Invariance from the Euclidean geometer's perspective

Luc J Van Gool, Theo Moons, Eric Pauwels

ESAT-MI2, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium

Johan Wagemans

Laboratory of Experimental Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium

Received 5 July 1993, in revised form 8 November 1993

Abstract. It is remarkable how well the human visual system can cope with changing viewpoints when it comes to recognising shapes. The state of the art in machine vision is still quite remote from solving such tasks. Nevertheless, a surge in invariance-based research has led to the development of methods for solving recognition problems still considered hard until recently. A nonmathematical account explains the basic philosophy and trade-offs underlying this strand of research. The principles are explained for the relatively simple case of planar-object recognition under arbitrary viewpoints. Well-known Euclidean concepts form the basis of invariance in this case. Introducing constraints in addition to that of planarity may further simplify the invariants. On the other hand, there are problems for which no invariants exist.

1 Introduction

Vision systems—biological or artificial—must be able to cope with the major variations in the way objects look in their two-dimensional (2-D) eye or camera projections. There are properties, however, that do not change over viewpoint, so-called invariants. The recent upsurge of interest in invariance has partly to do with the central role it plays in Gibson's (1950) theory of perception as the direct extraction of information.

Much of the early invariance-based work in perception research undoubtedly found its roots in the 19th century mathematical endeavour to look systematically for properties that remain invariant under different types of vision-related deformations. However, many of the early mathematical results would be overly sensitive to noise (eg requiring unrealistically high derivatives) or too restricted (eg only useful for collinear points). A gamut of recent approaches in computer vision allow sufficient efficiency and robustness but seem still largely unknown to the perception community.

One can distinguish two types of psychological research questions. One type is to find out whether or not a particular percept is changed by a particular transformation, whereas the other is aimed at revealing the kind of information that is used by the human visual system in the cases where invariance would have been obtained. Examples of the former type of studies are abundantly present in the recent literature (eg Allik et al 1991; Biederman and Cooper 1991; Nazir and O'Regan 1990), whereas examples of the latter type of studies are particularly scarce (Cutting 1986; Niall 1992; Niall and Macnamara 1990). This might be explained by the fact that the second type of research is more demanding as to the mathematical background needed. It is this kind of background that we hope to provide.

After invariant-based, viewpoint-independent recognition has been pinpointed as the main thrust of this paper, some further choices have to be made in order to select both an interesting, nontrivial test case for the use of invariance and yet to make the scope of the paper neither too wide nor too technical. Therefore the practical discussion will be restricted to the recognition of planar-object contours.

The paper is organised as follows. In section 2 we discuss the deformations induced by projection when a planar shape is viewed from different directions. These deformations are identified in terms of transformation groups. Why the group structure is such an important one in the study of invariance is explained in section 3. Although the deformations might be very pronounced, the properties that remain invariant under them can often be expressed in terms of simple Euclidean invariants. Examples are given in section 4. The strand of research that led to these results is then given a nonrigorous theoretical underpinning in section 5. In this section we also discuss the 'counting argument', a rule for estimating the required amount of information for calculating invariants. In section 6 we relate the invariance and non-accidentalness paradigms, but also illustrate the limitations of invariance. We conclude in section 7 by trying to sketch the implications of the invariance approach for visual recognition strategies.

2 The effects of projection

Objects—here assumed planar in order not to complicate further the discussion—are seriously deformed in the way they look when viewed obliquely. In this section we investigate how these deformations come about and what their exact nature is.

2.1 A projection model

In a way that depends on the relative positions of objects and the observer, deformations under projection will be more or less severe. First consider the perspective-projection model, as explained in figure 1. The image plane was placed between the centre of projection (physically the centre of the lens) and the object. This, of course, is not the case in reality. In fact, the image plane (eg, the fovea) lies behind the centre of projection. Mathematically speaking, it will make no difference, except for the elimination of sign reversals. Suppose we look at the point with three-dimensional (3-D) coordinates (X, Y, Z) . We see that the coordinates (x, y) of the image projection are found as

$$x = f \frac{X}{Z} \quad \text{and} \quad y = f \frac{Y}{Z},$$

where f is the focal length.

We will consider these equations as the precise model of projection, although in reality they are but approximations.

Figure 1. Perspective projection: f is the focal length, (X, Y, Z) are the coordinates of a scene point in the world coordinate frame, and (x, y) are the coordinates of the projection of the point on the image plane (eg the fovea). In most of the discussion the contour is supposed to be planar, as shown in the figure.

If we look at the above equations of perspective projection, some special cases can be distinguished. First, if Z is constant for all points of an object, then the projection model simplifies to $x = kX$ and $y = kY$, with k a constant equal to f/Z (pseudo-perspective or pseudoorthographic projection). The effect of projection is a simple scaling of the X and Y coordinates. The object then lies in a plane orthogonal to the viewing direction. Similar simplifications can be made when the range of Z -values of the object points is small compared with Z , eg when objects are viewed from a distance that is large compared with the object size. In that case f/Z will remain almost constant and the projection model can be simplified in the same way. In what follows we will consider in more detail the deformations these projection models imply.

2.2 Deformation of planar contours under projection

In this section we discuss the different types of deformations to be expected under projection for planar contours. We consider 3 cases.

