

# Invariants and Fixed Structures Lead the Way to Change

Luc Van Gool

ESAT-MI2, Kath. Universiteit Leuven  
Luc.VanGool@esat.kuleuven.ac.be

**Abstract.** The use of invariants quickly gained impetus in the computer vision community. The paper recapitulates why this strand of research has become so influential by summarizing the traditional advantages often highlighted in the context of object recognition. Then, however, the paper moves on to corroborate the importance of certain geometrical entities, called “fixed structures”. It is argued that looking at these entities forms a core idea, that can be considered central to several applications. These include as diverse subjects as grouping and three-dimensional scene reconstruction. A careful and systematic study of fixed structures, the corresponding subgroups, and their invariants is advocated.

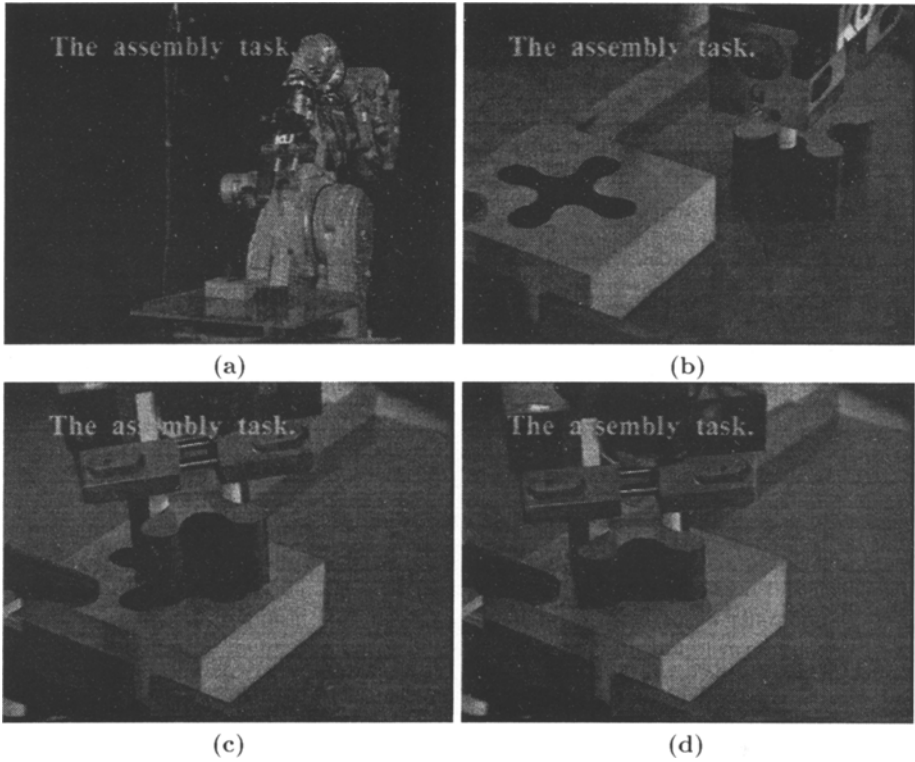
## 1 Introduction

Intensive research on the use of invariance in computer vision has started only a few years ago, yet it seems it will leave its mark in several areas. For one thing, it has strongly influenced the thinking about shape recognition. In the meantime, its ramifications are going well beyond this particular and most obvious area of application. The paper highlights the study of “fixed structures” as one of the key ideas underlying diverse trends in ongoing research.

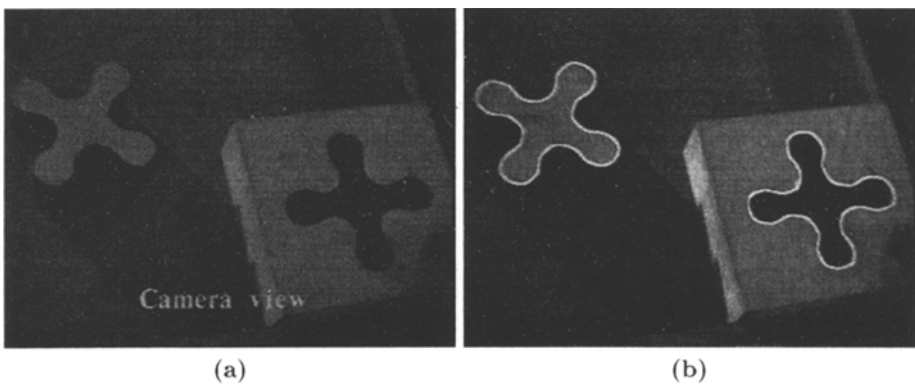
In section 2, the usefulness of invariants for recognition is summarised. These are the rationales that have triggered the interest in invariance in the first place. Section 3 discusses the concept of fixed structures and section 4 illustrates their use in two areas: grouping and three-dimensional reconstruction. Throughout the paper, mathematical derivations are avoided. Emphasis is on the general idea.

