PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Geometric structure and view invariant recognition

The Royal Society

Phil. Trans. R. Soc. Lond. A 1998 **356**, 1233-1250
doi: 10.1098/rsta.1998.0219

The Royal Society

Geometric structure and view invariant recognition

By Stefan Carlsson

Department of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden

By organizing object recognition as indexing a look-up table of model object features, it can be made effective in terms of time and memory complexity. However, for *general* three-dimensional (3D) objects it is not possible to compute non-trivial object-specific descriptors from a single view that are invariant to *general* choices of viewpoint. This raises the question of how to effectively organize recognition of 3D objects from single views by making various compromises w.r.t. invariance and efficiency of representations. For general shapes we can derive *shape constraints*, invariant to viewpoint and other camera parameters, that relate 3D and image structure. These relations can be used for verification of the presence in an image of a specific 3D object but they do not allow for the computation of view invariant indexes. In order to have an indexing system for recognition, there are basically two alternative options: if we want complete view invariance we have to restrict the class of objects. Alternatively, if we want methods that work for general unconstrained object types, we have to restrict the range of viewpoints over which invariant descriptors can be computed. We will see how this naturally leads to the introduction of *incidence* and *order* structure respectively, as a basis for shape description. The hierarchy of geometric structure descriptions: projective/affine, order and incidence can all be described in a unified way in terms of properties of bracket expressions of image coordinates in arbitrary frames.

Keywords: recognition; invariance; affine structure; projective structure; incidence geometry; order structure

1. Introduction

The geometric shape that a three-dimensional (3D) object projects to an image depends on the relative viewpoint, as well as on the internal parameters of the camera. For parallel and perspective projection, the variation due to internal camera parameters can be accounted for by the use of affine and projectively invariant image descriptors, respectively (Koenderink & van Doorn 1991; Sparr 1991; Faugeras 1992; Hartley *et al*. 1992). The variation due to changes in relative viewpoint of the camera is more complicated, however. Ideally one would like image descriptors that are discriminating between different objects and also independent of the viewpoint. It has been shown, however, that no such descriptors exists (Burns *et al*. 1993; Moses & Ullman 1992). For a *general* set of points in 3D it is not possible to compute a *general* view invariant, that is also discriminating w.r.t. other point-sets. Invariant descriptors are useful in that they can be used as indexing keys to look-up tables of object features, making the recognition efficient in terms of time and memory

complexity. The time efficiency comes from the fact that it is not necessary to match all objects in the model library to image data and the memory efficiency from the fact that all views of an object are represented by the invariants.

The non-existence of general view invariants of course raises the question of how to most effectively organize recognition of 3D objects. We will see that the options available imply the sacrifice of some of the efficiency of an ideal view invariant indexing system. We will look at three different cases:

1. There are *general view* invariant relations constraining affine, or projective 3D and image structure for *general shapes*.

   These relations can be used for viewpoint invariant verification but do not admit the construction of indexing keys. The time efficiency of recognition is therefore lost.

2. *General view* invariants can be computed for *restricted shapes*.

   This will require *a priori* information about the nature of the shape restriction. We will see that for sufficiently restricted shapes it is possible to identify the type of shape restriction from *incidence relations* of an image in a single view.

3. *Restricted view* invariants can be computed for *general shapes*.

   By extending the equivalence class of objects corresponding to a certain set of image descriptors we can increase their view invariance properties. The problem is then to choose descriptors so that equivalence classes are extended in a reasonable way. We will argue that the concept of *order structure* is a natural choice for a descriptor in this sense.

2. View invariant shape constraints for affine and projective structure

For points in general, position in three dimensions we will derive canonical projection relations involving affine and projective structure for parallel and perspective camera projection models, respectively. As will be seen, these relations will make explicit exactly how image structure depends on 3D structure and viewpoint. No other internal or external camera parameters will appear in these relations.

(*a*) *Parallel projection: affine structure*

In the parallel projection case we use four points with Cartesian† coordinate vectors $\bar{P}_1$, $\bar{P}_2$, $\bar{P}_3$, $\bar{P}_4$ to define affine coordinates in 3D (see Appendix A):

$$\bar{P}_n - \bar{P}_4 = \bar{X}_n(\bar{P}_1 - \bar{P}_4) + \bar{Y}_n(\bar{P}_2 - \bar{P}_4) + \bar{Z}_n(\bar{P}_3 - \bar{P}_4)$$

and the corresponding image points $\bar{p}_1$, $\bar{p}_2$ and $\bar{p}_4$ to define affine coordinates in the image:

$$\bar{p}_n - \bar{p}_4 = \bar{x}_n(\bar{p}_1 - \bar{p}_4) + \bar{y}_n(\bar{p}_2 - \bar{p}_4).$$

For parallel projection, the relation between any coordinate representation in three dimensions and the image can be written:

$$\bar{p}_n = \boldsymbol{M}\bar{P}_n + \bar{m}, \tag{2.1}$$

† Cartesian coordinates are denoted by $\bar{P}$ in order to distinguish them from homogeneous coordinates in projective space denoted P.

where $\boldsymbol{M}$ is a 2×3 matrix and $\bar{m}$ is a 2-vector. The vector $\bar{m}$ can be eliminated by taking differences:

$$\bar{p}_n - \bar{p}_4 = \boldsymbol{M}(\bar{P}_n - \bar{P}_4). \tag{2.2}$$

Using the affine coordinates defined in (2.1) we find that $\boldsymbol{M}$ must satisfy

$$\begin{pmatrix} 1 & 0 & \bar{x}_3 & 0 \\ 0 & 1 & \bar{y}_3 & 0 \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} 1 & 0 & 0 & \bar{X}_0 \\ 0 & 1 & 0 & \bar{Y}_0 \\ 0 & 0 & 1 & \bar{Z}_0 \end{pmatrix}, \tag{2.3}$$

where $\bar{X}_0$, $\bar{Y}_0$, $\bar{Z}_0$ are the affine coordinates of an arbitrary point on a line through the origin point P_4 in the view direction. The point $(\bar{X}_0, \bar{Y}_0, \bar{Z}_0)$ represents the view direction and projects to the origin point $(\bar{x}_4, \bar{y}_4) = (0, 0)$.

