartley[1] *and Roger Mohr*[2]

[1]GE - Corporate Research and Development,
P.O. Box 8, Schenectady, NY, 12301.
Ph : (518)-387-7333
Fax : (518)-387-6845
email : hartley@crd.ge.com

[2]: LIFIA - INRIA Rhône-Alpes
46, avenue F. Viallet 38031 Grenoble Cedex 1 - France

In this note, we would like to explore some of the issues suggested by the discussion paper of Pizlo et. al. The context is the study of image of planar objects taken with calibrated cameras. However, in this note, we shall consider the question using notation that has become somewhat standard in the literature of computer vision, particularly in the study of uncalibrated cameras.

1 Camera Model.

The projection from three-dimensional Euclidean space (or projective space) to the two dimensional image is conveniently expressed using formalism of projective geometry. Representing points in Euclidean space and in the image by homogeneous coordinates allows the mapping to be expressed in a very simple form as :

$$\mathbf{u} = M\mathbf{x} \tag{1}$$

where $\mathbf{u} = (u, v, w)^\top$ are the homogeneous coordinates in the image, $\mathbf{x} = (x, y, z, t)^\top$ is the point in space, expressed also in homogeneous coordinates, and M is a 3×4 matrix. Since we are using homogeneous coordinates, the matrix M is defined up to a constant factor only, and hence has 11 degrees of freedom.

The matrix M may be split up as a product

$$\begin{aligned} M &= K_{3\times3}(I_{3\times3} \mid \mathbf{0}) \begin{pmatrix} R_{3\times3} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \\ &= K(R \mid \mathbf{t}) \ . \end{aligned} \tag{2}$$

In this equation $R_{3\times3}$ is a rotation matrix, and the right-hand term in the equation is simply a Euclidean change of coordinates in object space. The central term, $(I \mid 0)$ represents the projection from 3 to 2 dimensions, and the left hand term K is a 3×3 upper triangular matrix representing an affine transformation of the image. The matrix K encodes the information about the calibration of the camera. In talking of calibrated cameras, we imply that K is known. Once again we may count 11 degrees of freedom : 6 for the Euclidean change of coordinates (three for rotation and three for translation) and 5 for the entries of the matrix K, which we defined only up to an irrelevant scale factor. All this is fairly standard. One may see [2] for a few more details.

Calibrated Cameras. If the camera is calibrated as in the discussion paper of Pizlo et. al. then the matrix K is know. In this case it is customary to correct for the calibration by replacing the point $\mathbf{u} = M\mathbf{x}$ by $\mathbf{u}' = K^{-1}u$, in which case we have $\mathbf{u}' = (R \mid \mathbf{t})\mathbf{x}$. Thus henceforth, one may assume that the calibration matrix K is in fact the identity. Correspondingly, we drop the notation $\mathbf{u}'$ and write in future $\mathbf{u} = (R \mid \mathbf{t})\mathbf{x}$. A calibrated camera has 6 degrees of freedom, three for the rotation and three for the translation of the camera.

Images of planar objects. Now we consider what happens when the object lies in a plane. We may assume for simplicity that the plane is the plane $z = 0$. In this case when we write

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = (R \mid \mathbf{t}) \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

the third column of the matrix R is irrelevant, since it is always multiplied by $z = 0$. Therefore, we may eliminate the third column of the matrix, and write

$$\mathbf{u} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}) \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the first two columns of the matrix R. This is the general form of the transformation from a planar object to the image by a calibrated camera. Notice that eliminating the third column, $\mathbf{r}_3$ from the matrix does not decrease the number of degrees of freedom of the transformation, since the vector $\mathbf{r}_3$ may always be retrieved by the formula $\mathbf{r}_1 \times \mathbf{r}_2 = \mathbf{r}_3$. This shows that the plane-to-plane transformation effected by a calibrated camera has 6 degrees of freedom. We will call such a transformation a *CP-transformation*, where C stands for "calibrated" and P stands for "planar". The set of all CP-transformations will be denoted by A. The set of CP-transformations does not form a group. We believe that the set of CP-transformations is the same as the FCDP transformations of the discussion paper.

2 Shape

Consider some object of interest in the in the object plane. To avoid overuse of the word object, we will refer to this object of interest as a *doodle*. A doodle may be a polygon, a set of points, a curve or some other geometrical entity. The image of a doodle will be called a doodle as well. Since the group of CP-transformations may be described by 6 parameters, the set of image doodles corresponding to a given object doodle is a 6-parameter family. How can this finding be reconciled with Pizlo's claim that the set of shapes that can be taken by the image of a planar object through projection with a calibrated camera form a 3-parameter family ?