## 2 An Early Example: Recognition

Consider fig. 1(a). It shows a robot before a table with workpieces. The scene is viewed by a camera mounted on the robot’s end effector, which is also equipped with force feedback. The task is to insert a pen into a similarly shaped hole, for objects as shown in fig. 2(a). The viewing angle is unknown, as well as the positions and orientations of the objects. There can be more objects than just two. In particular, irrelevant distractor objects are allowed to lie in the scene. The problem is then to recognize pairs of corresponding positive and negative objects,



**Fig. 1.** (a) Robot assembly cell; (b) robot grasping the pen object; (c) approach of the hole from visual estimates; (d) insertion based on force feedback.



**Fig. 2.** (a) Scene with pair of corresponding pen and hole objects; (b) recognized shapes superimposed in extracted pose.

to grasp the positive object, and to plug it into the negative. The problem was solved by using affine invariants to match the shape cross-sections. Rationales for using invariants in such application include:

Viewpoint independent operation: The use of the affine invariants made the difference in slants and tilts between objects and between views irrelevant. The invariant features are not affected by these changes.

Handling scene clutter: Invariants allow the computer to extract descriptions that are local and therefore allow for recognition of objects that are not completely visible or under conditions where edge extraction leaves gaps and other imperfections. Yet, at the same time, these descriptions are sufficiently rich to distinguish shapes.

Efficiency: The fact that the invariants do not change allows the system to use simple look-up tables for the matching. The invariants are used as the addresses to read-out the table entries, which contain the different model objects that are compatible with the measured invariants. Such process, usually called "indexing" [8], recognizes the shape without having to know its position and orientation. This avoids the hypothesis/verification loop that is typical of many recognition strategies and that tends to take a lot of time because of the enormous search spaces. After the indexing has suggested a few possible matches, a more detailed, but still invariant based comparison solves the remaining ambiguities and yields the position and orientation information by looking for the affine transformation that relates the model and the matched image projection.

No need for calibration: For the matching process no calibration of the camera parameters or its relative position with respect to the robot is needed. Time-to-contact type procedures can bring the gripper in the neighbourhood of the object, from where precise positioning of the gripper can be guided by a second camera, or as in the example, with force feedback.

These are the major advantages that have ensured invariants' initial success in vision. They have been described many times in the literature, that also includes recognition under more general conditions such as taking the full perspective effects into account by using projective invariants.

### 3 Fixed Structures

Invariance theory can be brought to bear if the changes that one wants the vision system to be immune against can be expressed as *group actions*, i.e. as transformations belonging to a group. The more parameters the group has, the more complicated the corresponding invariants will get. It is therefore advantageous to identify the smallest group that still applies. In the example of section 2 projective invariants could have been used, thereby including robustness against additional, but small perspective deformations. This, however, proved to be an unnecessary luxury, that would have required the use of invariants with a higher complexity. One can also actively simplify the situation by controlling the environment and lowering the variability in it. Note, however, that removing a degree

of freedom not necessarily simplifies the problem, i.e. not necessarily introduces a smaller group. As an example, if one is to match shapes under the conditions of the previous section, it does not help to fix the camera's focal length. The remaining variability will still call for dealing with the complete group of plane affine transformations. The theories that underlie the use of invariants, such as the analysis of Lie prolongations [12], allow one to identify the parameters that bring the most benefit if they can be measured or fixed.

Here I would like to focus on smaller groups that emerge because the transformations keep certain geometric entities in the scene fixed. As an example, consider a planar, mirror symmetric shape. If one considers the two symmetric halves, then the transformation joining symmetric points between their deformed image projections (assuming affine skewing) will not be a general affine transformation as considered in section 2. Actually, if one were to know the symmetry axis, the distances of a point and its symmetric counterpart to the symmetry axis are equal. This invariant can be calculated on the basis of a single point (if the axis is known), whereas a general affine transformation would require at least four points for the calculation of an invariant. It is clear that the crux of the matter lies with the assumption that we know the axis.