Using the relations in equation (2.3) we can solve for the elements of the projection matrix $\boldsymbol{M}$. If we take points 1, 2 and the 'viewpoint' 0 we get the projection equations:

$$\begin{pmatrix} \bar{x}_n \\ \bar{y}_n \end{pmatrix} = {\bar{Z}_0}^{-1} \begin{pmatrix} \bar{Z}_0 & 0 & -\bar{X}_0 \\ 0 & \bar{Z}_0 & -\bar{Y}_0 \end{pmatrix} \begin{pmatrix} \bar{X}_n \\ \bar{Y}_n \\ \bar{Z}_n \end{pmatrix}. \tag{2.4}$$

This is the canonical projection equation for parallel projection. It relates affine structure in 3D to that in the image depending on the view direction only. Using the fact that $(\bar{X}_3, \bar{Y}_3, \bar{Z}_3) = (0, 0, 1)$ we get

$$\bar{x}_3 = -\bar{X}_0/\bar{Z}_0, \qquad \bar{y}_3 = -\bar{Y}_0/\bar{Z}_0. \tag{2.5}$$

The fact that the view direction can be eliminated so easily in the parallel projection is coupled to the fact that it can actually be determined uniquely from the observation of four image points. We see that equation (2.5) implies

$$(\bar{X}_0, \bar{Y}_0, \bar{Z}_0) = \sigma(-\bar{x}_3, -\bar{y}_3, 1), \tag{2.6}$$

where σ is an arbitrary scale factor.

In parallel projection we can compute affine information about the view direction without any knowledge of the 3D structure of the point-set. Note that the sign of the arbitrary factor σ in the view direction is connected with a 'necker-reversal' of the 3D structure of the point-set.

If the view direction (2.6) is substituted in the projection equation (2.4) we get

$$\left.\begin{aligned} \bar{x}_n - \bar{X}_n - \bar{x}_3\bar{Z}_n = 0, \\ \bar{y}_n - \bar{Y}_n - \bar{y}_3\bar{Z}_n = 0, \end{aligned}\right\} \tag{2.7}$$

which are relations constraining affine 3D and image structure in a view invariant way. This form of the affine shape constraints was derived in Clemens & Jacobs (1991). Equivalent relations for affine shape constraints were also derived in Weinshall (1993). The linearity of the relations implies that image coordinates of any view can be expressed as a linear combination of those of two other views (Ullman & Basri 1991).

These constraint equations can be used to verify the presence of an object with known affine structure in a way that is totally independent of camera parameters and viewpoint. Unless the affine structure is constrained, however, they do not permit the explicit computation of view invariant descriptors from a single image.

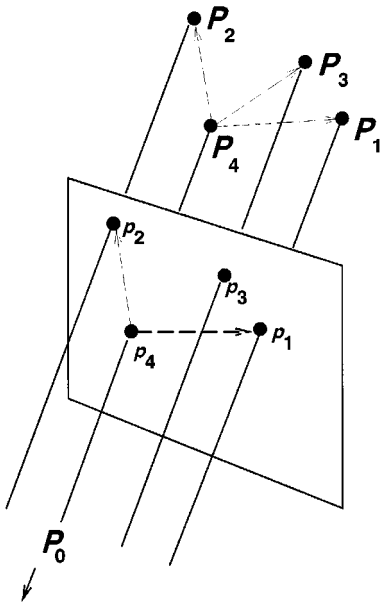

Figure 1. Affine basis in 3D and image for parallel projection.

If the affine coordinates are expressed in terms of brackets of normalized homogeneous coordinates in arbitrary frames (see Appendix A), using lower-case for image coordinates, these constraints can be written as the bracket polynomials:

$$\left.\begin{aligned}[2^*4^*5^*][1^*2^*3^*4^*] - [1^*2^*4^*][2^*3^*4^*5^*] + [2^*3^*4^*][1^*2^*4^*5^*] = 0,\\ [1^*4^*5^*][1^*2^*3^*4^*] - [1^*2^*4^*][1^*3^*4^*5^*] + [1^*3^*4^*][1^*2^*4^*5^*] = 0.\end{aligned}\right\} \tag{2.8}$$

This form of the constraint equations makes explicit the fact that the constraints are a property of a group of five points irrespective of the choice of coordinate frame.

(*b*) *Perspective projection: projective structure*

We use the homogeneous coordinates of five points $P_1, \ldots, P_5$ to define a projective coordinate system in 3D and the corresponding coordinates of image points p_1, p_2, p_3 and p_4 to define a projective coordinate system in the image (see Appendix B).

The projective coordinates in 3D and the image are then related by a general linear transformation given by the 3×4 matrix $\boldsymbol{M}$,

$$p = \boldsymbol{M}P,$$

which can be computed from the mappings:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} 1 & 0 & 0 & 0 & X_0 \\ 0 & 1 & 0 & 0 & Y_0 \\ 0 & 0 & 1 & 0 & Z_0 \\ 0 & 0 & 0 & 1 & W_0 \end{pmatrix}, \tag{2.9}$$

where X_0, Y_0, Z_0, W_0 are the projective coordinates of the perspective projection point in the five-point basis.

If we solve for $\boldsymbol{M}$ we find that 3D and image homogeneous projective coordinates are related by the canonical perspective projection equation:

$$\begin{pmatrix} x_n \\ y_n \\ w_n \end{pmatrix} = \sigma \begin{pmatrix} {X_0}^{-1} & 0 & 0 & -{W_0}^{-1} \\ 0 & {Y_0}^{-1} & 0 & -{W_0}^{-1} \\ 0 & 0 & {Z_0}^{-1} & -{W_0}^{-1} \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \\ Z_n \\ W_n \end{pmatrix}.$$

We can write these relations in the form of constraint equations as

$$\left.\begin{aligned} w_n \frac{Y_n}{Y_0} - y_n \frac{Z_n}{Z_0} + (y_n - w_n)\frac{W_n}{W_0} = 0, \\ w_n \frac{X_n}{X_0} - x_n \frac{Z_n}{Z_0} + (x_n - w_n)\frac{W_n}{W_0} = 0. \end{aligned}\right\} \tag{2.10}$$

Note that this equation constrains 3D point position and the inverse of camera position in exactly the same way. The problems of recovering 3D structure and multiple camera positions from image data are therefore computationally dual for the uncalibrated perspective camera (Carlsson 1995*b*; Carlsson & Weinshall 1998). Specifically, using multiple image data we can eliminate the coordinates of the 3D shape and get epipolar constraints relating camera positions and image data. Alternatively, using multiple points in one image, we can eliminate camera positions and get shape constraints, relating 3D and image structure. For multiple points, $5, 6, 7, \ldots$ the constraint equations (2.10) can be written as the system

$$\begin{pmatrix} 0 & w_5Y_5 & -y_5Z_5 & (y_5 - w_5)W_5 \\ w_5X_5 & 0 & -x_5Z_5 & (x_5 - w_5)W_5 \\ 0 & w_6Y_6 & -y_6Z_6 & (y_6 - w_6)W_6 \\ w_6X_6 & 0 & -x_6Z_6 & (x_6 - w_6)W_6 \\ 0 & w_7Y_7 & -y_7Z_7 & (y_7 - w_7)W_7 \\ w_7X_7 & 0 & -x_7Z_7 & (x_7 - w_7)W_7 \\ & & \vdots & \end{pmatrix} \begin{pmatrix} X_0^{-1} \\ Y_0^{-1} \\ Z_0^{-1} \\ W_0^{-1} \end{pmatrix} = 0. \tag{2.11}$$

Since this system has rank < 4 any determinant formed by taking four arbitrary rows of the system must vanish giving the shape constraint relations (Carlsson 1995*b*; Weinshall *et al*. 1995; Carlsson & Weinshall 1998). These constraints are dual to the multiple view matching constraints obtained in the same way (Faugeras & Mourrain 1995).