First of all, we ought to define shape. Two doodles that differ by a translation or rotation should be considered to have the same shape. This leads to a definition of shape to be the equivalence class (orbit) of a doodle under the action of the group of rigid translations of the plane. Maybe one may also consider two doodles that differ only by a scaling to have

the same shape as well, and one might adopt a different definition of shape accordingly. However, this point will not make much difference in the future discussion

Suppose that C is a CP-transformation and R is a rotation (about the origin), then one quickly verifies that RC is a CP-transformation. On the other hand, if T is a transformation, then in general, TC is not a CP-transformation. Furthermore, if S represents isotropic scaling, then SC is also not a CP-transform, unless the rotation involved in the CP-transform C is the identity. In fact, in general for a rigid transformation S, we may observe that SC is a CP-transformation only if S is a rotation about the origin. Thus, if s represents a doodle in the object plane, and As represents the corresponding set of all images doodles, then in general Cs and $C's$ have the same shape only if $C' = RC$, where R is a rotation. Since the group of rotations is a one-parameter group, and the set of all image doodles is a 6-parameter set, we deduce that the set of all image shapes is a 5-parameter set. This seems interesting, so we will display it.

Proposition 2.1. *The set of all shapes that may appear as the image (with a calibrated camera) of a given planar shape forms a 5-parameter set.*

So what is meant by Pizlo et. al. when they claim that the set of shapes is a 3-parameter set ?

3 Normalized shape.

One of the consequences of assuming a calibrated camera is that one knows the principal point of the camera, which acts as a distinguished point (in fact the coordinate origin) in the image. Now, consider a given shape in the object plane. In the image plane, it will appear as a certain given shape. If one holds the camera centre (and of course also the object) stationary while changing the camera orientation (panning), then of course the object will change position in the image. What is not quite so obvious, but is also true is that the shape of the imaged object will also change. This effect is readily perceived by looking through the view-finder of a camera with large field of view. This effect was exploited in [2] for camera calibration. In fact, it was shown in [2] that the image-to-image transformation that takes place when the camera is panned is a projective transformation of the image represented by a matrix H of the form $H = KSK^{-1}$, where K is the calibration matrix and S is a rotation matrix. In our case, we are assuming that K is the identity. Thus, the resulting transformation is a projective transformation represented by a rotation matrix S. (Note that this does not mean that the transformation is a rotation, since we are talking about a 3×3 transformation of *homogeneous* coordinates.) It may be verified easily that if S is a 3×3 rotation matrix, and C is a CP-transformation, of the form $(\mathbf{r}_1\mathbf{r}_2\mathbf{t})$ then the product SC is also a CP-transformation, since the two first columns of SC are the first two columns of the rotation matrix SR.

Now, suppose that we were to see a doodle in an image, placed far from the principal point. Suppose we asked the following question : "What would this doodle look like in the image if we were to point the camera straight at the object doodle?". The image doodle would then be placed near to the principal point. This question could easily be answered by choosing a 3×3 rotation matrix that, when interpreted as a 2D projective transformation, moves the image doodle to the principal point.

Let us make this more precise. Suppose that there is a distinguished point p_0 on the object doodle s (for instance a given vertex). Let point p_0 be fixed also as the origin of the object plane. Given the image of this doodle under a CP-transformation, C one may compute a further transformation S, represented by a 3×3 rotation matrix that moves the image of the distinguished point, Cp_0, to the origin of image coordinates (the principal point). The composition of the imaging transform, C and the normalizing transform S is a CP-transform taking the origin of the object plane to the origin of image coordinates. Since S does not represent a rigid transformation, it transforms the shape of the image doodle Cs to a *normalized shape* represented by the doodle SCs.

Given an image doodle and specification of a distinguished point, one can easily compute a representative normalized doodle by finding the appropriate normalizing transformation S. The transformation S that maps the distinguished point to the origin is not unique. However, it is easily seen that any two such transformations S and S' differ by a 2D rotation about the origin. Such a rotation preserves shape. Hence, the normalized shape is uniquely defined, and a representative normalized doodle is easily computed.

A CP-transformation C that takes the origin of the object plane to the origin of image coordinates (and hence a doodle to a normalized image doodle) must be of the form

$$C = \begin{pmatrix} & & 0 \\ \mathbf{r}_1 & \mathbf{r}_2 & 0 \\ & & k \end{pmatrix} \quad . \tag{3}$$

Transformations of this type form a 4-parameter set. Since a further rotation about the origin preserves shape, we deduce the following result

Proposition 3.2 Pizlo et. al.. *The set of normalized shapes corresponding to a given planar object doodle with a distinguished point constitutes a 3-parameter set.*

4 Questions

A number of questions naturally arise

Identification. Given a certain known object doodle s, how can one tell whether a given image doodle is (or may be) an image of s. This question is considered in Pizlo's paper.