In the symmetry example, the line corresponding to the axis is so special because it stays fixed under the symmetry transformation (i.e. under the transformation that corresponds to the skewed symmetry). As a matter of fact, every single point on that line remains fixed. A moment's thought shows that if one considers all the transformations in a group that keep a certain geometric structure fixed – as the axis in the example – then these particular transformations form a subgroup. The subgroup has fewer parameters and hence the invariants can be obtained with less information. Not only are the invariants simpler in that sense (the expressions need not be simpler, however), they are also better geared towards the task at hand, i.e. they are more selective in the kind of changes they allow for. In the sequel the symmetry example will be extended to perspectively skewed symmetries.

Summarizing the foregoing observations, it is interesting to find out the smallest group that applies to a problem. If fixed structures of the transformations are known, then these will deliver such a subgroup.

## 4 Exploiting Fixed Structures

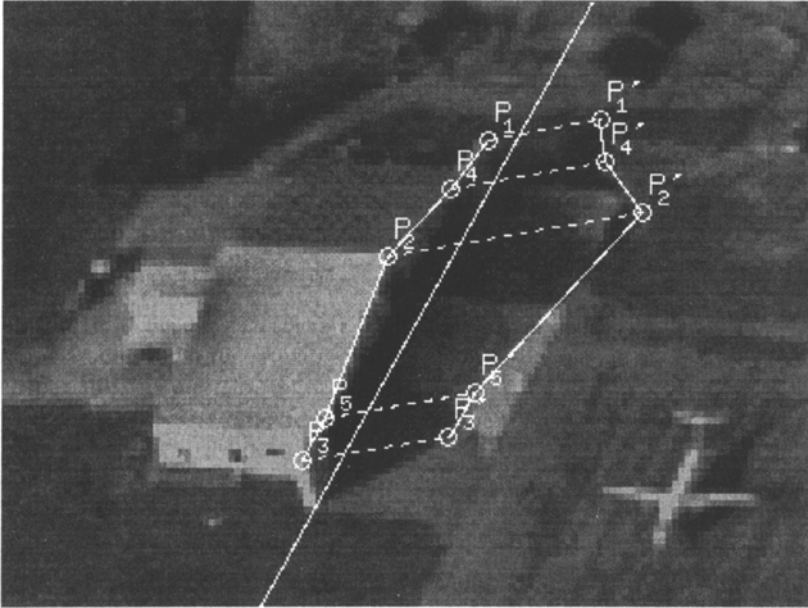
This section discusses how the fixed structure observation ties in with some recent strands of research. A first example is geometry-based grouping. As a second example, the problem of uncalibrated 3D reconstruction is sketched. In both cases, considering fixed structures of interest brings about a kind of stratification of possible subgroups of the most general group.

### 4.1 Grouping

The seminal non-accidental approach to grouping [4, 5] is based on lists of properties that very probably have not arisen by accident. Typically, these special

configurations correspond to transformations with fixed structures. Symmetry is an example that will usually figure prominently in these lists. In fact, basing the classification of non-accidental properties on the classification of fixed structures yields some interesting, additional cases [11]. A discussion of such classification is out of the scope of the presentation and it will therefore be limited to two related cases.

The first are so-called planar homologies. These are transformations that keep a line pointwise fixed, as well as a point off that line. A practical configuration that corresponds to such case is given in fig. 3. A planar wall casts a



**Fig. 3.** *A hangar and its shadow. Corresponding points of the facade and the shadow transform into each other through a planar homology.*

shadow onto a ground plane. The line that is pointwise fixed under the shadow casting transformation corresponds to the intersection of the facade plane with the ground plane. The fixed point off the line corresponds with the position of the sun (here assumed to be a point source) in the image. If both the fixed line and the fixed point are known, then only a single degree of freedom remains. Invariants under this 1D subgroup of the 8D plane projectivities are easy to derive. It is not without interest to have a closer look to the remaining d.o.f. It corresponds to the cross ratio of the fixed point, the pair of corresponding points, and the intersection of their join with the fixed line [9].

