Review
Geometric Invariants and Object Recognition

ISAAC WEISS
Center for Automation Research, University of Maryland, College Park, MD 20742


Abstract
We discuss the role of the general invariance concept in object recognition, and review the classical and recent
literature on projective invariance. Invariants help solve major problems of object recognition. For instance, different
images of the same object often differ from each other, because of the different viewpoint from which they
were obtained. To match the two images, common methods thus need to find the correct viewpoint, a difficult
problem that can involve search in a high dimensional space of all possible points of view and/or finding point
correspondences. Geometric invariants are shape descriptors, computed from the geometry of the shape, that remain
unchanged under geometric transformations such as changing the viewpoint. Thus they can be matched without
search. Deformations of objects are another important class of geometric changes for which invariance is useful.

1 Introduction

Object recognition is a major goal of computer vision, but many obstacles remain on the road toward effective
recognition systems. Here we discuss ways of overcoming many of the difficulties by using invariants of shapes.
A typical problem is that an object can be seen from different points of view, resulting in different images
which we would like to recognize as portraying the same object. In a typical recognition task we have one
image stored in a database, and we need to compare it with an image of an unknown object observed from
an unknown point of view. This difficult task can be greatly facilitated by using suitable invariants. These
are shape descriptors computed from the image which are independent of the viewpoint, that is, they are the
same regardless of which point of view the image was taken from.

It can be argued that object recognition is the search for invariants. Given an image of an object, we want
to extract one unique invariant: a name or a similar ultimate descriptor. Given another image of the same
object, differing from the first by, for example, viewpoint, we want to extract the same unique descriptor.
To do that, we have to eliminate in some way the effect of the transformations that gave rise to the differences
between the images.

There are several methods of eliminating transformations between images. The simplest is to perform
every possible transformation of one image to see if any of its transformed versions match the original image.
For instance, in template matching (Ballard & Brown 1982), it is assumed that a template and an image differ
only by translation, and the template is moved pixel by pixel over the image until a match is found. However,
when more complicated transformations are involved, such as rotation, projection, etc., the search space be-
comes overwhelmingly large.

To reduce the search space, “invariant features” can be used (Lowe 1985). These are features in the image
that stay invariant under some transformation and can be matched directly between the two images. For ex-
ample, an edge remains an edge, so edges can be used for matching. The problem here is that the kinds of
features usually used do not have much distinctiveness. Any edge in one image can match any edge in the other.
This leads to the correspondence problem, which can easily lead to a combinatorial explosion. Invariant con-
straints (Grimson & Lozano-Perez 1987) can help here but they still leave a large space for search.
Other methods aimed at viewpoint invariance have their own drawbacks. Fourier descriptors are not fully invariant and suffer from occlusion problems. Hashing methods such as the Hough transform break down when a large number of parameters is involved.

The correspondence problem can be solved by using more distinctive invariant descriptors—that is, descriptors that are invariant only to the transformation we are interested in and not to others. For instance, a shape descriptor of a fish should be distinct from a descriptor of a frog—that is, it should not be invariant to a transformation that maps the shape of the fish into that of the frog. Edges, of course, are invariant to this since they can appear in both shapes; they are “too” invariant, namely they are invariant to too wide a set of transformations. Thus, we must try to find features that are invariant only to the transformations that we want to eliminate and to no others, so they are distinctive enough to match without ambiguity.

As an example, the projective invariants are invariant only to the change of viewpoint, not to any other transformations. General projective invariants were first described for vision by Weiss (1988), although special cases were used before (section 3).

Change in the point of view is only one kind of geometric transformation that images can undergo. For instance, we would like to identify an object as a “fish” even if the particular example of a fish we are looking at is somewhat thinner or fatter than some standard fish. In this case we need invariants to deformations, that is, quantities that will not change under a not-too-great deformation of the object. It is again important not to seek invariance to transformations that are too general, because then the descriptors will blur the distinction between different objects.

A fundamental question immediately arises: what transformations do we want to eliminate? When do we decide that two images come from the same object, even though they are different? Viewpoint change is one example; other transformations will probably depend on the types of objects in question.

Another consideration in choosing the kind of invariance we need is that the larger the set of transformations, the harder it is to extract meaningful distinctive quantities that are invariant to it. (For example: distance, a Euclidean invariant, is not preserved under projections, a larger group.) Yet the need for invariants is much more acute, because the larger set of transformations has more unknown parameters and requires a search in a much bigger space. This consideration thus leads to the same conclusion as the distinctiveness argument: we have to find optimal invariants, that is, ones that will stay invariant under the set of transformations we want to eliminate, but not under a larger set.

A paradigm for object recognition can thus include the following:

1. Identify the transformations that an image can undergo and still describe the same object, that is, the transformations that we want to eliminate for particular classes of objects.
2. Find descriptors that are invariant to these transformations but not to others.
3. Use these descriptors to index shapes and match them.

The next section discusses point (1) above. The rest of the article carries out points (2) and (3) for projective transformations and their subsets, and finds a variety of invariants for them. For other transformations, these points are still being investigated.

2 Which Invariants?

Here we deal only with purely geometric invariants, that is, ones that can be calculated from the shape alone. Other surface properties such as shading, reflectance, color, etc., can also be considered as invariants, subject to the same considerations as above, but are not treated here.

The most obviously useful invariants in vision are the ones that are invariant to the Euclidean transformations—translation and rotation. A simple example is the length of a rod. In a simple world consisting of rods that lie in a plane, and images that can only move and rotate, we can identify a particular rod by measuring its length on the image and comparing it to a database of rod lengths. The rod’s position and orientation are irrelevant and can be ignored. As another example, when a 2-D curve is rotated or translated in the plane, its curvature at each point does not change. Thus curvature is an invariant of the Euclidean transformations. It is common to plot the curvature of such a curve as a function of its arc-length (which is invariant up to a starting point) to obtain a 2-D Euclidean invariant representation of the curve. Curvatures of surfaces have also been used when they can be measured, for example, from range data.

The formation of images in general involves a larger set of transformations (containing the Euclidean ones).