- (i) The objects are viewed from a perpendicular direction, can translate freely in any direction, but only rotate about an axis parallel to the viewing direction (ie the direction normal to the image plane).
- (ii) The objects can be viewed from any direction but should be at a sufficient distance from the viewer so as to make the differences in depth (Z -values) small relative to the distance to the viewer.
- (iii) The objects can be viewed from any direction and can be placed at arbitrary distances from the viewer.

It is clear that the viewer-object configurations become increasingly arbitrary as one goes down this list. As will be shown, the corresponding deformations are of increasing complexity.

Suppose that the contour point (X, Y, Z) projects to the image point (x, y) . Furthermore, we denote the image projection of this point after the contour has undergone a 3-D motion—and the point assumed new 3-D coordinates (X', Y', Z') —as (x', y') . We are interested in the transformation bringing the coordinates (x, y) into correspondence with (x', y') , ie in the type of transformation describing the deformations between the two views.

2.2.1 Case 1. The relationship between the 3-D object motion and the image deformation of the contours is particularly easy to derive in this case. Since the planes containing the two contours, ie the contour before and the contour after the motion, are both parallel to the image plane, points of each of the contours will have a constant Z -value. Hence, the projection equation simplify to a scaling:

$$\begin{cases} x = kX, \\ y = kY, \end{cases} \quad \text{and} \quad \begin{cases} x' = k'X', \\ y' = k'Y', \end{cases} \quad (1)$$

with $k = f/Z$ and $k' = f/Z'$ two constant scale factors. One finds that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{k'}{k} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + k' \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

where θ is the angle over which the contour rotated and T_1 and T_2 are the X and Y components of the translation. Hence, the image projection of the contour undergoes a similarity transformation. Different views are identical up to a combined scaling, rotation, and translation in the image plane. If the object does not come closer to or move away from the camera, then $Z' = Z$, hence $k' = k$, and the transformation becomes a Euclidean motion, ie a combination of only rotation and translation.

2.2.2 *Case 2.* In this case the curve can undergo arbitrary 3-D rotations and translations, but the projection model is simplified to a scaling:

$$\begin{cases} x = kX, \\ y = kY, \end{cases} \quad \text{and} \quad \begin{cases} x' = k'X', \\ y' = k'Y', \end{cases} \quad (2)$$

where in general $k \neq k'$, since different distances to the viewer may result in different scaling factors $k = f/Z$. Note that the projection model is an approximation in this case, whereas it was exact in case 1. Taking into account the 3-D motion (which now also includes general rotations) we arrive at a relationship of the type

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

where the a_{ij} and b_i are arbitrary real numbers. Such a relation is called a 2-D affine transformation.

2.2.3 *Case 3.* Here it is assumed that the curve undergoes an arbitrary 3-D motion (rotation plus translation) as under case 2, but the full perspective-projection model is used. The difference with case 2 is that the object may be large compared with its distance to the viewer.

$$\begin{cases} x = f \frac{X}{Z}, \\ y = f \frac{Y}{Z}, \end{cases} \quad \text{and} \quad \begin{cases} x' = f \frac{X'}{Z'}, \\ y' = f \frac{Y'}{Z'}. \end{cases}$$

We find a transformation of the type

$$x' = f \frac{p_{11}x + p_{12}y + p_{13}f}{p_{31}x + p_{32}y + p_{33}f},$$

$$y' = f \frac{p_{21}x + p_{22}y + p_{23}f}{p_{31}x + p_{32}y + p_{33}f},$$

with p_{ij} real numbers and normally referred to as a 2-D projective transformation.

2.2.4 *Remarks.* As already pointed out in section 2.2, the complexity of the relevant transformations goes up with the generality of the viewpoint. In order to give a more intuitive idea of the difference between these cases, figure 2 shows typical deformations for a square under the three types of transformation. Under similarities (case 1), angles as well as length ratios and area ratios remain unchanged. Of these, only ratios of areas survive an affine transformation (case 2). Also notice that parallelism is an

Figure 2. Effects of projection of the three types of transformation. The observed object is a square. As can be seen, similarity transformations do not alter the shape (but may alter the scale, the position, and the orientation). Affine transformations lead to deformations, but parallelism is preserved. Projective transformations lead to even stronger deformation.

invariant property in both cases. For the projective case (case 3), none of the aforementioned properties are invariant. This already illustrates the relative difficulty in finding invariants.

3 Invariance

Invariants were introduced in very loose terms as properties, measured in an image, that do not change when the viewpoint changes. More technically, invariants are properties that remain unchanged under certain sets of transformations, in particular under transformation groups. The similarity transformations, the affine transformations, and the projective transformations obtained in section 2 all form groups. What groups are and why they are important in this context are questions we intend to answer in this section.

3.1 *From images to invariants*

The problem of analysing the scene, and recognising the objects in particular, can be organised in the following hierarchy of steps. If one looks at different views of a scene, the same objects will typically take on different projected shapes. From these projections a series of measurements can be made. A natural choice are the coordinates and derivatives of the contour points. Other measurements such as the area enclosed by the contour and the coordinates of its centre of gravity could in principle equally well be used (though of less practical importance because of problems of occlusion). In the different views the measurements will typically be different. The challenge now is to combine such variable measurements so as to produce invariant properties. As an example, although points change their positions in the image (and consequently the measured coordinates) when the contour translates parallel to the image plane, the distance between the points is a function of the coordinates which is invariant.