The previous case can be further specialised to that of perspectively skewed mirror symmetry (and equally well point symmetry for that matter). What it

takes is that a pair of points is mapped onto each other, i.e. that the pair of points is a fixed *as a pair*. This addition to the fixed structures of the planar homology yield a so-called harmonic homology. It lifts the remaining degree of freedom and the fixed cross-ratio is -1. Suppose one wants to check whether two points (on contours) are a symmetric pair. In that case, the points form a fixed pair and the following invariant should apply [13]:

$$\frac{\kappa}{\sin^3 \theta} = \frac{\kappa'}{\sin^3 \theta'}$$

where  $\kappa$  and  $\kappa'$  are the curvatures of the symmetric contours in these points and the meaning of  $\theta$  and  $\theta'$  is explained in fig. 4. Readers interested in symmetry will

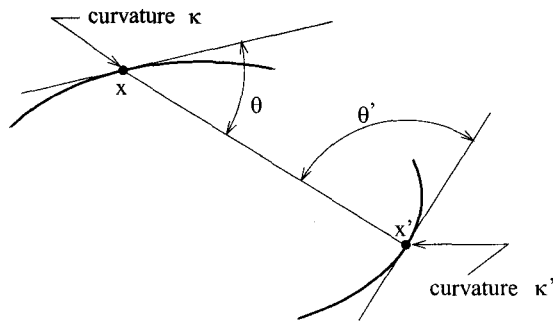
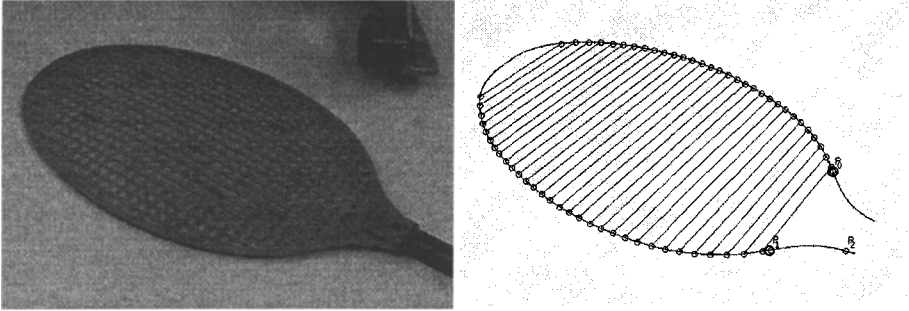


Fig. 4. variables used in Ponce's symmetry invariant.

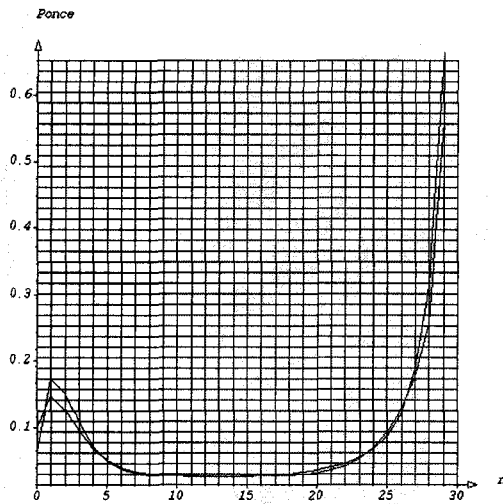
recognize this relation as the one introduced by Ponce [7] for symmetric points in the case of *affine* skewing. But as comes out, this relation also applies under perspective skewing. What this means is that  $\kappa/\sin^3 \theta$  is an invariant under the specific projective transformations that correspond to mirror symmetries as seen in the image, whereas such expression is by no means invariant under general projectivities. The fact that the pair of symmetric points is fixed (the pair is mapped to the pair, i.e. the points are swapped) has been instrumental in deriving this result. It allows to test a point pair for its potential symmetry.

Figure 5 shows a picture of a fly-flapper with a clear, perspective distortion. A spline has been fitted to the contour and was used to calculate  $\kappa$  and  $\theta$  and a collection of corresponding points. Figure 6 shows the value for the  $\kappa/\sin^3 \theta$  invariant for corresponding points on both halves. Clearly, they match quite well, indicating that the invariant can be used to generate point correspondences. In fact, two such correspondences suffice to completely fix the harmonic homology, i.e. the symmetry transformation in the image.

Symmetry is just one of the so-called "non-accidental" configurations used for grouping. The use of invariants can both speed up and specialize the search.



**Fig. 5.** *Left: Image of a fly-flapper; Right: corresponding points on the spline fit.*



**Fig. 6.** *The Ponce invariant for corresponding points (one curve for each half).*

The classification of subgroups according to the structures that they keep fixed yields a systematic approach for listing non-accidental configurations. There is a strong link between non-accidentalness and fixed structures in the sense that if a structure remains fixed in the original pattern, then it will also remain fixed under the transformation that includes the effects of image projection. Indeed, the transformations in the image are projection-conjugated versions of the original transformations and conjugations keep fixed structures intact (e.g. same eigenvalue multiplicities), in contrast to many other geometric properties.