Using the fact that $(X_5, Y_5, Z_5, W_5) = (1, 1, 1, 1)$ and taking the determinant of the first four rows of equation (2.11) we get the shape constraint for six points

$$\begin{aligned}(w_5y_6 - x_5y_6)X_6Z_6 + (x_5y_6 - x_5w_6)X_6W_6 + (x_5w_6 - y_5w_6)X_6Y_6 \\ + (y_5x_6 - w_5x_6)Y_6Z_6 + (y_5w_6 - y_5x_6)Y_6W_6 + (w_5x_6 - w_5y_6)Z_6W_6 = 0.\end{aligned}$$

This specific relation was derived in Quan (1994). Similar relations using a different representation can be found in the work of Sparr (1991). In the same way as equation (2.7) these equations can be used to verify the presence in an image of an object with a known projective structure independent of camera calibration and viewpoint. If the projective coordinates are expressed in terms of brackets of Cartesian coordinates in arbitrary frames (see Appendix B) and these are straightened using the

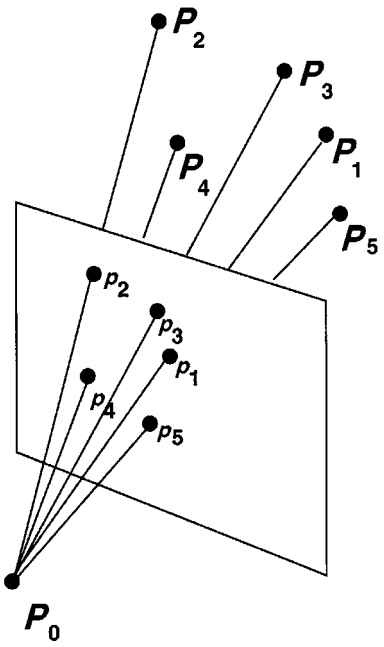

Figure 2. Projective basis in 3D and image for perspective projection.

Plucker–Grassmann relations (Hodge & Pedoe 1947), this constraint can be written as the bracket polynomial,

$$\begin{aligned}
&[1\ 3\ 6]\ [2\ 4\ 5]\ [1\ 2\ 4\ 6]\ [1\ 2\ 5\ 6]\ [1\ 3\ 4\ 5]\ [3\ 4\ 5\ 6]\\
-&[1\ 2\ 6]\ [3\ 4\ 5]\ [1\ 3\ 4\ 6]\ [1\ 3\ 5\ 6]\ [1\ 2\ 4\ 5]\ [2\ 4\ 5\ 6]\\
+&[1\ 3\ 4]\ [2\ 5\ 6]\ [1\ 2\ 4\ 6]\ [1\ 3\ 5\ 6]\ [1\ 2\ 4\ 5]\ [3\ 4\ 5\ 6]\\
-&[1\ 2\ 4]\ [3\ 5\ 6]\ [1\ 3\ 4\ 6]\ [1\ 2\ 5\ 6]\ [1\ 3\ 4\ 5]\ [2\ 4\ 5\ 6]\\
+&[1\ 3\ 5]\ [2\ 4\ 6]\ [1\ 3\ 4\ 6]\ [1\ 2\ 5\ 6]\ [1\ 2\ 4\ 5]\ [3\ 4\ 5\ 6]\\
-&[1\ 2\ 5]\ [3\ 4\ 6]\ [1\ 2\ 4\ 6]\ [1\ 3\ 5\ 6]\ [1\ 3\ 4\ 5]\ [2\ 4\ 5\ 6] = 0. \qquad (2.12)
\end{aligned}$$

This form of the six point constraint has the advantage that it does not assume the existence of a five point projective basis (Carlsson 1995*a*).

3. General view invariants for restricted shapes

The shape constraints (2.7) and (2.12) can be used to answer questions such as, can this image be the projection of this 3D affine/projective shape, independent of camera calibration and viewpoint? They do not, however, permit the construction of view invariant indexing keys for table look-up. In order for this to be possible, we would like to write the relations as

$$i(x_1, y_1 \ldots x_n, y_n) = I(X_1, Y_1, Z_1 \ldots X_n, Y_n, Z_n), \qquad (3.1)$$

where the dependence on image and 3D data has been separated. This can be achieved if we can solve for 3D structure in terms of image structure. This is of course not possible in the case of general shapes, a mere consequence of the fact that depth information is lost in the projection process. However, by adding extra shape

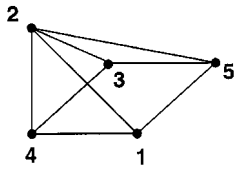

Figure 3. Five points, 1, 3, 4, 5 coplanar.

constraints to the relations (2.7) and (2.12), i.e. by considering restricted instead of general shapes, the 3D structure can be computed explicitly in terms of image structure. An important shape constraint is coplanarity, which is a projectively invariant property for a set of points and thereby also affine invariant. Indeed, by restricting all points in view to be coplanar, projective and affine image representations will be general view invariants for the case of perspective and parallel projection respectively. However, we will see that we can actually exploit also partial coplanarity constraints in order to compute general view invariants.

(*a*) *Parallel projection*

Suppose we have five points in parallel projection where point 5 lies in the plane spanned by points $1, 2$ and 3. This means that the affine coordinate $Y_5 = 0$. If we use this together with the affine shape constraints from the image equation (2.7), we get the linear system:

$$\left.\begin{aligned} Y_5 &= 0, \\ x_5 - X_5 - x_3 Z_5 &= 0, \\ y_5 - Y_5 - y_3 Z_5 &= 0. \end{aligned}\right\} \tag{3.2}$$

The 3D affine structure can be computed by solving this for X_5, Y_5 and Z_5

$$X_5 = \frac{x_5 y_3 - x_3 y_5}{y_3}, \qquad Y_5 = 0, \qquad Z_5 = \frac{y_5}{y_3}.$$

A single coplanarity constraint among five points therefore permits the computation of 3D affine structure from a parallel projection image. That is, we can compute a view invariant representation of the object from image data. Using equation (2.8) and straightening the bracket expressions, these relations can be expressed in terms of brackets of normalized homogeneous coordinates:

$$\frac{[2^*3^*4^*5^*]}{[1^*2^*3^*4^*]} = \frac{[3^*4^*5^*]}{[1^*3^*4^*]}, \qquad [1^*3^*4^*5^*] = 0, \qquad \frac{[1^*2^*4^*5^*]}{[1^*2^*3^*4^*]} = -\frac{[1^*4^*5^*]}{[1^*3^*4^*]}. \tag{3.3}$$

(*b*) *Perspective projection*

In the case of perspective projection if we take six points and constrain them so that we get two four-point coplanarities we get the so-called 'butterfly' configuration (see figure 4).