Whatever kind of measurements one chooses to work with, the exact way in which they change is a first issue to study carefully. In case of the coordinate measurements the changes (transformations) they undergo have already been identified in section 2. In order to compute the simplest invariants possible, however, one has to identify the smallest transformation group that includes all the relevant transformations. In order for a set of transformations to form a group, it has to comply with a number of conditions. First of all, if one applies one transformation of the set after another, the overall transformation, ie the net result, should also belong to the set. We can say that the group is closed under composition. If one takes the group of similarities as an example, one sees that when a first combination of rotation, translation, and scaling is applied, followed by a second such transformation, the resulting transformation can be decomposed as a combination of a single rotation, translation, and scaling. The result is a new similarity. Second, after application of a transformation, it should always be possible to 'undo' it by using a transformation from the same set. We can say that all elements of the set should have an inverse element within the set. For instance, each combination of scaling, rotation, and translation can be undone by a transformation of the same type. So each similarity has its own inverse similarity. Third, the set should always include an element that basically does nothing, the so-called identity element, and in our case keeps the contour unchanged. In the case of similarities, the element formed by a scaling with factor 1, a rotation with zero angle, and a translation with zero components clearly does not change the shape, and it is the identity element of the group of similarities. There is a fourth requirement, that of associativity, which we do not touch upon here, since it links up less directly with the invariance idea. In fact, invariants for algebraic structures other than groups are possible. However, it is for groups that the mathematical theory for their extraction is well developed.

3.2 Invariants and transformation groups

Suppose we find some property that remains invariant under a set of transformations. Application of a first transformation from the set will therefore not change the property. Neither will the subsequent application of a second transformation from that set. Hence, the composition of the two transformations preserve the invariant as well. We conclude that when a property remains invariant under a set of transformations, it necessarily remains invariant under all the compositions of these transformations. Moreover, if a property remains invariant under a transformation, it must be invariant under the inverse transformation as well. And, obviously, any property is invariant under a transformation that does not induce any change in the image, ie the identity. Making sure that one includes all compositions, inverses, as well as the identity element to the set of transformations preserving the invariance of a property, brings one close to the group concept.

In general, however, one wants to include as few nonintended transformations as possible in the set of transformations leaving the property invariant. Hence, one searches for the *smallest group* that includes all the relevant transformations. For cases 2 and 3 the affine and projective transformations, respectively, were highlighted as relevant. However, upon replacing the 3-D Euclidean motions of the objects by 3-D affine transformations, one would still find the same type of image transformations (see eg Van Gool et al 1993a). This has some serious consequences if we work with properties that remain invariant under affine and projective transformations. The original shapes can be skewed and yet they yield exactly the same values for the invariant properties. Therefore, properties that remain invariant under the 2-D affine or projective transformations can never fully describe the original shape of the planar object. So the question is, do we have a choice? Are there smaller groups that take the specific nature of the 3-D Euclidean motions into account and that exclude the transformations observed in the image upon 3-D affine transformations of the objects? Unfortunately, the answer is negative. The 2-D affine and projective transformations are the smallest groups we can find that contain all the relevant transformations for these cases. In section 5 it will be argued that there is another compelling reason for working with the smallest possible group. It will become clear that the complexity of the invariants directly depends on the complexity of the corresponding transformation group.

4 Invariants for planar contours

Although viewpoint-independent recognition even of planar objects has long been considered a very hard problem in computer vision, simple combinations of Euclidean concepts such as curvatures, lengths, and areas yield invariants. These combinations were called 'semidifferential invariants' because they usually include coordinate derivatives. Rather than going through the mathematics to obtain such invariants, we simply describe some of these building blocks and give examples of invariant combinations. For more extensive accounts the interested reader is referred to the literature (Brill et al 1992; Van Gool et al 1990, 1991, 1992b).

4.1 Euclidean building blocks

Suppose we have detected a contour in the image. Each point of the contour is characterised by its image coordinates (x, y) . If we were to traverse the contour, the points would be passed in a fixed order. A mathematical way of expressing the order in which contour points figure along the contour is through the introduction of a parameter t . As we traverse the contour, t will monotonically increase or decrease, according to the direction of traversal. Each point corresponds to a particular value of t . The starting point typically has $t = 0$. Thus, a contour is represented by a set of points $(x(t), y(t))$ where the parameter t is a real number. A special choice of the

parameter t is the 'natural arc-length' or, synonymously, 'Euclidean arc-length', and is usually denoted by s . This special parameter simply is the length from the starting point to the contour point, measured along the contour. We introduce the shorthand notation $\mathbf{x}(t)$ for $(x(t), y(t))^T$ with the superscript T indicating the transpose. Furthermore, $\mathbf{x}$ stands for an arbitrary contour point, and $\mathbf{x}_i$ will denote a specific 'reference' point, usually a point of some special nature such as an inflexion. Reference points are not necessarily points on the contour, but they must remain in a fixed position with respect to the contour and must lie in the plane of the contour. For instance, if a planar workpiece is lying on a table and does not change its position between different views, the corners of the table are valid reference points to be used for identifying the object from one view to the next. The distinction between the 'arbitrary point' and the 'reference point' is not really important at this stage. Last, $\mathbf{x}^{(i)}$ stands for the i th derivative $((d^i x/dt^i), (d^i y/dt^i)^T)$ of the coordinates $(x, y)^T$ with respect to the contour parameter t . These values will actually depend on the choice of the parameter t . The use of such derivatives therefore adds another dimension to the problem: the choice of the parameter should be left free. This is because it will be hardly possible to choose the same parameterisation in every view, since this would require knowing point correspondences beforehand. Rather, we would like to select properties that also remain invariant under changes in the choice of t as well as under the influence of changing viewpoints.