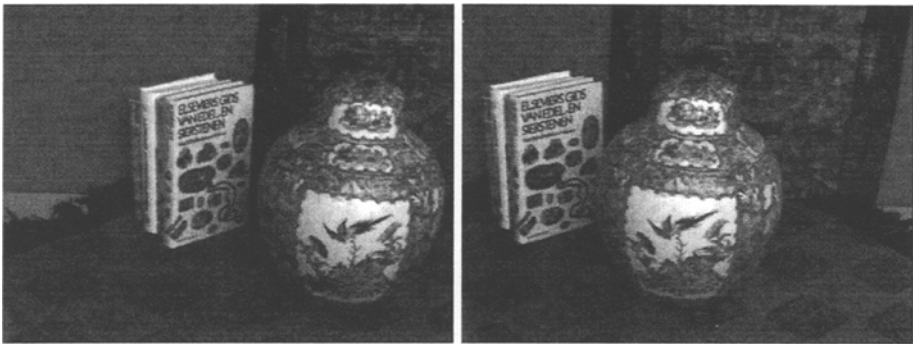
## 4.2 Uncalibrated Reconstruction

In the past, the use of range finders and stereo rigs would require a substantial time for their installation. This was hampering progress in the 3D area. The

development of systems for 3D reconstruction from uncalibrated sets of images would yield an empowering technology, where 3D reconstruction could be brought at about anybody's fingertips. Moving around with a camcorder would suffice to generate the necessary imagery.

In the discussion on grouping, it was suggested that the study of fixed structures helps to stratify a group (like the group of plane projective transformations) into subgroups. A very similar strategy lies at the heart of quite a few papers in the field 3D scene reconstruction from uncalibrated images. From two such images, only 3D projective structure can be recovered [2, 3]. Such a reconstruction can be upgraded, however, if the plane at infinity can be identified, which is kept fixed by all affine transformations of the 3D world. Moving this plane to infinity allows to preserve properties such as parallelism and length ratios along a line. A further upgrade to metric structure then follows by identifying the absolute conic in that plane, because all Euclidean transformations keep that conic fixed (see [9]).

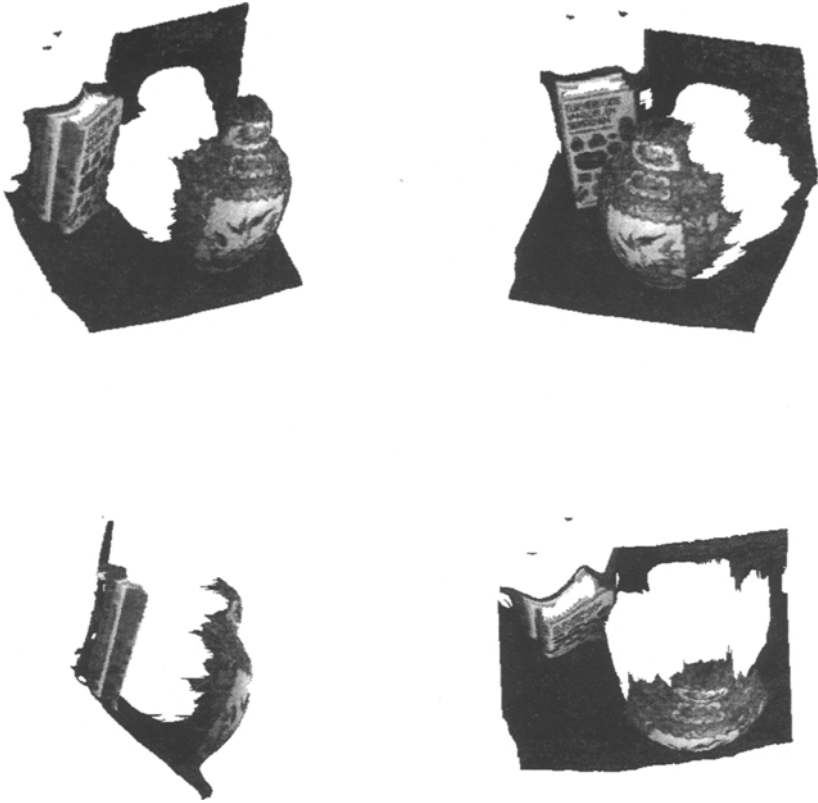
A discussion of the work that exploits such fixed structures again is outside our current scope, but a simple example is given by the opportunities offered by a translating camera. In that case it is easy to find vanishing points of several 3D orientations. All these points lie in the plane at infinity. Having identified this plane it is possible to build an affine reconstruction [10]. Note that the direction of translation or the translation distance between the two views need not be known. The camera model is fully perspective and its parameters are left uncalibrated. Fig. 7 shows two images obtained with a translating camera. Four different views of the affine reconstruction are shown in fig. 8. Note that