The coplanarity constraints can be expressed in terms of brackets and after using straightening relations, in terms of the projective coordinates as

$$\left.\begin{aligned} [1346] = 0 \quad &\Longrightarrow \quad Y_6 = 0, \\ [2456] = 0 \quad &\Longrightarrow \quad Z_6 - X_6 = 0. \end{aligned}\right\} \tag{3.4}$$

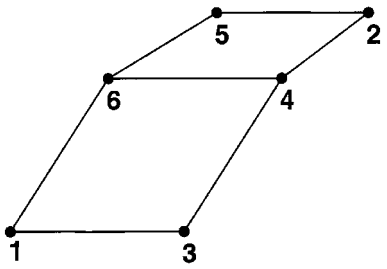

Figure 4. Six points, 1, 3, 4, 6 and 2, 4, 5, 6 coplanar.

If we add these constraints to the perspective projection six-point constraint (2.12) we can solve for the projective structure of the six-point configuration:

$$\frac{X_6}{W_6} = \frac{Z_6}{W_6} = \frac{x_6 w_5 - y_6 w_5 + x_5 y_6 - x_5 w_6}{x_5 y_6 - y_6 w_5}. \tag{3.5}$$

Using the bracket expressions for the projective coordinates (Appendix B) this can be written after straightening:

$$\frac{X_6}{W_6} + 1 = \frac{Z_6}{W_6} + 1 = \frac{[1256][1345]}{[1245][1356]} = \frac{[134][256]}{[136][245]}. \tag{3.6}$$

This can alternatively be derived directly from the bracket expression for the shape constraint (2.12) using the coplanarity constraints equation (3.4).

The effect of coplanarity constraints in 3D has been analysed extensively for polyhedral type objects in Sugihara (1986), Sparr (1992) and Rothwell *et al.* (1993), using various projection and calibration models.

(*c*) *Bilateral symmetry: incidence structure*

By introducing the extra constraint of coplanarity in three dimensions we saw that it is possible to compute view invariants from image data and that these view invariants are actually the affine/projective structure of the shape in three dimensions. This introduces the problem, however, of deciding that a certain coplanarity is present in three dimensions. For a general planar set of points in three dimensions, there is no constraint in a single image that can be used to decide coplanarity among the points. If the set of points is constrained even further, however, we will see that this induces view invariant constraints in the image. This is the case for bilaterally symmetric structures.

We take a six-point butterfly configuration and constrain it so that there is a symmetry plane dividing the points into two groups, $1, 3, 5$ and $2, 4, 6$ (see figure 5).

By intersecting lines connecting various points we can generate points a, b, c and d. The symmetry constraint in three dimensions implies that the three-dimensional lines ab and cd both lie in the symmetry plane. They therefore intersect the line 34 in a common point q. The fact that q lies on the line 34 can be expressed as a collinearity constraint of three image points as

$$[3\ 4\ q]{=}0. \tag{3.7}$$

The point q is the intersection of lines ab and cd in the image. The image coordinates of q can therefore be expressed using the Grassmann–Cayley algebra meet operation

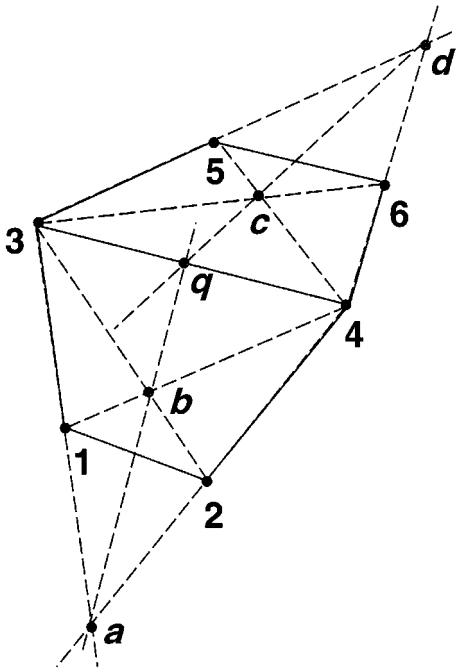

Figure 5. Six points 1 . . . 6 in bilateral symmetry.

(Barnabei *et al*. 1985; Carlsson 1994), as

$$q=ab\wedge cd=[a\ b\ c]\ d-[a\ b\ d]\ c. \tag{3.8}$$

If this is substituted in (3.7) we get

$$[a\ b\ c]\ [3\ 4\ d]-[a\ b\ d]\ [3\ 4\ c]=0. \tag{3.9}$$

Points a, b, c and d can be expressed as intersections:

$$a=13\wedge 24=[1\ 2\ 3]\ 4+[1\ 3\ 4]\ 2,$$
$$b=14\wedge 23=[1\ 2\ 4]\ 3-[1\ 3\ 4]\ 2,$$
$$c=36\wedge 45=[3\ 4\ 6]\ 5-[3\ 5\ 6]\ 4,$$
$$d=35\wedge 46=[3\ 4\ 5]\ 6+[3\ 5\ 6]\ 4.$$

If these are substituted in (3.9), we get after factoring out common factors

$$[1\ 2\ 3]\ ([2\ 4\ 5]\ [3\ 4\ 6]-[2\ 4\ 6]\ [3\ 4\ 5])+[1\ 2\ 4]\ ([2\ 3\ 5]\ [3\ 4\ 6]-[2\ 3\ 6]\ [3\ 4\ 5]\ )$$
$$=[1\ 2\ 3]\ [4\ 5\ 6]-[1\ 2\ 4]\ [3\ 5\ 6]=0, \tag{3.10}$$

where we have used a straightening operation in the last step. This is a viewpoint invariant constraint on six image points 1 . . . 6 that is fulfilled if the six points form a bilateral symmetry as defined above. It can therefore be used to verify the presence of a bilateral symmetry from image data in one perspective projection.