As will be illustrated later, such invariants can be constructed by using expressions of the following determinant types:

$$|\mathbf{x} - \mathbf{x}_i \quad \mathbf{x} - \mathbf{x}_j|,$$

$$|\mathbf{x} - \mathbf{x}_i \quad \mathbf{x}^{(1)}|,$$

$$|\mathbf{x}^{(1)} \quad \mathbf{x}^{(2)}|,$$

where

$$|\mathbf{x} - \mathbf{x}_i \quad \mathbf{x} - \mathbf{x}_j| = \begin{vmatrix} x - x_i & x - x_j \\ y - y_i & y - y_j \end{vmatrix} = (x - x_i)(y - y_j) - (y - y_i)(x - x_j), \text{ etc.}$$

Since the choice of the parameter in the envisaged invariants should be arbitrary anyway, we can just as well interpret the resulting expressions by picking Euclidean

Figure 3. If the Euclidean arc-length is used for their evaluation, the building blocks underlying the construction of projective and (some) affine semidifferential invariants yield simple Euclidean interpretations.

arc-length, as illustrated in figure 3. The last expression in the above list then yields Euclidean curvature. The second expression can be given an equally simple Euclidean interpretation in that case: it is the distance of the reference point x_i from the line tangent to the contour in the contour point x . The first expression has a simple Euclidean interpretation irrespective of the particular choice of the parameter: it is the area of the parallelogram defined by taking the points x , x_i , and x_j as its vertices. An additional building block for affine (but not for projective) invariants is

$$|x^{(1)} x_i^{(1)}|,$$

which, again with Euclidean arc-length as a choice for the parameter, amounts to the sine of the angle subtended between the two tangential lines at the points x and x_i .

4.2 Invariant combinations

With these building blocks, it is possible to construct invariants for the description of planar shapes, both under pseudo-orthographic and under perspective projection. The following lists contain some examples, but are by no means exhaustive enumerations. Moreover, it is important to note that invariant parameters, ie a kind of invariant 'length' measures along the contour, can be obtained in much the same way.

Affine invariants (for use with pseudoorthographic projection):

$$\frac{|x - x_1 \ x - x_2|}{|x - x_1 \ x - x_3|},$$

$$\frac{|x^{(1)} \ x - x_1|}{|x^{(1)} \ x - x_2|},$$

$$\frac{|x - x_1 \ x_1^{(1)2} |x^{(1)} \ x^{(2)}|}{|x - x_1 \ x^{(1)}| |x^{(1)} \ x_1^{(1)2}|}.$$

Notice that using invariants of the first type amounts to the alignment strategy (Huttenlocher and Ullman 1987; Ullman 1989).

Projective invariants (for use with perspective projection):

$$\frac{|x - x_1 \ x - x_4| |x_2 - x_3 \ x_2 - x_4|}{|x - x_2 \ x - x_4| |x_1 - x_3 \ x_1 - x_4|},$$

$$\frac{|x - x_1 \ x^{(1)}| |x_1 - x_2 \ x_1^{(1)}| |x - x_2 \ x_2^{(1)}|}{|x - x_2 \ x^{(1)}| |x - x_1 \ x_1^{(1)}| |x_1 - x_2 \ x_2^{(1)}|},$$

$$\frac{|x^{(1)} \ x^{(2)}| |x_1^{(1)} \ x - x_1|^3}{|x_1^{(1)} \ x_1^{(2)}| |x^{(1)} \ x - x_1|^3}.$$

The first of these invariants is sometimes referred to as the five-point cross ratio and is the direct extension of the alignment strategy for the projective case (Efimov 1981; Rothwell et al 1992a).

As an example of the Euclidean interpretation of these expressions, we consider the third projective invariant,

$$\frac{|x^{(1)} \ x^{(2)}| |x - x_1 \ x_1^{(1)3}|}{|x_1^{(1)} \ x_1^{(2)}| |x - x_1 \ x^{(1)3}|}. \quad (3)$$

This invariant will be used repeatedly as an example case since it is of particular interest in more than one respect. As can be seen, this expression contains the ratio of two curvatures, calculated at the points x and x_1 . It also contains the ratio of two distances to tangent lines, raised to the third power, as follows from the interpretations of the

different chunks outlined earlier. The resulting overall interpretation is illustrated in figure 4, where the curvature is represented as the reciprocal of the radius of the osculating circle.

Figure 4. Affine and projective invariants can often be given a simple Euclidean interpretation. This is illustrated for the projective invariant of equation (3).

4.3 Complete descriptions

The completeness of a description is an issue not touched upon yet. A description is said to be complete if it yields an unambiguous representation of a shape, up to the kind of transformations the description is supposed to be invariant under. It is, for example, clear that an affine invariant description will be the same for a contour and an affine distortion of that contour. But if we carefully pick our affine invariant description such that it is complete, it will be different for any nonaffine distortion. A complete description of a planar shape requires two independent invariants. In particular, a complete affine description requires two independent affine invariants. Invariants are said to be independent if none of them can be calculated from the others (see Van Gool et al 1993a for further information).