For the case where both object and image doodles are triangles, it seems likely that there is always a CP-transformation that takes the object to the image triangle. One can always assume that corresponding points lie at the origin of object and image coordinates. In this case, the CP-transformation can be assumed to have the form (3). Every other pair of points give rise to 2 linear equations in the entries of the matrix. In addition, there are two quadratic constraints indicating that the two first columns are orthogonal and of the same length. This gives sufficiently many equations to determine the entries of the matrix M. Given that we have two quadratic equations, it is possible to have up to four solutions. It is also possible to have no real solutions. It would be interesting to know if there actually exist two triangles for which no real solution exists.

For point configurations consisting of more than 3 points, one can answer the identification question by first finding the CP-transformation or transformations that realize the

correspondence between three points, and then see if any of the possible transformations gives the required correspondence on the rest of the points.

Correspondence. Suppose one has two images of doodles. Is it possible that they are images of the same (unknown) object doodle ? The answer to this question seems to be yes if and only if there is a projective transformation taking one of the doodles to the other. The *only if* part of this claim is obvious, since two projections of the same planar doodle must be projectively equivalent. To justify the *if* part, suppose there is a projective transformation H taking one doodle to the other. Let $\mathbf{u}_i$ be four points in the one image, and $\mathbf{u}'_i$ be the four corresponding points in the other image. The projective transform is determined by the correspondence of the four points. We seek a set of four object points $\mathbf{x}_i$ and two CP-transforms C and C' such that $C\mathbf{x}_i = \mathbf{u}_i$ and $C'\mathbf{x}_i = \mathbf{u}'_i$. Counting the number of unknowns, we see that there are

1. For each transformation, 9 matrix entries.
2. For each of the 4 object points, 2 coordinates.

This gives a total of $9 \times 2 + 2 \times 4 = 26$ unknowns. On the other hand, counting equations, there are

1. For each transformation, 3 quadratic constraints expressing the fact that the first two columns are the columns of a rotation matrix.
2. For each point correspondence, 2 linear equations expressing the correspondence.

This gives a total of $3 \times 2 + 4 \times 2 \times 2 = 22$ equations. The system is therefore underdetermined, and one can expect a 4-parameter family of solutions. Allowing 3 parameters for the placement and orientation of the object doodle, we are reduced to a 1-parameter family of solutions. It may also be observed that the overall size of the object can not be determined from two views. Allowing for scale ambiguity, we are reduced to a 0-parameter family of solutions, namely a finite set of solutions. This of course ignores such issues as the existence of only complex solutions and cheirality ([1, 3]).

Thus, we deduce that there exists at least one pair of transformations C and C' and points $\mathbf{x}_i$ that satisfy $C\mathbf{x}_i = \mathbf{u}_i$ and $C'\mathbf{x}_i = \mathbf{u}'_i$ for $i = 1, \ldots, 4$. From this it follows that $C'C^{-1}\mathbf{u}_i = \mathbf{u}'_i = H\mathbf{u}_i$ for $i = 1, \ldots, 4$. Since H and $C'C^{-1}$ correspond on four points, they must be equal. So $H = C'C^{-1}$. Now, for points u_i for $i \geq 5$ we may define $\mathbf{x}_i$ by $\mathbf{x}_i = C^{-1}\mathbf{u}_i$. One then verifies that $C\mathbf{x}_i = \mathbf{u}_i$ and also that $C'\mathbf{x}_i = C'C^{-1}\mathbf{u}_i = H\mathbf{u}_i = \mathbf{u}'_i$. In summary, there exist points $\mathbf{x}_i$ for all i that map to the given image points under the two constructed CP-transformations. These points $\mathbf{x}_i$ form the desired object doodle.

Reconstruction. From the previous argument, it appears that a planar doodle can be reconstructed up to rigid transformation from its projection in two images, as long as 4 point correspondences are known. This is much the same as in the case of three dimensional objects where 8 point correspondences are enough to give a unique solution ([3]). In the case of 4 planar point correspondences, We do not know how many distinct solutions are possible, though the number should be at most 16.

References

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

[2] Richard I. Hartley. Self-calibration from multiple views with a rotating camera. In *Computer Vision - ECCV '94, Volume I, LNCS-Series Vol. 800, Springer-Verlag*, pages 471–478, May 1994.

[3] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Sept 1981.