This expression can be given another even simpler geometric interpretation. by factorizing into the expression

$$12\ \wedge\ 34\ \wedge\ 56=[1\ 2\ 3]\ [4\ 5\ 6]-[1\ 2\ 4]\ [3\ 5\ 6]=0. \tag{3.11}$$

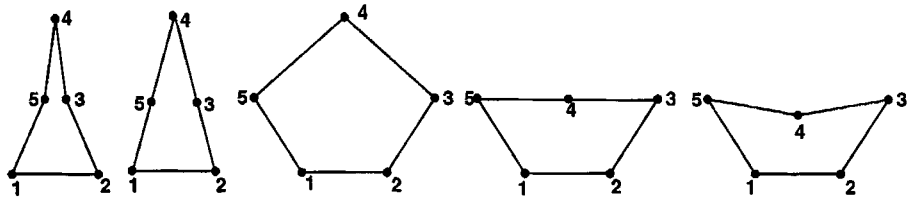

Figure 6. Shapes with different order and incidence structure.

Geometrically this means the condition for the three lines 12, 34 and 56 to intersect in a common point. In our case this point is the vanishing point since the three lines are parallel in three dimensions due to the symmetry constraint.

Note that we can carry out exactly the same reasoning for the case where we have a symmetry plane along the line 34 with points 12 and 56 symmetrically related, giving the same symmetry constraint. The constraint (3.10) represents an example of view invariant image *incidence structure*. In the next section we shall see how the concept of qualitative structure can be extended from incidence structure of projected constrained shapes to *order structure* of generic shapes in projection.

4. Restricted view invariants for general shapes

(*a*) *Order structure*

The fact that general view invariants can be defined for restricted 3D shapes is of limited use when considering general shapes. For general shapes, any image descriptor will have to display variation w.r.t. change of viewpoint or will otherwise be useless for discrimination purposes. We could always assign the same descriptor to any view of any object, giving us complete view invariance but no discrimination between different objects. This illustrates the fact that in general, discriminability and view invariance are conflicting objectives in recognition. A direct indexing method must therefore be a compromise between these objectives.

Any descriptor that is used as an indexing key to a look-up table will by definition be discrete. A straightforward way to define discrete descriptors is to quantize continuous image measurements, possibly linear invariants, into discrete bins as is done in geometric hashing (Lamdan *et al*. 1988). Any quantization strategy will then define equivalence classes of image structure that give rise to the same discrete index. Ideally these equivalence classes should contain as many views as possible of a specific object. This inevitably implies loss in discriminability relative to other objects, i.e. the equivalence classes have to be enlarged.

A crucial step in the design of an indexing system based on image measurements is therefore to ensure that the necessary enlargement of the equivalence classes is based on a relevant similarity measure. A natural extension of equivalence classes is to consider object categories. Although it is not possible to give a strict geometric definition of the concept of object category, it is generally believed that one needs representations based on qualitative, relational properties as opposed to metrical descriptions used to define object instances (Marr & Nishihara 1981; Biederman 1985). The introduction of incidence structure in the previous section is a step in this direction but it applies only to restricted shapes. For general shapes the concept of *order structure* of groups of features can be used to describe qualitative properties.

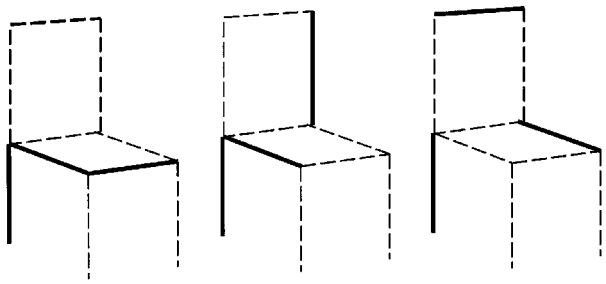

Figure 7. Examples of order and incidence types for triplets of line segments.

Figure 6 shows a sequence of five-point configurations that are successively deformed. These configurations are all completely characterized by the *order* and *incidence* structure given by the mapping

$$\chi(i,j,k) = \mathrm{sgn}[i^* \; j^* \; k^*] \longrightarrow \{-1,0,1\} \tag{4.1}$$

for all combinations of points $i,j,k \in \{1,2,3,4,5\}$, where i^* denotes the oriented homogeneous coordinates of point i as defined in Appendix A (Björner *et al*. 1993). Any planar n-point configuration in three dimensions has a specific order type (Goodman & Pollack 1983), which is invariant to changes in viewpoint that do not intersect the 3D plane containing the configuration. It therefore shares essentially the same view invariance properties as affine and projective representations of planar point-sets. For non-planar point-sets the order type of the image will be invariant over *restricted* changes of viewpoint. The change of order type with viewpoint coincides with the accidental alignment of three points. The concept of order type therefore captures the idea of qualitative image structure as defined for the aspect graph of an object (Koenderink & van Doorn 1979).

(*b*) *Using order and incidence structure for indexing*

The use of qualitative geometric properties as a basis for recognition is generally based on perceptual grouping of low level primitives into higher order structures (Lowe 1984; Havaldar *et al*. 1996). This is often a difficult problem, mainly due to the fact that geometric features extracted from image data are often noisy and fragmented. The fact that order structure can be defined for general groups of features implies in principle that we can avoid the step of perceptual grouping. In practise, however, it turns out that the order types are unevenly distributed and we have to discard the most frequent ones since they are too ambiguous and give too many false matches. This can be seen as a soft way to introduce perceptual grouping, i.e. selecting only the most informative order types.

Order structure is a property of combinations of features, encoded by bracket expressions of coordinates in the same way as projective affine and incidence structure. The invariance properties of the order and incidence type of groups of features make them interesting for use as an index to a table of model features for a restricted view invariant recognition system. In Carlsson (1996), order and incidence type of configurations for line segments were considered. Using the endpoints of the segments, order and incidence structure can be defined in the same way as for point configurations (see figure 7).

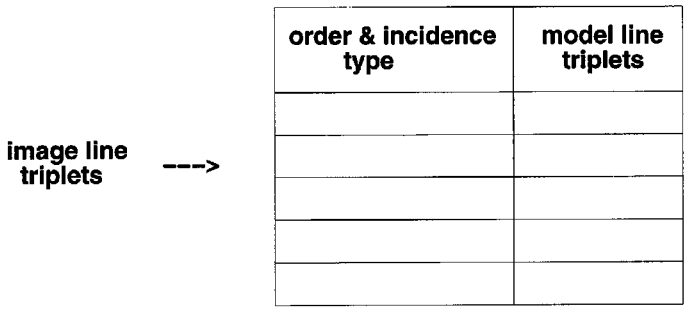

Figure 8. Order and incidence type indexing.

Recognition using order and incidence type for indexing proceeds in the same way as in geometric hashing (Lamdan *et al*. 1988) (see figure 8).