5 The counting argument

In the invariant expressions of section 4, there is some trade-off to be observed. If no derivatives but only point coordinates are used, we need several points. In particular, four points are used in the affine invariant without derivatives and even five for the corresponding projective invariant (the first cases in the lists). It can actually be proven that no invariants with fewer points in generic positions can be found if only coordinates and no derivatives are to be used. The introduction of derivatives—ie the use of semidifferential invariants—seems to lower the number of reference points needed, as shown by the other examples. In fact differential invariants are found if this exchange is carried out to its extreme. Then only coordinates and derivatives at a single point are used.

One should realise that the introduction of reference points brings in a correspondence problem, which can be reduced in two ways. The first is to choose as reference points special points that can easily be found in different projections. Good candidates therefore are points characterised by some invariant aspect—so invariance pops up again. Inflexions or tangent points obtained from bitangent lines are good examples for both the affine and the projective case (see section 6). But, even then, there will typically be a lot of ambiguity when trying to figure out which inflexions or bitangent points correspond to similar points in another view. The search can be reduced further if the correspondence of a single point leads to the correspondence of

the other points, by using some invariant construction. For instance, knowing one reference point, one might choose a sequence of subsequent, neighbouring inflexions as the other reference points. Search is basically reduced to searching the first inflexion. Such schemes have been applied successfully in computer vision (eg Rothwell et al 1992a; Van Diest et al 1994) and are akin to the codon schemes propounded in perception research (Hoffman and Richards 1984).

Constructions such as the foregoing ones will not always be possible, however, or might even lead to reference points lying in some accidental configuration that precludes their use (eg three subsequent inflexions being collinear). The introduction of derivatives is a viable alternative (Brill et al 1992; Van Gool et al 1990, 1991, 1992b). If the number of reference points is reduced, other invariant constructions can be considered as illustrated by the list of invariants given in section 4.2. Explaining in detail why the use of derivatives has this effect would lead us too far, but the overall principle is rather simple: *in order to find invariants, the number of measurements should surpass the number of independent group parameters.*

Using not only coordinates but also their derivatives simply introduces more measurements without having to introduce new points. Every pair of point coordinates as well as every derivative thereof adds two measurements. Before we dwell any further on the effect of the derivatives, the above 'counting argument' should first be considered a little further.

First, the number of independent group parameters is considered. In the case of pseudoorthographic projection we found the affine transformations as the relevant transformation group. A 2-D affine transformation is characterised by six parameters, each of which has an independent effect. Projective transformations, on the other hand, have nine parameters, but their effect will not differ if they are all multiplied by a common factor. Hence, there are only eight *independent* parameters in that case. It is also useful to illustrate the above counting rule on a simpler group. If a planar object is looked at from a perpendicular direction and it can only rotate and translate in its plane, the relevant group is that of the 2-D Euclidean motions. This group has three independent parameters: one rotation angle and two translation components. The counting rule prescribes the use of at least four measurements, eg the coordinates of two points. It is well known that this choice allows the use of the distance between the points as a Euclidean invariant. For the affine group with its six independent parameters one should thus combine at least four points, and for the eight-parameter projective group at least five (if no derivatives are used). These predictions tally with the invariants given earlier.

To return to the introduction of derivatives, the counting argument should be adapted slightly for reasons beyond the scope of the paper (Van Gool et al 1992b). Let it suffice to mention that reparameterisation can also be cast in the form of a group action, thereby introducing other transformations and their parameters. *In order to find invariants, the number of measurements should surpass the number of independent group parameters incremented with the highest order of derivative taken in each point.*

As an example, trying to find an invariant for a curve under affine transformations and using information in only two points, we will have to introduce coordinate derivatives, since two points yield only four measurements, which is insufficient to surpass the minimum of six affine group parameters. Using up to the second derivative in the first, but only the first derivative in the second point, we have ten measurements (six + four), exceeding the incremented number of nine parameters (six + two + one) by one, and find, as an affine invariant:

$$\frac{|\mathbf{x} - \mathbf{x}_1| \, |\mathbf{x}_1^{(1)}|^2 \, |\mathbf{x}^{(1)} \, \mathbf{x}^{(2)}|}{|\mathbf{x} - \mathbf{x}_1| \, |\mathbf{x}^{(1)}| \, |\mathbf{x}_1^{(1)}| \, |\mathbf{x}_1^{(1)}|^2}.$$

This single invariant does not allow us to build a complete affine description. The selected measurements do not allow the extraction of an independent, second affine invariant, as suggested by a similar rule: *the difference between the number of measurements and the incremented number of independent group parameters yields the number of independent invariants*.

As a simpler example, consider rotations and translations in the plane. All areas are invariants. If we take three points, the three distances between any pair of these points is invariant, as well as the area of the triangle they form. However, these invariants are not independent, as predicted by the aforementioned rule: the six measurements minus the three transformation parameters leaves us with only three independent invariants. Indeed, knowing the three lengths of the sides of a triangle fixes the shape of the triangle and therefore its area.

In order to find a second, independent invariant in the affine example, we have to introduce further measurements. The introduction of the second-order derivative in the second point suffices. We find, as an additional, independent affine invariant:

$$\frac{|\mathbf{x}^{(1)} \quad \mathbf{x}^{(2)}| |\mathbf{x}_1^{(1)} \quad \mathbf{x} - \mathbf{x}_1|^3}{|\mathbf{x}_1^{(1)} \quad \mathbf{x}_1^{(2)}| |\mathbf{x}^{(1)} \quad \mathbf{x} - \mathbf{x}_1|^3}.$$

The reader might have noticed something strange at this point. The last invariant has been introduced before as a *projective* invariant. Using the counting argument for the projective case leads to the alarming conclusion that the existence of this projective invariant contradicts it, since the number of measurements (twelve) does not surpass but equals the incremented number of parameters (eight + two + two). Nevertheless, the above expression really is a projective invariant. This example shows the care that should be taken in using the counting argument. It is but a simplified formulation of some mathematical rule that always holds. Usually it works, but it must not be applied blindly. Rather, it serves as a conservative estimate.