*Fig. 7. Two images of a Chinese vase scene obtained with a translating camera.*

the scene is skewed, but this is normal since an affine reconstruction does not preserve the angles. On the other hand, it is not always necessary to know about absolute distances or metric structure. An affine reconstruction can e.g. be quite sufficient for guiding autonomous vehicles. Important information such as the planarity of the road, its midline, the parallelism of its sides, etc. are





**Fig. 8.** *Four views of the affine reconstruction of the vase scene. The bottom row reconstructions are taken from an extreme side and top position.*

all preserved under affine reconstructions. Translation has also been used to yield an affine initialisation, followed by an upgrade to metric structure from a subsequent, general motion [1, 6]. Depending on the degree to which the fixed structures have been identified, one would then have to use 3D projective, affine, or metric invariants for 3D shape recognition.

Exploiting the plane at infinity and its absolute conic as the fixed structures of the 3D world can be considered as the most basic stratification. An obvious extension is to look at additional fixed structures that have to do with shape symmetries, for instance, and to extend the work on grouping to 3D.

## 5 Conclusions

There is an interesting interplay between fixed structures, subgroups, and invariants. Looking from this vantage point to geometric problems in vision can help to get a more systematic handle on different issues. In the case of uncalibrated 3D reconstruction most effort so far went into detecting the fixed structures, whereas the invariants are relatively easy to derive. In the case of grouping, most attention has been paid to deriving dedicated invariants, whereas detecting the fixed structures needs to be considered and exploited further. Hough like strategies are one promising route, because the fixed structures have a simple shape such as a line, however complex the shapes of e.g. a mirror symmetric pattern happen to be.

**acknowledgements:** This work has been supported by Esprit BRA 'VIVA' and Belgian DWTC project IUAP-50 (Inter-University Attraction Pole).

## References

1. M. Armstrong, A. Zisserman, and P. Beardsley, Euclidean structure from uncalibrated images, 5th BMVC, 1994
2. O. Faugeras, What can be seen in three dimensions from an uncalibrated stereo rig, Proc. 2nd ECCV, pp. 321-334, 1992
3. R. Hartley, Estimation of relative camera positions for uncalibrated cameras, proc. 2nd ECCV, pp. 579-587, 1992
4. T. Kanade, Recovery of the 3-dimensional shape of an object from a single view, Artificial Intelligence, Vol. 17, pp. 75-116, 1981
5. D. Lowe, Perceptual organisation and visual recognition, Kluwer Academic Publishers, 1985
6. M. Pollefeys, L. Van Gool, and M. Proesmans, Euclidean 3D reconstruction from image sequences with variable focal lengths, Proc. European Conf. Computer Vision, Vol.I, pp. 31-42, 1996
7. J. Ponce, On characterising ribbons and finding skewed symmetries, Proc. Int. Conf. Robotics and Automation, pp. 49-54, 1989
8. C. Rothwell, Recognition using projective invariance, PhD Thesis, Univ. Oxford, 1993
9. J. Semple and G. Kneebone, Algebraic projective geometry, Oxford Univ. Press, 1952
10. L. Van Gool, T. Moons, M. Proesmans, and M. Van Diest, Affine reconstruction from perspective image pairs obtained by a translating camera, Proc. Int. Conf. Pattern Recognition, Jerusalem, pp. A/290-A/294, oct. 1994
11. L. Van Gool, M. Proesmans, and T. Moons, Groups for grouping, SPIE Int. Symp. on Optical Science, Appl. of Digital Image Processing XVIII, Vol.2564, pp.402-413, 1995
12. L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck, Vision and Lie's approach to invariance, *Image and Vision Computing*, Vol. 13, No. 4, pp. 259-277, may 1995
13. L. Van Gool, T. Moons, and M. Proesmans, Mirror and point symmetry under perspective skewing, to be published in prod. CVPR, 1996