1. The line triplets of a certain geometric view model of an object are stored in a look-up table with the order and incidence type as the look-up key.
2. Line triplets are extracted from an image and the order and incidence type is computed.
3. Line triplets with most common order and incidence types are discarded due to their poor discrimination properties.
4. This order and incidence type is used to index the look-up table and a vote is given for each association of image segment and model segment in that specific table position.
5. A matching score between the image and a certain model is computed based on the total number of votes between image and model segments, normalized w.r.t. the total number of segments.

It should be noted that the order and incidence type of a triplet can be ambiguous due to imperfect data and the necessity to threshold when evaluating incidence relations. In these cases multiple weighted order and incidence type hypotheses are generated.

The resulting matching score for various views of a chair using a simple chair model are shown in figure 9. Also shown are the matching scores for various views of an unrelated object. The results verify that order and incidence structure have interesting restricted view invariant properties. Note that the views of the chair that are 'qualitatively similar' to the model view receive relatively higher matching scores. (It should be pointed out that the model has a dual 3D interpretation due to a necker reversal.) Different instances of chairs give similar responses of matching scores to this model (Carlsson 1996).

5. Summary and conclusions

I have discussed the view invariance properties of various geometric structure representations. For projective/affine structure it is possible to derive shape constraints that relate 3D and image structure in a view invariant way. These constraints can be

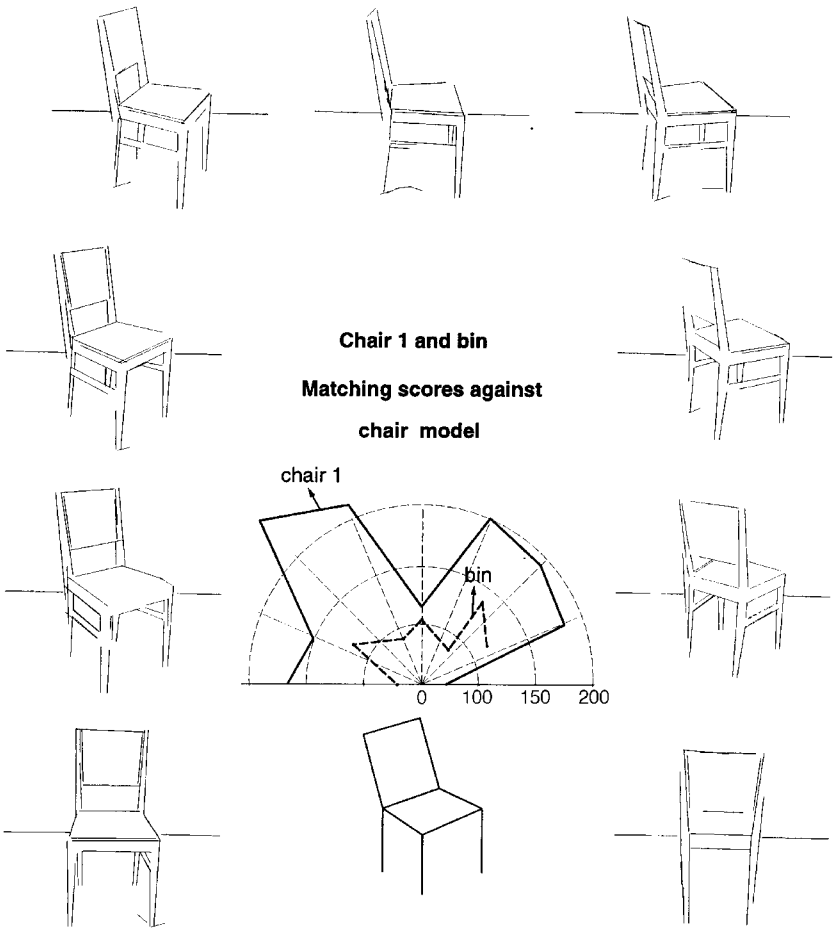

Figure 9. Matching scores for various views using indexing-based order and incidence structure.

used for verification of the presence of a specific object but not for direct indexing. In order to perform indexing-based recognition we have to restrict either the objects or the range of viewpoints. Restricted objects give rise to view invariant incidence structure in the image and order structure of geometric image features can be used for restricted view invariant recognition. All structure descriptors and constraints can be represented in a unified way using properties of brackets of image coordinates in arbitrary coordinate frames.

This work was supported by the ESPRIT-BRA project VIVA and by the Swedish Foundation for Strategic Research under the Centre for Autonomous Systems contract.

Affine coordinates

The affine transformation group in two dimensions can be expressed using non-homogeneous Cartesian coordinates as

$$\bar{p}' = \boldsymbol{A}\bar{p} + b, \tag{1}$$

where $\boldsymbol{A}$ and b are a general 2×2 matrix and 2-vector respectively.

The Cartesian coordinate vectors $\bar{p}_i$ of three points can be used to construct an *affine basis*. Any other vector can then be expressed as

$$\bar{p}_n - \bar{p}_3 = \bar{x}_n^a(\bar{p}_1 - \bar{p}_3) + \bar{y}_n^a(\bar{p}_2 - \bar{p}_3), \tag{2}$$

where $\bar{x}_n^a$, $\bar{y}_n^a$ are affine invariants, or *affine coordinates*. If we solve for these in terms of the Cartesian coordinates $\bar{p}_1 \ldots \bar{p}_n$, we get

$$\left.\begin{aligned} \bar{x}_n^a &= \frac{[\bar{p}_n - \bar{p}_3,\ \bar{p}_2 - \bar{p}_3]}{[\bar{p}_1 - \bar{p}_3,\ \bar{p}_2 - \bar{p}_3]} = \frac{\begin{bmatrix} \bar{p}_2 & \bar{p}_3 & \bar{p}_n \\ 1 & 1 & 1 \end{bmatrix}}{\begin{bmatrix} \bar{p}_1 & \bar{p}_2 & \bar{p}_3 \\ 1 & 1 & 1 \end{bmatrix}} = \frac{[2^*3^*n^*]}{[1^*2^*3^*]}, \\ \bar{y}_n^a &= \frac{[\bar{p}_n - \bar{p}_3,\ \bar{p}_1 - \bar{p}_3]}{[\bar{p}_1 - \bar{p}_3,\ \bar{p}_2 - \bar{p}_3]} = \begin{bmatrix} \bar{p}_1 & \bar{p}_3 & \bar{p}_n \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \bar{p}_1 & \bar{p}_2 & \bar{p}_3 \\ 1 & 1 & 1 \end{bmatrix} = \frac{[1^*3^*n^*]}{[1^*2^*3^*]}. \end{aligned}\right\} \tag{3}$$

The bracket $[\ldots]$ is used for the determinant function $\det(\ldots)$. The invariance over affine transformations follows from the fact that the determinant of the transformation matrix $\boldsymbol{A}$ can be factored out and cancelled. Note that we have used the notation i^* for the homogeneous coordinate p_i normalized with $w_i = 1$.