6 Invariance, nonaccidental properties, and accidental successes

In the foregoing account, no special constraints were imposed except for the structures being planar. Affine invariants can, for example, be calculated from any four coplanar points. Imposing further constraints—often related to so-called nonaccidental properties (Biederman 1987; Lowe 1987; Wagemans 1992, 1993)—can simplify the situation. For instance, if we assume that three points are collinear, the ratios of lengths between the point pairs are affine invariants. Thus, by imposing this additional restriction of collinearity, invariance can be obtained with fewer points. Therefore, not only are nonaccidental properties interesting invariant properties in their own right, they also often support simpler, quantitative analysis.

As a further example, we consider again the invariant formula (3). In the affine situation, a special case arises when the two points have parallel tangent lines. Then the two distances (the determinants raised to the third power) are equal and the invariant can be simplified to the ratio of curvatures:

$$\frac{|\mathbf{x}_1^{(1)} \quad \mathbf{x}_1^{(2)}|}{|\mathbf{x}^{(1)} \quad \mathbf{x}^{(2)}|}$$

for such points. This simplification holds for arbitrary viewpoints, since parallelism is an affine invariant property. Under projective transformations this expression is no longer invariant, since parallelism is not preserved. However, it can be shown that if two curves are tangent in the same point, then the ratio of the curvatures in that point is invariant under projective transformations. In this situation, where the two tangent

lines are reduced to the same line in that single point, the invariance of tangency under projective transformations is used.

The previous examples all use restrictions on the scene configurations. One might just as well consider constrained relative motions of the objects and the viewer. As an example, if we assume that the planar curve lies in a plane that is purely slanted (not tilted) with respect to the observer and that it undergoes pure 3-D translations in between views (or, equivalently, that the observer only translates),

$$\frac{(y_2 - y_1)(y_4 - y_3)}{(y_3 - y_1)(y_4 - y_2)}$$

is an invariant (for details see Van Gool et al 1993b). Note that this expression would simply be the cross ratio if the points were collinear, which they not necessarily are in this case. These examples show that restrictions on the scene configurations or the relative motions of objects and cameras can also simplify the invariants. Such simplifications certainly are not guaranteed to result from just any kind of constraint on configurations and possible relative motions.

One often needs constraints in order to find invariants at all. It can, for example, be shown that no invariants can be found for an arbitrary 3-D point cloud from a single view. One of the major challenges in the coming years will be to tabulate explicitly which situations defy the invariance approach, which invariants exist when, and what the precision of reconstruction is that they allow (eg up to Euclidean motion, similarity, affine transformation, projective transformation). Important steps have been taken already for the analysis of nonplanar 3-D structures from multiple views (eg Faugeras 1992; Koenderink and van Doorn 1992; Mohr et al 1992; Van Gool et al 1992a). Also, constraints have been found which yield invariants for 3-D scenes from single views (eg Forsyth et al 1992; Rothwell et al 1993) and investigations into the invariant-based recognition of articulated objects have started (Moons et al 1993).

Last, there is a direct connection between the building blocks and the non-accidental properties themselves. If we look back at the Euclidean interpretations in figure 4, nonaccidental properties arise when these building blocks vanish. Although the building blocks themselves are not invariant, their zero-crossings are. Therefore,

$$|\mathbf{x}_1 - \mathbf{x}_2 \quad \mathbf{x}_1 - \mathbf{x}_3| = 0,$$

which occurs when the three points are collinear and since this determinant will vanish just the same after an affine or projective transformation, collinearity is an invariant property. Similarly,

$$|\mathbf{x}_1 - \mathbf{x}_2 \quad \mathbf{x}_1^{(1)}| = 0$$

corresponds to tangency being an invariant property, ie the line tangent to a transformed curve is the transform of the line tangent to the original curve.

$$|\mathbf{x}_1^{(1)} \quad \mathbf{x}_1^{(2)}| = 0$$

illustrates that inflexions are invariant points (which is not the case for curvature extrema, not even under affine transformations). Last,

$$|\mathbf{x}_1^{(1)} \quad \mathbf{x}_2^{(1)}| = 0$$

corresponds to the preservation of parallelism under affine transformations (this determinant is not a building block of projective invariants).

7 Discussion and conclusions

Apart from the very existence of all these marvellous properties that are undisturbed by the rather dramatic changes in the way objects present themselves to an observer, there are several intriguing consequences as to the way recognition can be expected to proceed when they are being used.

By using invariants as a basis for recognition, one escapes from the traditional computer-vision approach of testing joint hypotheses about the identity of objects and the way they are positioned in space. After such a hypothesis-and-test paradigm, the visual system should almost be endowed with graphics capabilities. The way an object ought to look when positioned in the hypothesised way has to be predicted and compared with the actual projection of the scene. Is every man's homunculus a Leonardo da Vinci? This is most doubtful. Instead, the invariance-based paradigm decouples identity from pose hypotheses. Objects can be recognised without assumptions being made on their position and orientation in space. The determination of position and orientation can then be attacked as a second step, since the invariance-based match also yields information concerning the particular deformation that occurred. This deformation can be attributed to the particular orientation and relative position under which the object was observed. Decoupling recognition and pose determination of objects might seem a strange way of dealing with scene structure. However, an analogous dichotomy has been suggested by neurophysiologists (Ungerleider and Mishkin 1982; Vaina 1990) where the temporal visual pathway has been accredited with recognition, and pose has been suggested to be determined in the parietal pathway.