Projective coordinates

Projective transformations are general linear, non-singular transformations $\boldsymbol{T}$ of points in homogeneous coordinates:

$$p' = \boldsymbol{T}p. \tag{1}$$

The invariants of this transformation can be represented using *projective coordinates*. These can be defined using four points p_1, p_2, p_3, p_4 in projective 2-space as a basis in the following way. An arbitrary point p_n can be expressed in this basis as

$$p_n = x_n^p[2\ 3\ 4]p_1 - y_n^p[1\ 3\ 4]p_2 + w_n^p[1\ 2\ 4]p_3, \tag{2}$$

where the factors $[ijk]$ are used to make p_n homogeneous, i.e. arbitrary scalings of the homogeneous coordinates p_i imply the same scaling of p_n.

x_n^p, y_n^p, w_n^p are the homogeneous projective coordinates of the point p_n in the basis p_1, p_2, p_3, p_4. They can be expressed explicitly in terms of the Cartesian coordinates p_i by considering equation (2) as a linear system and solving it. We can directly compute the determinants

$$\left.\begin{aligned} [2\ 3\ n] &= x_n^p[2\ 3\ 4][1\ 2\ 3], \\ [1\ 3\ n] &= y_n^p[1\ 3\ 4][1\ 2\ 3], \\ [1\ 2\ n] &= w_n^p[1\ 2\ 4][1\ 2\ 3]. \end{aligned}\right\} \tag{3}$$

Disregarding the common factor $[1\ 2\ 3]$ we get

$$x_n^p = \frac{[2\ 3\ n]}{[2\ 3\ 4]}, \qquad y_n^p = \frac{[1\ 3\ n]}{[1\ 3\ 4]}, \qquad w_n^p = \frac{[1\ 2\ n]}{[1\ 2\ 4]}. \tag{4}$$

By taking ratios we can form the absolute projective coordinates

$$\left.\begin{aligned} \frac{x_n^p}{w_n^p} &= \frac{[2\ 3\ n][1\ 2\ 4]}{[2\ 3\ 4][1\ 2\ n]}, \\ \frac{y_n^p}{w_n^p} &= \frac{[1\ 3\ n][1\ 2\ 4]}{[1\ 3\ 4][1\ 2\ n]}. \end{aligned}\right\} \tag{5}$$

It is easy to verify that any linear transformation applied to the Cartesian coordinates p_i will leave these coordinates invariant since the determinant of the transformation can be factored out and cancelled.

$$[\boldsymbol{T}p_1\ \boldsymbol{T}p_2\ \boldsymbol{T}p_3] = [\boldsymbol{T}][p_1\ p_2\ p_3]. \quad (6)$$

Each invariant is the ratio of two polynomials in the Cartesian coordinates p_i. The polynomials are homogeneous in the same degree of each coordinate which means that their scaling will not affect the ratio.

References

Barnabei, M., Brini, A. & Rota, G.-C. 1985 On the exterior calculus of invariant theory. *J. Algebra* **96**, 120–160.

Biederman, I. 1985 Human image understanding: recent research and a theory. *CVGIP* **32**, 29–73.

Björner, A., Las Vergnas, M., Sturmfels, B., White, N. & Ziegler, G. 1993 *Oriented matroids. The encyclopedia of mathematics and its applications* (ed. C. G. Rota), vol. 46. Cambridge University Press.

Burns, J. B., Weiss, R. S. & Riseman, E. M. 1993 View variation of point-set and line-segment features. *IEEE Trans. Pattern Analysis Machine Intellig.* **15**, 51–68.

Carlsson, S. 1994 The double algebra: an effective tool for computing invariants in computer vision. *Applications of invariance in computer vision* (ed. J. L. Mundy, A. Zisserman & D. A. Forsyth), pp. 145–164. Lecture Notes in Computer Science, vol. 825. Springer.

Carlsson, S. 1995*a* View variation and linear invariants in 2D and 3D. *Tech. Rep.* KTH/NA/P-95/22-SE, Royal Institute of Technology.

Carlsson, S. 1995*b* Duality of reconstruction and positioning from projective views. In *Proc. IEEE Workshop on Representations of Visual Scenes, Cambridge, MA, June 1995.*

Carlsson, S. 1996 Combinatorial geometry for shape representation and indexing. *Object representation in computer vision II* (ed. J. Ponce & A. Zisserman), pp. 53–78. Lecture Notes in Computer Science, vol. 1144. Springer.

Carlsson, S. & Weinshall, D. 1998 Dual computation of projective shape and camera positions from multiple images. *Int. J. Computer Vision* **27**, 1–15.

Clemens, D. & Jacobs, D. 1991 Space and time bounds on model indexing. *IEEE Trans. Pattern Analysis Machine Intellig.* **13**, 1007–1018.

Faugeras, O. D. 1992 What can be seen in three dimensions with an uncalibrated stereo rig? *Proc. 2nd ECCV*, pp. 563–578. Lecture Notes in Computer Science, vol. 588. Springer.

Faugeras, O. D. & Mourrain, B. 1995 Algebraic and geometric properties of point correspondences between N images. *Proc. 5th ICCV*, pp. 951–956. Los Alamitos, CA: IEEE Computer Science Society Press.

Goodman, J. E. & Pollack, R. 1983 Multidimensional sorting. *SIAM J. Comput.* **12**, 484–507.

Hartley, R., Gupta, R. & Chang, T. 1992 Stereo from uncalibrated cameras. *Proc. Computer Vision and Pattern Recognition*, pp. 761–764.

Havaldar, P., Medioni, M. & Stein, F. 1996 Perceptual grouping for generic recognition. *Int. J. Computer Vision* **20**, 59–80.

Hodge, W. V. & Pedoe, D. 1947 *Methods of algebraic geometry*, vol. 1. Cambridge University Press.

Koenderink, J. J. & van Doorn, A. J. 1979 The internal representation of solid shape with respect to vision. *Biol. Cybernetics* **32**, 211–216.

Koenderink, J. J. & van Doorn, A. J. 1991 Affine structure from motion. *J. Optic. Soc. Am.* A **2**, 377–385.

Lamdan, Y., Schwartz, J. T. & Wolfson, H. J. 1988 Object recognition by affine invariant matching. *Proc. Computer Vision and Pattern Recognition*, pp. 335–344. Los Alamitos, CA: IEEE Computer Science Society Press.

Lowe, D. G. 1984 *Perceptual organization and visual recognition*. Boston: Kluwer.

Marr, D. & Nishihara, K. 1981 Representation and recognition of the spatial organization of three dimensional shapes. *Proc. R. Soc. Lond.* B **200**, 269–294.