The invariance-based paradigm also has a direct bearing on times needed for recognition. In the computer-vision literature, emphasis has been put on how to render recognition times almost independent of the number of different objects known to the system (Rothwell et al 1992b). This is achieved through a kind of lookup table and stands in sharp contrast to the usually worse than linear growth of recognition times typical for the hypothesis-and-test techniques. Confronted with too many objects, a straightforward lookup-table strategy cannot be maintained either, since more and more features will have to be invoked to reach sufficient discriminating power. Nevertheless, the invariance-based approach seems more promising in avoiding exploding recognition times, which clearly do not plague human perception (Sternberg 1966; Van Zandt and Townsend 1993).

Recognition times in perception sometimes vary linearly with some pose parameter. Examples are mental rotation and scaling (Shepard and Cooper 1982). These results are not necessarily in contradiction to the use of invariants. A possible explanation of these effects would be the time needed to find corresponding reference points used in the extraction of the invariants. These will move further away from their original position and finding a correct correspondence may take more time.

The discussed invariants are also invariant under a change in focal length (f). This invariance was not intended, but nevertheless is very useful. This is another aspect worth some careful study. According to the group that is considered, it may 'absorb' some or all of the intrinsic parameters, such that changing these parameters too would only yield the same group. The knowledge a system—be it biological or artificial—needs concerning its own parameter settings can often be tuned by selecting smaller or larger groups, exchanging complexity of calibration and precision of reconstruction for complexity of invariant extraction.

Of course, the mathematical study of the transformations and their invariants is only a first step. Students of human perception will have to try to discover which of these invariants are used, if any. Although this kind of research is only in its infancy, the preliminary conclusions do not seem to be optimistic. Niall and Macnamara

concluded that "The deeply-rooted tradition that explains shape constancy by appeal to projective invariance has no adequate empirical support" (1990, page 658). In a similar vein, Cutting (1992) recently argued that the search for perceptually useful invariants has yielded a lean harvest. On the other hand, ongoing research (Wagemans et al 1993) shows that affine-transformed versions of planar four-dot patterns can quite successfully be labelled as same or different. Recent results (Viviani and Stucchi 1992), indicating that the power $\frac{1}{3}$ of curvature seems to govern *perceived* velocity of a point along a curve, might suggest the relevance of affine arc-length (the authors of that paper prefer another explanation). Our results (Wagemans 1992; Wagemans 1993; Wagemans et al 1993), also, suggest that nonaccidental properties such as collinearity and parallelism are important cues in affine shape equivalence for human perceivers. Perhaps, the negative results obtained by Niall (1992) and Niall and Macnamara (1990) might partly be based on the exclusion of these properties from the set of shapes they used in their experiments. In any event, we agree with Cutting (1983, page 314) that "Unless one can specify that some invariants are used in perception and some are not, ... the use of the concepts of group, transformation, and invariance in perception may be little more than whistling in the breeze". We hope that our paper may serve as a useful source of background information to help to foster this kind of research.

Acknowledgements. This work was supported by Esprit BRAs 3274 'FIRST' and 6448 'VIVA', the Belgian National Fund for Scientific Research project FKFO 2.0065.90, and the IUAP-50 project. Theo Moons, Eric Pauwels, and Johan Wagemans gratefully acknowledge the post-doctoral research grants offered by the National Fund for Scientific Research (NFWO).

References

- Allik J, Tuulmets T, Vos P G, 1991 "Size invariance in visual number discrimination" *Psychological Research* **53** 290-295
- Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115-147
- Biederman I, Cooper E E, 1991 "Evidence for complete translational and reflectional invariance in visual object priming" *Perception* **20** 585-593
- Brill M, Barrett E, Payton P, 1992 "Projective invariants for curves in two and three dimension", in *Geometric Invariance in Computer Vision* Eds J Mundy, A Zisserman (Cambridge, MA: MIT Press) pp 193-214
- Cutting J E, 1983 "Four assumptions about invariance in perception" *Journal of Experimental Psychology: Human Perception and Performance* **9** 310-317
- Cutting J E, 1986 *Perception with an Eye for Motion* (Cambridge, MA: MIT Press)
- Cutting J E, 1992 "Beyond Gibson's legacies" *International Journal of Psychology* **27** 15 (SY005.4)
- Efimov N, 1981 *Géométrie Supérieure* (Moscow: Mir)
- Faugeras O, 1992 "What can be seen in three dimensions with an uncalibrated stereo rig?", in *Proceedings of the Second European Conference on Computer Vision* Ed. G Sandini (Berlin: Springer) p 563-578
- Forsyth D, Mundy J, Zisserman A, Rothwell C, 1992 "Efficient recognition of rotationally symmetric surfaces and straight homogeneous generalized cylinders", in *Proceedings of the Second European Conference on Computer Vision* Ed. G Sandini (Berlin: Springer) pp 639-648
- Gibson J J, 1950 *The Perception of the Visual World* (Boston, MA: Houghton Mifflin)
- Hoffman D, Richards W, 1984 "Parts of recognition" *Cognition* **18** 65-96
- Huttenlocher D, Ullman S, 1987 "Object recognition using alignment", in *Proceedings of the First International Conference on Computer Vision* (Los Alamitos, CA: IEEE Computer Society Press) pp 102-111
- Koenderink J, van Doorn A, 1992 "Affine structure from motion" *Journal of the Optical Society of America A* **8** 377-385
- Lowe D G, 1987 "Three-dimensional object recognition from single two-dimensional images" *Artificial Intelligence* **31** 355-395
- Mohr R, Morin L, Grosso E, 1992 "Relative positioning with uncalibrated cameras", in *Geometric Invariance in Computer Vision* Eds J Mundy, A Zisserman (Cambridge, MA: MIT Press) pp 440-460

- Moons T, Pauwels E, Van Gool L, Oosterlinck A, 1993 "Viewpoint invariant characterization of objects composed of different rigid parts: a mathematical framework", in *Geometry and Topology of Submanifolds* volume 5 Eds F Dillen, L Verstraelen (Singapore: World Scientific)
- Nazir T A, O'Regan J K, 1990 "Some results on translation invariance in the human visual system" *Spatial Vision* 5 81-100
- Niall K K, 1992 "Projective invariance and the kinetic depth effect" *Acta Psychologica* 81 127-168
- Niall K K, Macnamara J, 1990 "Projective invariance and picture perception" *Perception* 19 637-660
- Rothwell C, Forsyth D, Zisserman A, Mundy J, 1993 "Extracting projective information from single views of 3D point sets", in *Proceedings of the Third International Conference on Computer Vision* (Los Alamitos, CA: IEEE Computer Society Press) pp 573-582
- Rothwell C, Zisserman A, Forsyth D, Mundy J, 1992a "Canonical frames for planar object recognition", in *Proceedings of the Second European Conference on Computer Vision* Ed. G Sandini (Berlin: Springer) pp 757-772
- Rothwell C, Zisserman A, Mundy J, Forsyth D, 1992b "Efficient model library access by projectively invariant indexing functions", in *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Los Alamitos, CA: IEEE Computer Society Press) pp 109-114
- Shepard R, Cooper L, 1982 *Mental Images and their Transformations* (Cambridge, MA: MIT Press)
- Sternberg S, 1966 "High-speed scanning in human memory" *Science* 153 652-654
- Ullman S, 1989 "Aligning pictorial descriptions: An approach to object recognition" *Cognition* 32 193-254
- Ungerleider L, Mishkin M, 1982 "The cortical visual systems", in *Analysis of Visual Behavior* Eds D J Ingle, M A Goodale, R J W Mansfield (Cambridge, MA: MIT Press) pp 549-586
- Vaina L, 1990 "What and where in the human visual system: Two hierarchies of visual modules" *Synthese* 83 49-91
- Van Diest M, Van Gool L, Moons T, Pauwels E, 1994 "Projective invariants for planar contour recognition", in *Proceedings of the Third European Conference on Computer Vision* Ed. J-O Eklundh (Berlin: Springer) pp 527-534
- Van Gool L, Brill M, Barrett E, Moons T, Pauwels E, 1992a "Semi-differential invariants for nonplanar curves", in *Applications of Invariance in Vision* Eds J Mundy, A Zisserman (Cambridge, MA: MIT Press) pp 293-309
- Van Gool L, Kempenaers P, Oosterlinck A, 1991 "Recognition and semidifferential invariants", in *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Los Alamitos, CA: IEEE Computer Society Press) pp 454-460
- Van Gool L, Moons T, Pauwels E, Oosterlinck A, 1992b "Semi-differential invariants", in *Applications of Invariance in Vision* Eds J Mundy, A Zisserman (Cambridge, MA: MIT Press) pp 157-192
- Van Gool L, Moons T, Pauwels E, Wagemans J, 1993a "Invariance from the Euclidean geometer's perspective" University of Leuven, technical report KUL/ESAT/MI2/9310
- Van Gool L, Moons T, Van Diest M, Pauwels E, 1993b "Matching perspective views of parallel plane structures", in *Proceedings of the NSF-Esprit Workshop on Applications of Invariance in Computer Vision II* Eds J Mundy, A Zisserman (Cambridge, MA: MIT Press) pp 351-363
- Van Gool L, Wagemans J, Vandeneede J, Oosterlinck A, 1990 "Similarity extraction and modeling", in *Proceedings of the Third International Conference on Computer Vision* (Los Alamitos, CA: IEEE Computer Society Press) pp 530-534
- Van Zandt T, Townsend J, 1993 "Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluation review" *Perception & Psychophysics* 53 563-580
- Viviani P, Stucchi N, 1992 "Biological movements look uniform: Evidence of motor-perceptual interactions" *Journal of Experimental Psychology: Human Perception and Performance* 18 603-623
- Wagemans J, 1992 "Perceptual use of nonaccidental properties" *Canadian Journal of Psychology* 46 236-279
- Wagemans J, 1993 "Skewed symmetry: A nonaccidental property used to perceive visual form" *Journal of Experimental Psychology: Human Perception and Performance* 19 364-380
- Wagemans J, Van Gool L, De Troy A, Foster D H, Wood J R, 1993 "Minimal information to determine affine shape equivalence" *Perception (Supplement)* 22 107-108