Moses, Y. & Ullman, S. 1992 Limitations of non model-based recognition schemes. *Proc. 2nd European Conf. on Computer Vision*, pp. 820–828. Lecture Notes in Computer Science, vol. 888. Springer.

Quan, L. 1994 Invariants of 6 points from 3 uncalibrated images. *Proc. 3rd European Conf. on Computer Vision*, pp. 459–470. Lecture Notes in Computer Science, vol. 801. Springer.

Rothwell, C. A., Forsyth, D. A., Zisserman, A. P. & Mundy, J. L. 1993 Extracting projective structure from single perspective views of 3D point sets. *Proc. of 4th Int. Conf. on Computer Vision*, pp. 573–582. Los Alamitos, CA: IEEE Computer Science Society Press.

Sparr, G. 1991 Projective invariants for affine shapes of point configurations. In *Applications of invariance in computer vision, DARPA-ESPRIT workshop* (ed. J. L. Mundy & A. Zisserman). Reykjavik, Iceland.

Sparr, G. 1992 Depth computations from polyhedral images. *Image Vision Computing* **10**, 683–688.

Sugihara, K. 1986 *Machine interpretation of line drawings*. Cambridge, MA: MIT Press.

Ullman, S. & Basri, R. 1991 Recognition by linear combination of models. *IEEE Trans. Pattern Analysis Machine Intellig.* **13**, 992–1007.

Weinshall, D. 1993 Model-based invariants for 3D vision. *Int. J. Computer Vision*, **10**, 27–42.

Weinshall, D., Werman, M. & Shashua, A. 1995 Shape tensors for efficient and learnable indexing. *Proc. IEEE Workshop on Representations of Visual Scenes*. Los Alamitos, CA: IEEE Computer Science Society Press.

Discussion

A. Zisserman (*Department of Engineering Science, University of Oxford, UK*). The 3D symmetric point configuration Dr Carlsson described was a 'butterfly'. It is known that a projective invariant can be measured for this configuration from a single perspective image. Are there any examples of non-butterfly bilateral symmetric point configurations for which an invariant can be measured from a single image?

S. Carlsson. Yes, for a configuration of six points with four in one plane, defining a plane of symmetry, and with the other two symmetrically placed on each side. For example, points placed at the the vertices of a regular octahedron. It is possible to write down an invariant image constraint for this structure.

J. L. Mundy (*GE Corporate Research and Development, Niskayuna, NY, USA*). Stefan, you were somewhat critical of Jacobs's idea of storing the affine observations in a look-up table because of the cost of memory. However, I would argue the following: first, I believe that this idea can be generalized quite a bit to where we store affine information about more arbitrary surfaces than just point-sets and also that we can recover the viewpoint from the stored observations. Then as long as two objects don't look the same from the same viewpoint we will have successful indexing. It seems to me actually a very small price to pay for the memory and storage to achieve that. Secondly, if you look today, people are storing image manifolds of hundreds of images and using that essentially as a database, so storing the affine point information pales into insignificance in cost by comparison with that.

S. Carlsson. Jacobs notes that you run into problems with too many point combinations unless you perform some kind of perceptual grouping first. I suppose it is not the cost of memory that is the problem but rather the fact that there will be too many false matches.

T. Kanade (*Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA*). Is more than one concept, like stability or robustness, of the invariants needed? In the sense that in reality, the detection may or may not be stably robust. Once Katsushi Ikeuchi and I did some work in the area of synthetic aperture radar image detection. We didn't systematically pursue it, as Dr Carlsson has, but that kind of concept seems to be useful and I wonder whether there has been any work to study that type of idea.

S. Carlsson. As far as I know there has been no systematic evaluation of the robustness of incidence constraints. There will be many situations where the presence of an incidence is uncertain. In those cases it is necessary to generate multiple hypotheses, i.e. allowing both for the presence and non-presence of the incidence.

D. Foster (*Aston University, UK*). The notion of an intermediate class of representations, somewhere between metric and incidence, seems particularly appropriate for describing human object and pattern recognition. Order relations have been used in modelling the approximate invariance of human recognition to small global transformations, i.e. limited translations, dilations, and rotations within the frontoparallel plane (e.g. $< 20°$), as well as to large global transformations, such as reflections about the vertical or horizontal midline, or 180° rotations in the plane (Foster 1991). They can also be used to explain the efficient discrimination of patterns subjected to local reflections (Hummel & Stankiewicz 1996), and failures in recognition under global 90° planar rotations. These order relations do, however, depend on the marked anisotropy of human vision, in which the vertical and horizontal midlines and the point of gaze have a special status (Attneave & Curlee 1977).

S. Carlsson. The experiments reported in Carlsson (1996) actually used a partly calibrated camera aligned parallel to the ground which permitted the definition of a vertical direction and ordering relations relative to this. This turned out to be very effective in terms of reducing ambiguity in the feature matching and thereby increasing overall performance.

A. Fitzgibbon (*Department of Engineering, University of Oxford, UK*). There is a sense in which view-based vision explores the aspect graph of the point-set and has to store some sort of invariant per node in the aspect graph. Now, even at full complexity, the aspect graph is only polynomial in k, whereas Dr Carlsson is suggesting that the number of order relations may be exponential in k. Is there not some similarity between searching the aspect graph and generating all these order relations?

S. Carlsson. The problems are related but not identical. The aspect graph problem for $k-1$ points is essentially the problem of finding various order types for k points given that the first $k-1$ points are fixed, i.e. a restricted version of the order type problem for k points. The complexity of the k-point aspect graph is therefore lower than for the k-point order type.

W. Triggs (*INRIA, France*). Intuiting the way that I recognize things, it seems to me that besides order structure I also use a qualitative idea of how far things are from each other, without really going to strict metric invariant type structures. Would Dr Carlsson like to speculate on this?

S. Carlsson. There is some kind of region between strict order structure and metric that we don't really know how to characterize. What do we mean by large and things like that?

W. Triggs. One hesitates to say the word fuzzy in a meeting like this ...

S. Carlsson. Size concepts are in general relative. When we say large we essentially always mean large relative to something else, and that means that we're talking about order structure.

Additional references

Attneave, F. & Curlee, T. E. 1977 Cartesian organization in the immediate reproduction of spatial patterns. *Bull. Psychonomic Soc.* **10**, 469–470.

Foster, D. H. 1991 Operating on spatial relations. In *Pattern recognition by man and machine* (ed. R. J. Watt), pp. 50–68. Houndmills, Basingstoke: Macmillan.

Hummel, J. E. & Stankiewicz, B. J. 1996 Categorical relations in shape perception. *Spatial Vision* **10**, 201–236.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES