

Object Recognition using Affine and Projective Invariants of Planar Sets *

Kalle Åström

Dept. of Mathematics, Lund Institute of Technology
Box 118, S-221 00 Lund, Sweden

Abstract

A human with its highly complex and fine tuned visual system can easily recognize objects. It is however difficult to construct a computer program that can recognize general objects using visual input, e.g. from a CCD-camera. This paper treats recognition of planar non-algebraic curves. It is shown how invariant features, which are robust to occlusion, changes in lighting and viewpoint, can be extracted from a single gray-scale image. The invariant features can then be used in a fast recognition algorithm. Geometrical configurations are selected using edge detection and segmentation. Robust invariant features are then extracted from these configurations by choosing first a canonical or normal reference frame, and then features in this reference frame. These features automatically become invariant. In this work methods are given where a global and stable approach is used both in choosing the normal reference frame and in extracting invariants under affine and projective transformations. These invariants can be used to recognize fairly general feature configurations, and they can be used in a semi-local way to recognize occluded objects. It is demonstrated how an essentially unique affine reference frame can be chosen using weak isotropy, maximal compactness or moments. Using moments a unique distinguished affine reference frame can be chosen in a continuous way with respect to the Hausdorff metric. In the projective case it is shown how a non-unique reference frame can be chosen using moments. It is also shown that the method of maximal compactness in the projective case might give a unique reference frame for curves with a polygonal representation, but the method is unstable. It is not even well defined for smoothly curved curves. Finally it is shown that it is in fact impossible to select a unique projective normal reference frame in a continuous way with respect to the Hausdorff metric.

1 Introduction

The human visual system is highly complex. The retina of the eye contains approximately 100 million photoreceptors. Humans can easily obtain simple but quite accurate descriptions of the world with this system. It is possible to perceive depth almost instantaneously, reconstruct a scene and recognize things under different lighting and viewing conditions. Some parts of the human visual system, like the physical construction of the eye, are well known. The low-level massively parallel processing in the visual cortex are also known to some extent, cf. [HW1] for neurophysiological and [Ma1] for computational aspects. Less is known about the processes at higher levels in the brain.

*The work has been partially supported by the Swedish National Board for Technical and Industrial Development (NUTEK). The work is performed as part of the ESPRIT-BRA project VIVA (Viewpoint Invariant Visual Acquisition).

A camera can capture an image faithfully but it is not clear how the image should be processed to obtain information useful for machine vision. This is partly a conceptual problem and partly a problem of lacking computer power. One reason is the large amount of information. It is difficult to automatically distinguish between important and irrelevant pieces of information. Another reason is the way pictures change under different viewpoint and lighting conditions. This does not seem like a major problem for humans.

The present work is focused on the recognition problem in computer vision. The goal is to find and recognize structures in an image that are known to the system by a model. The idea is to reduce the amount of information in the image by extracting features that are invariant under changes in viewpoint and different lighting conditions. Geometrical features, like edges between regions of different intensity and texture, are invariant under different lighting conditions. Viewpoint invariant features are then extracted from these edges, and used in recognition. The theme is to explore how invariants can be used in this context. Invariants are part of classical algebra, cf. [Sch1], and classical geometry, cf. [Cox1].

The object recognition problem is discussed in Section 2, where an algorithm for object recognition based on invariants is presented. The section ends with a brief introduction to image processing. Some mathematical models are also discussed. The pinhole camera model is described in detail since it is necessary to understand that viewpoint invariance is equivalent to invariance under group action. The simple but useful principle of normalization with respect to group action is presented in Section 3. The key idea is that any choice of ‘normal’ reference frame naturally leads to a complete invariant description of an equivalence class of features. This normal reference frame can be used to define invariant indices and invariant metrics. Two affine normalization schemes are presented in Section 4. The first scheme uses weak isotropy or, equivalently, maximal compactness as a criterion for selecting the normal reference frames. This scheme appears to be relatively robust against perturbations of the curve, but it requires estimation of the tangent direction at each point of the curve. The other normalization scheme, which is based on moments, is continuous even with respect to the Hausdorff metric. Projective normalization schemes are investigated in Section 5. A method based on moments works for a large class of curves, but the normal reference frame obtained is not unique. It is shown that the normal reference frame based on maximal compactness is unstable and therefore unsuitable. It is not even well defined for smooth curves. Finally it is shown that it is impossible to select a unique reference frame in a continuous way with respect to the Hausdorff metric.

The algorithms have been implemented in MATLAB and tested on real images.

2 The recognition problem

The recognition problem is described in this section. The goal is to present the ideas of object recognition using invariants and to put up a framework on which Sections 3, 4 and 5 rely.

2.1 Object recognition using invariant features

The pinhole camera model, described in Section 2.3, is a fairly adequate model of a real camera. Points in three dimensions are projected onto a plane. Using this model it is straightforward to predict the image of a collection of objects in specified positions. The inverse problem, to identify and to determine the three dimensional positions of possible objects from a two-dimensional image, is however much more difficult. Traditionally recognition has been done by matching each model in a model data base with parts of the image. Model based recognition using viewpoint invariant features has recently attracted much attention, cf. [MZ1]. Invariant features are computed directly from the image and used as index functions in a model data base.

This gives algorithms which are faster than the traditional methods. These techniques cannot, however, be used to recognise general curves or point features in three dimensions by means of one single two-dimensional image. For point configurations it was shown in [BW1, Ås1] that only trivial invariants exist in the general case. Additional information, e.g. that the object is planar, is needed.

The ideas described here are inspired by the ESPRIT-project VIVA, and are heavily influenced by the recognition system that has been developed in Oxford, cf. [RZ1, Ro1]. The core of invariant based recognition is a chain of algorithms that reduce the enormous amount of information in a gray-scale image, typically $512 \times 512 = 262144$ bytes, to a smaller amount, typically $100 \times 3 = 300$ bytes. The idea is to throw away¹ information that varies with lighting, occlusion and viewpoint, and to keep the invariant features that allow recognition. This information reduction process is called the *extraction of invariant indices* and can be described in the following steps. It is interesting to note how the amount of information is reduced, how the information becomes less self-contained and how higher degrees of invariance are achieved.

Extraction of invariant indices

1. A gray-scale image is captured by a camera. The intensity depends very much on lighting conditions. The position and visibility of objects depends on occlusion and viewpoint. Two such gray-scale images are shown in Figure 1a and b. These contain $512 \times 512 = 262144$ bytes of data.
2. Edges between regions of different intensity or texture can be detected using an edge detector. The result is a number of curve fragments. Edge detection is briefly described in Section 2.2. In Figure 1c-d, data is reduced to 20 000 kbytes. Detected edges are less sensitive to changes in lighting conditions, but the position and visibility still depends on occlusion and viewpoint.
3. The curve fragments can be grouped or segmented into configurations of points, curves and regions. This important step, described in Section 2.2, is the key to obtain robustness under occlusion. The segments are less sensitive to lighting conditions and occlusions than the curve fragments from the edge detector. Only a geometrical ambiguity due to different viewpoints remain. Two such segments, in this case regions enclosed by a curve and a bitangent, are shown in Figure 1e-f, where data is reduced to 2 000 bytes.
4. Finally invariant indices are extracted from the configurations above. This is illustrated in Figure 1e-h. In Sections 3, 4 and 5, methods are presented on how to extract a small number of invariant indices from each segment. This is a purely geometrical/algebraical problem. Understanding of the transformation between the object and the image is essential. A model of the camera and the transformations involved are given in Section 2.3. In Figure 1g-h three invariant indices have been extracted from each region in Figure 1e-f. These numbers are invariant to changes in viewpoint, lighting conditions and occlusions. In Figure 1g-h, data is reduced to 6 bytes.

The extraction of invariant features are used both in *learning* a new object and in *recognizing* objects.

¹The other information is not necessarily thrown away, but kept for verification of hypotheses at different levels.

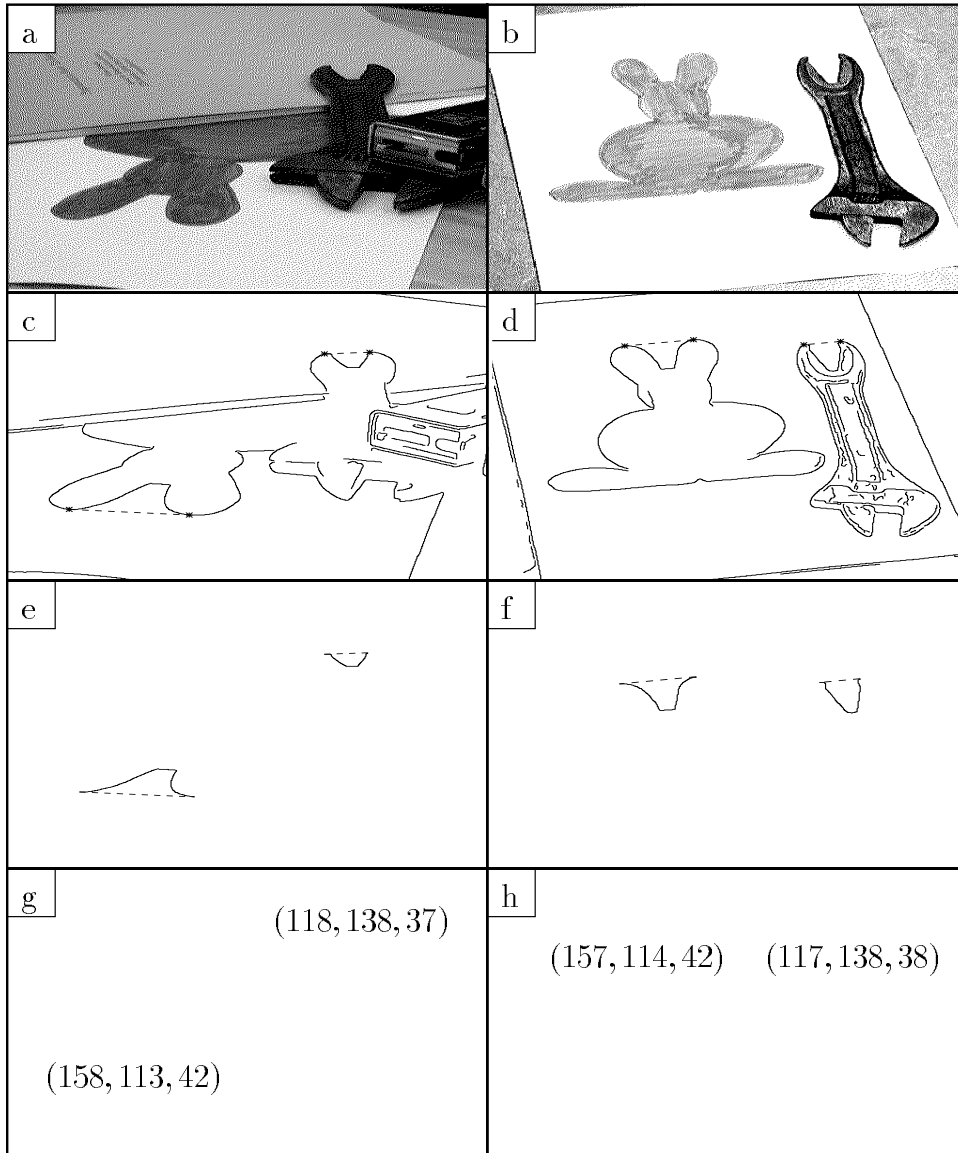


Figure 1: Two gray-scale images of a scene are shown in a and b. These contain objects that have approximately planar features. Edges, extracted using a Canny-Deriche edge detector are shown in c and d. Distinguished points on edge fragments are used to segment the edges into pieces in a projectively invariant way. Both points and curve segments can be used in recognition. Distinguished points and lines can also be used to extract small regions in a projectively invariant way. Two such regions are shown in each of the figures e and f. Viewpoint invariant features are extracted from the regions in e and f. Three invariant numbers from each concavity are shown in g and h. These serve as a label. A recognition table seems feasible.

Learning

1. Extract invariant indices from a model object by viewing it under favourable conditions as in Figure 1b or from an explicit description of the model, e.g. a CAD-model.
2. Store information about a model object at the position given by the indices in a ‘recognition table’. This step is done once and for all, before applying the recognition scheme to actual images.

Recognition

1. Extract invariant indices from an image. For each set of indices use table-lookup in the recognition table, and get either a negative answer, meaning that the segment is unknown to the system, or a positive answer, meaning that the segment could correspond to a segment in the recognition table.
2. For each hypothesis do additional verification/rejection using all information in the model segment and the viewed segment. We are now working our way up in the information tree.
3. Try to verify joint hypothesis or verify larger portions of an object by comparing the whole model for evidence in a scene.

In this work emphasis is made on the geometrical/algebraical construction of invariant indices from configurations of points, curves and compact regions.

2.2 Pre-processing of an image

Some pre-processing methods are described in this section. For general introductions into methods used in image processing, see [Ho1, GW1].

Edge and point detection

The first step in reducing the amount of information in an image in order to obtain invariance to variations in lighting, is to extract geometrical features in the image. This is typically done by filtering techniques. This problem has been studied for a long time. Edge detectors localize discontinuities in the gray-level function of the image, cf. the general references, or [HM1, Ma1, Can1, PM1]. Points of interest detectors like [WB1, Fr1] can also be used. An efficient implementation of Canny’s edge detector is described in [De1]. The experiments in this work uses this ‘Canny-Deriche’ edge detector. The input to the edge detector is a gray-scale image. The result is a binary image. This is illustrated in Figure 1a-d. The edge pixels of the binary image can be linked up to form polygonal curve patches, which will be called *curve fragments* in the sequel.

Segmentation

It may happen that a detected curve fragment correspond to a whole closed contour of one object. This is possible if the whole contour is unoccluded as in Figure 1b. This whole curve can then be used for recognition. A contour may however be fragmented because of errors. This can be seen in Figure 1d. Notice the fragmentation at the base of the rabbits head and also at the point where the body meets the leg. Some of these errors can be corrected directly. When an object is partly concealed or occluded by another, it is possible that a curve fragment is a

mixture of contours from different objects, as is shown in Figure 1c. Underlying texture can also cause problems.

Curve fragments thus should be grouped in some appropriate way to enable recognition. Such groups can consist of points or curve pieces and are called *segments*. This is illustrated in Figure 1c-f. Some pairs of points on the curve have a common tangent, a *bitangent*. This property is preserved under projections, and is thus suitable to cut the curve into segments which are in correspondence. Such segments are shown in Figure 1e and f. The bitangent, dotted in Figure 1e and f, together with the curve segments also form the boundary of a compact region. The term *configuration* will be used for a combination of segments.

It is an open question how to select configurations that can be used for recognition. By choosing a small curve segment it is more likely that it belongs to one object, but a small segment has a weak discriminatory power. A larger curve segment with several concavities has more discriminatory power but it may originate from several objects.

Smoothing

Edges extracted from an image, using a Canny edge detector, are polygonal edge-trains with orientations 0° , 45° , 90° or 135° . This is particularly noticeable in Figure 3c and d on page 20. Sometimes it is desirable to extract differential properties from these edges. A smoother representation is then needed. This can be achieved using a spline approximation of the curve, or by using a smoothing scheme. Smoothing schemes that commute with affine transformations have recently been developed by [ST1, ST2] and [AL1]. In our work we have emphasized methods that are robust to errors even in the first derivative. These methods do not require smooth curves as input.

2.3 Camera geometry

Three different camera models are discussed in this section. The first one uses R^2 as a model of the image plane. The usual extension of this model is the projective camera model. In this the camera plane is extended to form the projective plane P^2 . A third model is the spherical camera model, using a sphere S^2 as a model for the image ‘plane’. From these models the group of affine transformations of the plane and the group of linear transformation of the sphere, are derived.

Pinhole camera model

Consider a pinhole camera with ideal lens, focal length one, focus at the origin and an image plate $\pi = \{(\hat{y}_1, \hat{y}_2, 1) \in R^3 \mid (\hat{y}_1, \hat{y}_2) \in [a, b] \times [c, d]\}$. The central projection of a point

$$\hat{X} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) \in \{R^3 \mid \hat{x}_3 > 0, \quad a \leq \hat{x}_1/\hat{x}_3 \leq b, \quad c \leq \hat{x}_2/\hat{x}_3 \leq d\}$$

onto the plate along the half-line through the focus is

$$(\hat{x}_1/\hat{x}_3, \hat{x}_2/\hat{x}_3, 1).$$

or in image coordinates

$$\hat{Y} = (\hat{y}_1, \hat{y}_2) = (\hat{x}_1/\hat{x}_3, \hat{x}_2/\hat{x}_3) \tag{1}$$

First notice that points on the same half-line from the center of the lens is projected to the same point on the plate. Notice also that some points are not projected onto the plate at all. This incompleteness in the model is often inconvenient. Even if we extend the plate to an infinite plane, i.e. R^2 , the transformation is not defined for points with $\hat{x}_3 \leq 0$. We will later consider two extensions to this model.

Equation (1) describes the imaging transformation in the case when the image plane and the object are described in the same coordinate system. Since this is not possible in practice, it is desirable to uncouple these systems. For any other coordinate system in the scene holds

$$\hat{X} = RX + t,$$

for some 3×3 rotation matrix R and translation vector t . New coordinate system in the image plane are introduced by

$$Y = A\hat{Y} + b,$$

where A is a 2×2 matrix and b a translation vector. Altogether, this means that (1) is equivalent to

$$X \mapsto Y = \left(\frac{m_{11}x_1 + m_{12}x_2 + m_{13}x_3 + m_{14}}{m_{31}x_1 + m_{32}x_2 + m_{33}x_3 + m_{34}}, \frac{m_{21}x_1 + m_{22}x_2 + m_{23}x_3 + m_{24}}{m_{31}x_1 + m_{32}x_2 + m_{33}x_3 + m_{34}} \right). \quad (2)$$

Let in particular ω be a planar configuration described by the coordinates (x_1, x_2) . Then the imaging transformation can be written

$$X = (x_1, x_2) \mapsto Y = \left(\frac{m_{11}x_1 + m_{12}x_2 + m_{13}}{m_{31}x_1 + m_{32}x_2 + m_{33}}, \frac{m_{21}x_1 + m_{22}x_2 + m_{23}}{m_{31}x_1 + m_{32}x_2 + m_{33}} \right). \quad (3)$$

We will call this a **physically realisable** transformation if

$$m_{31}x_1 + m_{32}x_2 + m_{33} > 0, \quad \forall (x_1, x_2) \in \omega.$$

For certain physically realisable transformations the relative depth in a configuration with respect to the image plane is small, i.e.

$$m_{31}x_1 + m_{32}x_2 + m_{33} \approx C, \quad (x_1, x_2) \in \omega \quad (4)$$

If this is the case, the transformation acting on this configuration, can be approximated by a planar affine transformation.

$$Y = (y_1, y_2) \approx \left(\frac{m_{11}x_1 + m_{12}x_2 + m_{13}}{C}, \frac{m_{21}x_1 + m_{22}x_2 + m_{23}}{C} \right). \quad (5)$$

By identifying points in the image plane and the object plane through their coordinates we get the group of *proper affine transformations of the plane*

$$G_{aff} = \left\{ g = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ 0 & 0 & 1 \end{pmatrix} \in R_{3 \times 3} \mid \det(g) > 0 \right\}.$$

where each element acts on subsets of R^2 according to

$$g(\omega) = \left\{ \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} g_{13} \\ g_{23} \end{pmatrix} \mid \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \omega \right\}.$$

These transformations are called proper since they preserve orientation.

Projective camera model

The projective camera model is the one most used in the computer vision literature. It is often computationally convenient since it benefits from the machinery of projective geometry. Some disadvantages will be pointed out below. In this model the 3-dimensional object space is

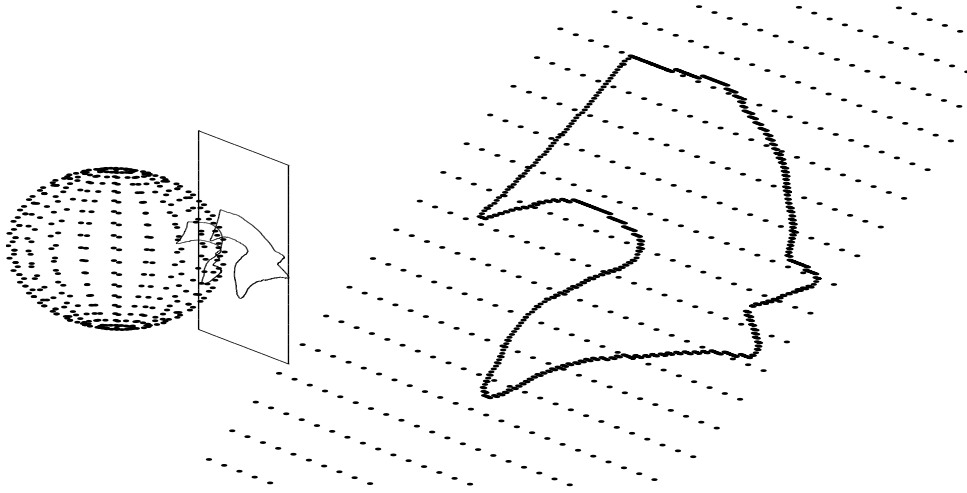


Figure 2: The figure illustrates the central projection of a planar configuration onto a planar image and the projection onto the sphere in the spherical camera model. The image of the planar configuration after a rigid motion is related to the original image by a linear transformation of the sphere.

embedded in a projective space P^3 , and the image plane in P^2 by the assignment of **homogeneous coordinates**

$$X = (x_1, x_2, x_3) \mapsto t(x_1, x_2, x_3, 1), \quad t \in R,$$

$$Y = (y_1, y_2) \mapsto t(y_1, y_2, 1), \quad t \in R.$$

Here separate coordinate systems in the scene and image plane are allowed. Observe that a ‘point’ in the extended image plane P^2 is a line through the origin in R^3 . Such a point with homogeneous coordinates $t(y_1, y_2, y_3)$, $y_3 \neq 0$, corresponds to a point in the ordinary image plane $(y_1/y_3, y_2/y_3)$. But P^2 also contain so called points at the infinity, i.e. points with homogeneous coordinates $(y_1, y_2, 0)$. These can be thought as points infinitely far off in direction $\pm(y_1, y_2)$.

By an imaging transformation is meant a mapping

$$Y = MX$$

where M is a 4×3 matrix of rank 3. Since we are working with homogeneous coordinates, the transformations given by M and tM with $t \neq 0$ must be considered equal. To see the geometric significance, a modified singular value decomposition of M gives

$$Y = A^{-1}PSX,$$

with

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Hence making the coordinate transformation $\hat{Y} = AY$ and $\hat{X} = SX$, we get,

$$\hat{Y} = P\hat{X},$$

or

$$\hat{X} = t(\hat{x}_1, \hat{x}_2, \hat{x}_3, 1) \mapsto t(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \hat{Y}, \quad (6)$$

which, for $\hat{x}_3 > 0$ is the same as (1), but written in homogeneous coordinates.

Spherical camera model

In the projective camera model, the two different points $t(x_1, x_2, x_3, 1)$ in front of the camera and $t(-x_1, -x_2, -x_3, 1)$ behind the camera are mapped onto the same point $t(x_1, x_2, x_3)$ in the image plane. This is often inconvenient. In the spherical model the 3-dimensional object space is imbedded in S^3 , and the image plane in sphere S^2 , by the assignments of **directional homogeneous coordinates**

$$X = (x_1, x_2, x_3) \mapsto t(x_1, x_2, x_3, 1), \quad t > 0,$$

$$Y = (y_1, y_2) \mapsto t(y_1, y_2, 1), \quad t > 0.$$

Again separate coordinate systems in the scene and image plane are allowed. Observe that a ‘point’ in the extended image plane S^2 is a half-line from the origin.

Just as in the projective case an imaging transformation is a mapping

$$Y = MX,$$

But now two matrices correspond to the same transformation if and only if one is a positive multiple of the other, i.e.

$$M \sim M' \iff M = tM', t > 0 \quad (7)$$

For planar configurations the transformation between object directional homogeneous coordinates $X = t(x_1, x_2, x_3)$, $t > 0$ and image coordinates is

$$\begin{aligned} X &= t(x_1, x_2, x_3) \mapsto \\ Y &= t(m_{11}x_1 + m_{12}x_2 + m_{13}x_3, m_{21}x_1 + m_{22}x_2 + m_{23}x_3, m_{31}x_1 + m_{32}x_2 + m_{33}x_3). \end{aligned}$$

Concepts like positive, negative and zero determinant are well defined on equivalence classes, (7), since

$$\text{sign}(\det(tM)) = \text{sign}(t^3 \det(M)) = \text{sign}(\det(M)), \quad \text{if } t > 0$$

A matrix M with positive determinant corresponds to a viewing situation where both images are taken from the same side of the plane. A singular matrix M corresponds to the singular viewing situation where the focus lies in the same plane as the configuration. A negative determinant corresponds to the situation where the images are taken from different sides. One image is a mirror image of the other. By omitting the last two cases the following group of transformations of the sphere is obtained,

$$G_{\text{sphere}} = \{g \in (R^{3 \times 3} \setminus 0) / \sim \mid \det(g) > 0\}$$

where the equivalence relation is defined by (7). This group will be called *the group of proper linear transformations of the sphere*. Each element g of the group acts on subsets of the sphere according to

$$g(X_1) = \{y \in S^2 \mid \exists x \in X_1, y \sim gx\}.$$

There is no difference between these three model as long as physically realisable transformations are considered. So in practice (3) is very useful. There are several reasons for using the spherical model instead of the projective for our purposes.

- A distinction is made between points ‘in front’ of the camera and points ‘behind’.
- Many vision systems actually have all-around sight and are therefore more accurately modeled with S^2 .
- Using P_R^2 , there is no notion of mirror image. Positive or negative determinant is not well defined since a matrix M is equivalent to its negative $-M$.
- For subsets of S^2 that lie on one half of the sphere there is a natural notion of convex hull. This is not well defined for a subset of P_R^2 .

3 Normalization principles

The normalization principle and some notations are introduced in this section. Normalization is presented as a way to construct invariants. This can be viewed as introducing an abstract coordinate system, on the set of configurations. It is shown how the normal reference frame induces a metric on equivalence classes of configurations. The extraction of invariant indices from a, possibly non-unique, normal reference frame is also discussed.

It is relatively easy to construct invariants of objects consisting of a finite combination of points and lines. A survey of different methods can be found in [Gr1, MZ1]. It is also relatively easy to supply the configuration space, usually R^n , P_R^n or S^n , with full topological and differentiable structures. The probabilistic behaviour of invariant features can also be studied, relatively easy.

One difficulty, when working with curves and regions is that they have much more complicated structure than point configurations. We will first consider objects as elements of an abstract set Ω on which a group G of transformations act. It is still meaningful to talk about invariants and normalization under group action, without introducing additional structure. By adding topology on Ω and differentiable structure on the group G , it is possible to discuss continuity. It is also interesting to study the probabilistic behaviour of invariants.

3.1 Group action

A group G is said to act on a set Ω if there exists a mapping

$$(G, \Omega) \ni (g, \omega) \longrightarrow g(\omega) \in \Omega$$

with properties

$$\begin{aligned} 1(\omega) &= \omega, & \forall \omega \in \Omega, \\ g_1(g_2(\omega)) &= (g_1g_2)(\omega), & \forall \omega \in \Omega, \quad \forall g_1, g_2 \in G. \end{aligned}$$

The notation for group action is either $g\omega$ or $g(\omega)$.

Two elements ω_1 and ω_2 are said to have the same *shape* with respect to the group G if $\omega_1 = g\omega_2$ for some transformation $g \in G$. This is an equivalence relation, because of the group structure of G . We write

$$\omega_1 \sim \omega_2 \iff \exists g \in G, \quad \omega_1 = g\omega_2 \tag{8}$$

The equivalence relation divides Ω into disjoint equivalence classes. Denote the equivalence class containing ω by $G\omega = \{g\omega | g \in G\}$. This is also called the *orbit* of ω under the group action G .

Let $T : \Omega \longrightarrow W$ be a function defined on Ω with values in some feature set W . This function is called an **invariant** if

$$\omega_1 \sim \omega_2 \implies T(\omega_1) = T(\omega_2) \quad (9)$$

An invariant is called **complete** if

$$\omega_1 \sim \omega_2 \iff T(\omega_1) = T(\omega_2) \quad (10)$$

In computer vision invariants are used to determine the shape, under some transformation group, of a configuration in the image. Completeness is an important property. For complete invariants no information about shape is lost when going from ω to $T(\omega)$. Any function of $T(\omega)$ is also an invariant. Furthermore all invariant features of ω can be calculated from $T(\omega)$. Useful invariants should also be easy to compute and be stable under 'small' distortions in ω , where the notion of small will be commented upon below.

Example. The trivial mapping $T \equiv 0$ is an invariant, but it cannot be used to discriminate between any elements of different shape. \square

Example. The mapping $T : \omega \longrightarrow G\omega$ is a complete invariant. It is however difficult to work with if the set $G\omega$ is large. \square

Example. If Ω is the set of ordered linear 4-point configurations (a, b, c, d) , then the cross ratio

$$\text{cross ratio} = \frac{a-c}{b-c} / \frac{a-d}{b-d} \quad (11)$$

is a complete invariant under the projective transformation group. \square

3.2 Normalization

Normalization schemes can be constructed using properties that change 'a lot' under transformations in G . These latter properties will make it possible to select a small number of representatives from each equivalence class, defining the invariants.

A typical such property can be obtained in the following way. Take any mapping $P : \Omega \longrightarrow R^n$. This will in general not be an invariant. The element ω is said to have the *property* P if $P(\omega) = 0$. Assume that P can be chosen so that at least one element from each equivalence class has the property P , and let

$$\Omega_P = \{\omega | P(\omega) = 0\}.$$

Then projection of ω onto Ω_P along the equivalence class $G\omega$,

$$T(\omega) = G\omega \cap \Omega_P, \quad (12)$$

is a complete invariant. Here $T(\omega)$ is a set-valued function, containing the representatives for the equivalence class $G(\omega)$. In the sequel, the term **normal** reference frame will be used for any one element in Ω_P .

Assume for simplicity that $T(\omega)$ contains only one element. An element ω_1 can then be represented as

$$\omega_1 = g_1 \omega_1^{inv} \quad (13)$$

with $g_1 \in G$ and $\omega_1^{inv} = T(\omega_1)$. The element ω_1 is thus represented as a product of two factors, one transformation and one element of Ω . In this way an abstract coordinate system is

introduced on $\Omega = G \times \Omega_P$. An element ω_1 has ‘shape coordinate’ ω_1^{inv} and ‘group coordinate’ g_1 . Notice that ω_1^{inv} carries all information about the shape of ω_1 and that g_1 carries all information about where on the equivalence class ω_1 lies. A second element ω_2 is also represented as

$$\omega_2 = g_2 \omega_2^{inv}$$

with $g_2 \in G$ and $\omega_2^{inv} = T(\omega_2)$. These two elements have same shape if and only if $\omega_1^{inv} = \omega_2^{inv}$. If this is the case, then $\omega_1 = g_1 g_2^{-1} \omega_2$, as is seen in the following diagram.

$$\begin{array}{ccc} \omega_1 & \longleftarrow & \omega_2 \\ \uparrow g_1 & & \uparrow g_2 \\ \omega_1^{inv} & \equiv & \omega_2^{inv} \end{array}$$

In this way, we have developed a tool to determine the transformation between two elements of the same shape.

Each choice of Ω_P gives a *normalization scheme*. Such a scheme is useful for recognition even if $T(\omega)$ consists of several elements. If a model object ω is to be recognized, and $T(\omega)$ consist of say three elements $\{\omega^1, \omega^2, \omega^3\}$, then information about the element in its three normal reference frames are stored during the *learning* phase. When ω' is to be recognized, it is normalized into one of its normal reference frames, and recognition can be done using invariant features in this reference frame.

Example. Affine normalization of four points; affine coordinates. Let Ω be the class of ordered 4-point configurations,

$$\omega = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\} \subset R^2.$$

Let G be the group of planar affine transformations acting on ω pointwise, and let the normal reference frames be defined by

$$\Omega_P = \{\omega \mid (x_1, y_1) = (0, 0), (x_2, y_2) = (1, 0) \text{ and } (x_3, y_3) = (0, 1)\}. \quad (14)$$

If the first three points of ω are not collinear there is exactly one representative in each equivalence class. Notice that other normal reference frames are obtained by choosing other points as basis points. \square

3.3 Adding structure: continuity and probabilistic behaviour

So far the set Ω has been considered without any structure. Suppose that we have some notion of how close two elements are to each other, i.e. a metric or topology on Ω . Then it becomes meaningful to discuss continuity of the normalization scheme. In computer vision, G is often a finite dimensional transformation group. It is assumed that we equip G with topological and differentiable structures.

Useful invariants should be complete, continuous and easy to compute. These considerations give claims on the choice of Ω_P , or alternatively, on the features used to define Ω_P . For each $\omega \in \Omega$,

- $G\omega \cap \Omega_P$ should be a small set, but not empty
- $G\omega \cap \Omega_P$ should be insensitive to distortion in ω
- $G\omega \cap \Omega_P$ should be easy to compute

In the sequel we will say that a normalization scheme has the *uniqueness property* if there is only one normal reference frame, i.e. $G\omega \cap \Omega_P$ has only one element. A normalization scheme is called *continuous* if the normal reference frames depend continuously on ω .

One direct result of the implicit function theorem is that g is defined implicitly by $F_\omega(g) = P(g^{-1}\omega) = 0$ and will depend continuously on ω , in a neighbourhood of (g, ω) if the following conditions hold.

- $F_\omega(g) = 0$
- F_ω is continuous in both g and ω
- The differential dF_ω/dg is non-singular and continuous in g and ω

If errors due to discretization and edge detection are small and can be modelled so that the distribution of $P(\omega)$ is known, then the functional matrix dF_ω/dg can be used to estimate the robustness of the normal reference frame.

3.4 Group invariant metrics

Assume that d is a metric on Ω . Assume also that we have a normalization scheme with uniqueness. Then a G-invariant metric on the class of shapes is defined by

$$d_G(\omega_1, \omega_2) = d(T(\omega_1), T(\omega_2)) \quad (15)$$

In the sequel the following **Hausdorff metric** will be used for compact subsets of R^2 .

$$d(\omega_1, \omega_2) = \max_{z_1 \in \omega_1} \min_{z_2 \in \omega_2} \|z_1 - z_2\| + \max_{z_2 \in \omega_2} \min_{z_1 \in \omega_1} \|z_1 - z_2\| \quad (16)$$

where $\|z\|$ is the Euclidean norm.

Rotations

Rotationally symmetric shapes are difficult to normalize with respect to rotations. Any rotational normalization scheme will have trouble with shapes that are close to rotational symmetry. On the other hand rotations and translations do not affect Euclidean distance between a pair of points. It is therefore possible to define a metric on shapes with respect to the group G_{rot} of rotational transformations

$$G_{rot} = \left\{ g = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \mid \theta \in R \right\}$$

acting on R^2 by left multiplication. The following induced Hausdorff metric will do

$$d_{rot}(\omega_1, \omega_2) = \min_{g \in G_{rot}} d(g\omega_1, \omega_2) \quad (17)$$

Normalization schemes can be constructed for compact subsets of R^2 with respect to rotations. One example is

$$\Omega_P = \left\{ \omega \mid \max_{z \in \omega} \|z\| = \max_{z \in \omega \cap (R_+, 0)} \|z\| \right\} \quad (18)$$

It is clear that some subsets have several possible *normal* reference frames. It is also clear that the normal reference frames are sensitive to distortions.

3.5 Invariant indices

Object recognition can be done using invariant metrics. But this requires a linear search through all the objects, $\omega_1, \dots, \omega_N$, calculating the distances $d_G(\omega, \omega_1), \dots, d_G(\omega, \omega_N)$. The real use of normalization schemes is the possibility of extracting invariant indices that can index a recognition table directly.

Once in a normal reference frame any feature defines an invariant. For curves or regions, moments or Fourier coefficients can be used for this purpose. Another very simple but effective idea for extracting features from a region is the following. Let $R(\phi) \subset R^2$ be the sector

$$R(\phi) = \{x | x = (r \cos(\alpha), r \sin(\alpha)) \text{ with } r > 0, 0 < \alpha < \phi\}$$

and let

$$A_\omega(\phi) = \int_{R(\phi) \cap \omega} dx.$$

The feature $A_\omega(\phi)$ is thus the area of ω that lies in a ϕ -sector with vertex at the origin. If ω is star shaped or convex, no information is lost when going from ω to $A_\omega(\phi)$. For such regions we have the following relation.

$$A'_\omega(\phi) = r(\phi)^2/2$$

where $r(\phi)$ is the radius of the region at angle ϕ . For a general region $A_\omega(\phi)$ does not contain all information about the original curve but it is still something that can be used to discriminate between different objects. In the experiment in Sections 4 and 5, $n = 101$ bins were used. Each bin represents one sector and contains

$$a(k) = A\left(\frac{k+1/2}{2n\pi}\right) - A\left(\frac{k-1/2}{2n\pi}\right)$$

These n features are then used to discriminate between different shapes. Usually 2 or 3 linear combinations of these 101 numbers are used as invariant indices. The linear combinations are chosen or designed to separate the model features as much as possible, e.g. using singular value decomposition.

4 Affine normalization

Affine invariant descriptors have been studied earlier. The most common approach is to use affine coordinates with respect to three fiducial points, cf. (14) and [LS1]. One disadvantage of this method is that three points have to be selected and ordered in some manner. The affine invariant features depend crucially on this choice. Another idea is to use Fourier coefficients of a closed curve and calculate invariant features from these, cf. [Ar1, ASBH1].

The present work has been inspired by ideas used in algorithms that find shape from texture, cf. [DH1, Wit1, BY1, BM1, Gã1]. Two ideas, maximal compactness and weak isotropy, used in these algorithms, are presented in Section 4.1. It is shown how they can be used for affine normalization of curve segments. The normal reference frames obtained in this way are quite robust with respect to random distortions, but they require that the first derivative of a curve is estimated. The normalization is therefore not continuous in the Hausdorff metric.

Another idea, using integral instead of differential features, is presented in Section 4.2. This method uses moments of segmented regions to select the normal reference frames. This scheme is continuous with respect to the Hausdorff metric. The normal reference frame is unique up to rotation and can be computed directly from the first three moments of the original region. The robustness of the normalization scheme and the discriminatory power of invariant indices are illustrated with experiments.

4.1 Maximal compactness and weak isotropy

In this subsection Ω will be the set of piecewise smooth curves in R^2 with finite arc-length. It will be assumed that the curves can be parametrized as

$$\omega = \left\{ \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \mid t \in [t_0, t_1] \right\}$$

for some piecewise differentiable functions x_1 and x_2 . The group G of transformations acting on Ω will be the group of proper affine transformations, cf. Section 2.3.

For each curve it is possible to calculate a relative distribution of directions $f(\alpha)$ such that $f(\alpha)d\alpha$ is the fraction of the arc-length for the set of points of the curve where the direction of the tangent lies in the interval $[\alpha, \alpha + d\alpha]$. The curve is not oriented, so the directions α and $\alpha + \pi$ are considered identical. One can show that the directional distribution is independent of the choice of parametrization. It is positive and normalized so that

$$\int_0^\pi f(\alpha)d\alpha = 1 \tag{19}$$

The distribution $f(\alpha)$ is a π -periodic function which can be described by its Fourier series

$$f(\alpha) = \frac{1}{\pi} + \frac{2}{\pi} \sum_{k \geq 1} (a_k \cos(2k\alpha) + b_k \sin(2k\alpha)) = \frac{1}{\pi} + \frac{2}{\pi} \sum_{k \geq 1} A_k \cos(2k\alpha + d_k)$$

where A_k are positive amplitudes and d_k are phase shifts. In particular a_1 and b_1 will be used later. Let ω be a planar, piecewise differentiable parametric curve, given by

$$\omega(t) = \left\{ \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \mid t \in [t_0, t_1] \right\}$$

Then, cf. [Gå1], the coefficients a_1 and b_1 are given by

$$\begin{cases} a_1(\omega) = \frac{1}{L} \int_{t_0}^{t_1} \frac{x_1^2 - x_2^2}{\sqrt{x_1^2 + x_2^2}} dt \\ b_1(\omega) = \frac{1}{L} \int_{t_0}^{t_1} \frac{2x_1x_2}{\sqrt{x_1^2 + x_2^2}} dt \end{cases} \tag{20}$$

where $L = \int_{t_0}^{t_1} \sqrt{\dot{x}_1^2 + \dot{x}_2^2} dt$ is the arc-length of ω .

The coefficients a_k, b_k or A_k, d_k transform in a simple way when the curve is translated, rotated and scaled. A translation does not change the distribution at all, neither does a change of scale. A rotation changes the phase shifts. A curve that is rotationally symmetric of order n has $A_k = 0$, when k is a multiple of n . A circle has constant directional distribution and all A_i 's are zero. This case where all directions are equally represented, is called *strong isotropy* and has been treated by [Wit1].

In the literature some related notions can be found:

- A curve is called **weakly isotropic** if $A_1 = 0$, i.e. $a_1 = b_1 = 0$. This condition has successfully been used by J. Gårding, to determine the shape of a surface using surface markings, see [Gå1].
- A *closed curve* is said to have **maximal compactness** if the ratio of the square of the arc-length and the area enclosed is minimal with respect to a set of transformations. Neither translation, rotation nor a change of scale affect this ratio. It is therefore possible to extend the concept of maximal compactness to all curve segments. A *curve segment* is said to have **maximal compactness** with respect to affine transformations if the arc-length is minimal with respect to a all area preserving affine transformations.

It can be shown that $a_1 = 0$ and $b_1 = 0$ are equivalent to the condition that a curve has local maximal compactness with respect to affine transformations, i.e. weak isotropy is equivalent to local maximum compactness. This follows from an observation that a_1 and b_1 are the partial derivatives of arc-length, with respect to area preserving affine transformations, for the right choice of parametrization.

The global properties are now examined. These are based on some facts that have been known in the shape from texture literature for some time. Since formal proofs are difficult to find in the literature, we find it worthwhile to write them down.

By transforming a curve with a stretch transformation

$$S_{(k,\theta)} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

it is possible to change the directional distribution. Such a transformation gives more weight to the curve patches whose tangential direction is close to θ . By rotating the coordinate system all stretch transformation have the form

$$S_{(k,0)} = \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$$

It will now be investigated how the coefficient a_1 changes under such transformations.

Lemma 1 *If $\omega \in \Omega$ is non-linear, then*

$$\frac{d}{dz}(a_1(S_{(z,0)}\omega)) > 0, \quad \forall z > 0. \quad (21)$$

Proof: In this proof, we will temporarily write $a_1(1,\omega)$ instead of $a_1(\omega)$, and more generally define

$$a_1(z,\omega) = a_1(S_{(z,0)}\omega) = \frac{\int \frac{z^2 \dot{x}_1^2 - \dot{x}_2^2}{\sqrt{z^2 \dot{x}_1^2 + \dot{x}_2^2}} dt}{\int \sqrt{z^2 \dot{x}_1^2 + \dot{x}_2^2} dt}$$

using (20). The functions x_1 and x_2 and the integration limits are given by the choice of parametrization. Notice however that $a_1(z,\omega)$ is independent of this choice. Differentiation gives

$$\left. \frac{\partial a_1}{\partial z}(z,\omega) \right|_{z=k} = \frac{\int (2k\dot{x}_1^2/s - (k^2\dot{x}_1^2 - \dot{x}_2^2)k\dot{x}_1^2/s^3) dt \int s dt - \int (k^2\dot{x}_1^2 - \dot{x}_2^2)/s dt \int k\dot{x}_1^2/s dt}{(\int s dt)^2} \quad (22)$$

where we have used $s = \sqrt{k^2\dot{x}_1^2 + \dot{x}_2^2}$. Notice that the denominator is positive. It is easy to show that

$$\left. \frac{\partial a_1}{\partial z}(z, S_{(k,0)}^{-1}\omega) \right|_{z=k} = \frac{1}{k} \left. \frac{\partial a_1}{\partial z}(z,\omega) \right|_{z=1}.$$

Since $S_{(k,0)}$ is a bijective map, to prove

$$\left. \frac{\partial a_1}{\partial z}(z,\omega) \right|_{z=k} > 0$$

for every $\omega \in \Omega$ and every $k > 0$, is equivalent to prove

$$\left. \frac{\partial a_1}{\partial z}(z,\omega) \right|_{z=1} > 0$$

for every $\omega \in \Omega$. It is thus sufficient to study the numerator N of (22) for $z = k = 1$, which using $s^2 = \dot{x}_1^2 + \dot{x}_2^2$ becomes

$$N = \int \underbrace{(2\dot{x}_1^2(\dot{x}_1^2 + \dot{x}_2^2) - (\dot{x}_1^2 - \dot{x}_2^2)\dot{x}_1^2)}_{\dot{x}_1^2 s^2 + 2\dot{x}_1^2 \dot{x}_2^2} / s^3 dt \int s dt - \int (\dot{x}_1^2 - \dot{x}_2^2) / s dt \int \dot{x}_1^2 / s dt$$

or

$$N = \int \dot{x}_1^2 / s dt \int s dt + 2 \int \dot{x}_1^2 \dot{x}_2^2 / s^3 dt \int s dt - (\int \dot{x}_1^2 / s dt)^2 + \int \dot{x}_1^2 / s dt \int \dot{x}_2^2 / s dt$$

The first term of the right hand side can be written

$$\int \dot{x}_1^2 / s dt \int s dt = \int \dot{x}_1^2 / s dt \int (\dot{x}_1^2 + \dot{x}_2^2) / s dt = (\int \dot{x}_1^2 / s dt)^2 + \int \dot{x}_1^2 / s dt \int \dot{x}_2^2 / s dt$$

It follows that

$$N = 2 \int \dot{x}_1^2 \dot{x}_2^2 / s^3 dt \int s dt + 2 \int \dot{x}_1^2 / s dt \int \dot{x}_2^2 / s dt$$

which obviously is non negative. The derivative $\frac{\partial a_1}{\partial z}$ is zero if and only if $\dot{x}_1 = 0$ at every point or $\dot{x}_2 = 0$ at every point of the curve. This means that the curve is linear. ■

It is easy to see that a weakly isotropic curve remains weakly isotropic under similarity transformations. The following lemma tells that these are the only affine transformations with this property.

Lemma 2 *A proper affine transformation maps a weakly isotropic curve into a weakly isotropic curve if and only if it is a similarity transformation.*

Proof: Every proper affine transformation may be written $x \rightarrow Oa(Sx + b)$, where O is a rotation matrix, a is a change of scale, S a stretch transformation and b a translation. Translations and scale changes do not affect the directional distribution. A rotation changes d_1 but not A_1 . Thus if ω is a weakly isotropic curve, then $Oa(S\omega + b)$ is weakly isotropic if and only if $S\omega$ is weakly isotropic. To investigate when the latter happens, rotate the coordinate system so that

$$S = S_{(\kappa, 0)} = \begin{pmatrix} \kappa & 0 \\ 0 & 1 \end{pmatrix}$$

It is sufficient to study the function $a_1(S_{(\kappa, 0)}\omega)$. By Lemma 1, it is strictly increasing for $\kappa > 0$. By the assumption, ω is weakly isotropic, so it vanishes for $\kappa = 1$. Hence $a_1(S_{(\kappa, 0)}\omega) = 0$ if and only if $\kappa = 1$, i.e. $S = I$. This proves the assertion. ■

We are now ready to use weak isotropy to construct invariants. The following lemma will be useful.

Lemma 3 *Given a non-linear curve $\omega \in \Omega$, there is a unique stretch transformation S so that $S\omega$ is weakly isotropic.*

Proof:

We first show the **existence** of a stretch transformation S that makes ω weakly isotropic. It is easy to see that for any $\omega \in \Omega$, holds $A_1(S_{(k, \theta)}\omega) \rightarrow 1$ pointwise in θ as $k \rightarrow \infty$. Intuitively the reason is that a higher fraction of the arc-length has a tangential direction close to the direction θ as k is increased. The set of directions is compact and the limiting function is continuous so the convergence is uniform in θ .

For any $\omega \in \Omega$ we have

$$A_1(\omega) = \sqrt{a_1^2 + b_1^2}$$

where

$$a_1(\omega) = \int_0^\pi f(\alpha) \cos 2\alpha \, d\alpha,$$

$$b_1(\omega) = \int_0^\pi f(\alpha) \sin 2\alpha \, d\alpha.$$

By means of (19) and the assumption that ω is non-linear it follows that $A_1(\omega) = \beta < 1$ for some β . Let Σ denote the set of stretch transformations. Consider the subset Σ_K of Σ

$$\Sigma_K = \{S_{(k,\theta)} | k \leq K\},$$

where K is chosen so that $A_1(S_{(k,\theta)}\omega) > \frac{1+\beta}{2}$ for all $k > K$ and all θ . This is possible by the uniform convergence. Using the Frobenius norm on the set of matrices, Σ_K is compact. By the construction holds

$$\inf_{S \in \Sigma} A_1(S\omega) = \inf_{S \in \Sigma_K} A_1(S\omega)$$

Since Σ_K is compact there exists a stretch transformation S_{\min} so that

$$A_1(S_{\min}\omega) = \inf_{S \in \Sigma} A_1(S\omega).$$

If this minimal A_1 is greater than zero, a contradiction can be obtained in the following way. Rotate the coordinate system so that $a_1 < 0$ and $b_1 = 0$. Study how $A_1(S_{(z,0)}S_{\min}\omega)$ changes with z , where $S_{(z,0)}$ is defined in this new coordinate system. According to Lemma 1 we have $\frac{d}{dz}(a_1) > 0$ so that

$$\frac{d}{dz}(A_1^2) = \frac{d}{dz}(a_1^2 + b_1^2) = 2a_1 \frac{d}{dz}(a_1) + 2b_1 \frac{d}{dz}(b_1) < 0,$$

contradicting the minimality. Hence $S_{\min}\omega$ is weakly isotropic.

The stretch transformation S minimizing $S\omega$, is also **unique**. To see this assume that both $S_1\omega$ and $S_2\omega$ are weakly isotropic. Then, by Lemma 2, $S_1 \circ S_2^{-1}$ is equal to a similarity transformation Oa . Now $S_1 \circ S_2^{-1} = Oa$ gives $S_1 = OaS_2$. This factorization is unique which implies $a = 1, O = I, S_1 = S_2$. ■

Theorem 1 *Given a curve $\omega \in \Omega$ there exist unique S , a and b so that $a(S\omega + b)$ is weakly isotropic, has mass center in the origin and mean distance one to the origin.*

Proof: First choose the stretch transformation S according to Lemma 3, so that $S\omega$ is weakly isotropic. Then choose b so that the mass center coincides with the origin, and a so that the mean distance to the origin is one. Notice that $a(S\omega + b)$ remains weakly isotropic, with mass center in the origin. ■

An affine normalization for planar smooth curves can be constructed taking Ω_P as the planar non-linear smooth curves with finite arc-length that (i) have mass centers at the origin, (ii) are weakly isotropic, (iii) have mean distance one to the origin and (iv) the maximum distance from the origin to the curve occurs on the positive x_1 -axis. According to Theorem 1 the translation, stretch transformation and scale change component of the affine transformation are uniquely determined by properties (i-iii) of Ω_P . For some images the rotation is not uniquely determined by the last property, in which case $T(\omega)$ consists of more than one element. Curves that are affinely normalized using this scheme are shown in Figure 3c.

4.2 Affine normalization using moments

In this section, let

$$\Omega = \{\omega \subset \mathbb{R}^2 \mid \omega \text{ is compact with positive area}\}.$$

The results below generate corresponding results for the closed boundary curves of such regions. A typical example is an extracted concavity as in Figure 11.

For $\omega \in \Omega$, let the **moments** be defined as

$$\begin{aligned} m_0(\omega) &= \int_{x \in \omega} dx \\ m_1(\omega) &= \int_{x \in \omega} x dx \\ m_2(\omega) &= \int_{x \in \omega} xx^T dx \end{aligned} \tag{23}$$

Notice that m_0 is a scalar, m_1 a vector and m_2 a matrix. Notice also that these quantities depend continuously on ω in the Hausdorff metric. The moments change in a simple way when a region is transformed. For instance we have

$$\begin{aligned} m_1(\omega + b) &= \int_{y \in \omega + b} y dy = \int_{x \in \omega} (x + b) dx = m_1(\omega) + bm_0(\omega) \\ m_1(A\omega) &= \int_{y \in A\omega} y dy = \int_{x \in \omega} Ax |\det A| dx = |\det(A)| Am_1(\omega) \\ m_2(A\omega) &= \int_{y \in A\omega} yy^T dy = \int_{x \in \omega} Ax(Ax)^T |\det A| dx = Am_2(\omega)A^T |\det(A)| \end{aligned}$$

Therefore it is easy to use the moments to select representatives from each equivalence class.

Theorem 2 *Given $\omega \in \Omega$, there is a proper affine transformation unique up to rotations that transforms ω into a region $\omega' \in \Omega_P$, where*

$$\Omega_P = \{\omega \mid m_0(\omega) = 1, \quad m_1(\omega) = 0, \quad m_2(\omega) = aI, \quad a \in \mathbb{R}\}. \tag{24}$$

Furthermore the complete invariant $T(\omega) = G_{aff}\omega \cap \Omega_P$ is a continuous mapping from (Ω, d) to $(\Omega/G_{rot}, d_{rot})$.

Proof: The condition $m_1(\omega + b) = 0$ gives

$$m_1(\omega + b) = m_1(\omega) + bm_0(\omega) = 0.$$

Since $m_0(\omega) \neq 0$, b is uniquely determined as

$$b = -\frac{m_1(\omega)}{m_0(\omega)}.$$

Assuming that $m_1(\omega) = 0$, A has to be chosen so that the second moment is the identity matrix. Observe that $m_1(\omega) = 0$ implies that $m_1(A\omega) = 0$, i.e. the mass center is not affected by multiplication with A . The condition

$$m_2(A\omega) = |\det(A)| Am_2(\omega)A^T = aI$$

gives

$$m_2(\omega) \sim BB^T$$

with $B = A^{-1}$. Observe that $\det(m_2(\omega)) = \det(B)^4$. Let $|B|$ be the positive square root of the positive definite matrix BB^T , i.e. $|B|^2 = BB^T$. Then

$$|B| = \sqrt{m_2(\omega)}.$$

The matrix A is thus given as

$$A = |B|^{-1}.$$

It is determined uniquely up to rotation and scale. Finally fix the scale by $m_0(\omega') = 1$. It is easy to see that all transformations are continuous in the Hausdorff metric. ■

The normalization scheme based on (24) has the following properties.

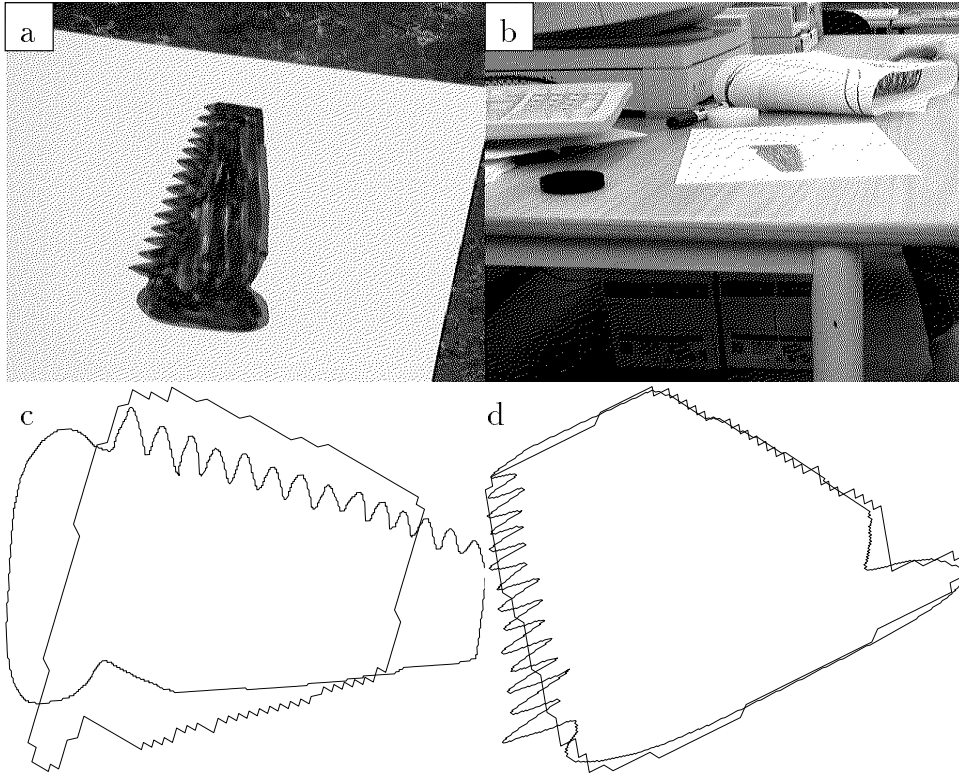


Figure 3: 3a and 3b. Images of a sawlike planar curve. In 3b, interesting features disappear because of low resolution. 3c: The two curves can be affinely normalized using maximum compactness or weak isotropy, but the normal reference frame depends crucially on how the curve is approximated. 3d: Moment based normalization on the other hand is very robust to these kind of digitization errors. Notice the stability of this normal reference frame.

- Uniqueness. In (13), g and ω^{inv} are unique up to rotation.
- Continuity. Both g and ω^{inv} depend continuously on ω .
- Easy to compute. As can be seen from the proof of Theorem 2, the transformation g can be directly computed from the moments of order 0, 1 and 2 of the region ω .
- Robust to digitization errors.
- No distinguished points are needed.

This scheme can be very useful in the recognition of planar curve segments, obtained from concavities. The affine approximation (4) is often valid since these concavities often occupy a small region in the image. Examples of its use are given in Figures 3, 4 and 5. It can be seen that the method has good robustness properties, in comparison with maximum compactness and weak isotropy. Figure 4 shows different concavities in affine normal reference frames. The concavities were segmented automatically from images as in Figure 1d. The result from the edge detector is jagged because of digitization effects. These edges could be smoothed before

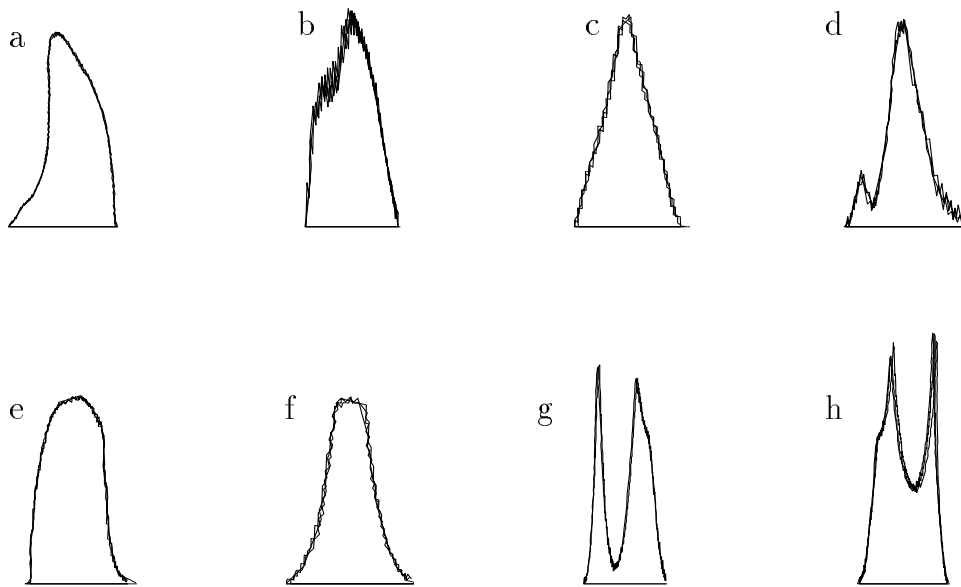


Figure 4: Extracted edges of eight concavities from three different gray-scale images, in their affine normal reference frame. In each of the eight plots there are three superimposed concavities, computed from different images. The edges from the edge detector were not smoothed. These jagged edges do not affect the moment based normalisation scheme. Notice particularly the digitization errors in Figure 4b, 4c, 4d and 4f.

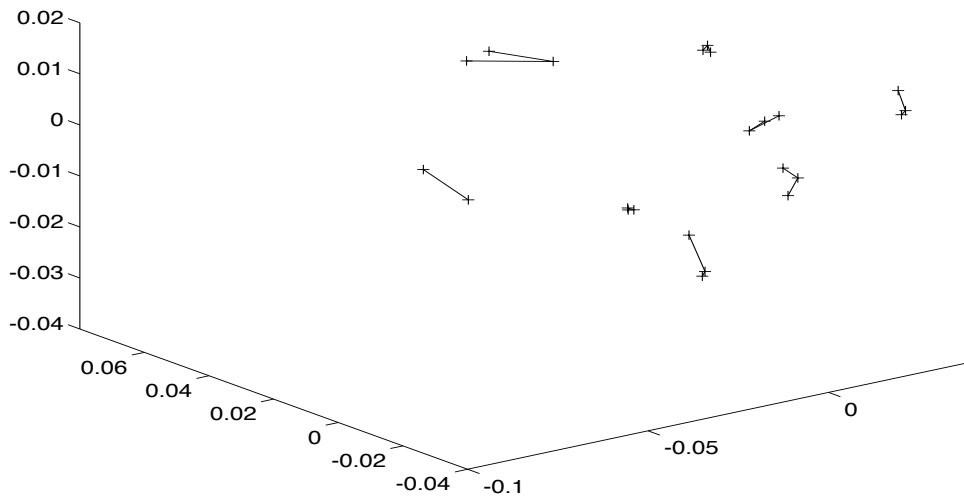


Figure 5: Three principal components of the invariant features extracted from the 24 concavities in Figure 4. These are used to index a recognition table. Corresponding invariant of the same shape are joined with a line. Notice the nice separation between the eight different shapes.

normalization, but this has not been done in the figures, to illustrate the size of digitization errors in different images. Figure 5 is a plot of invariant indices extracted from the concavities in their normal reference frame, cf. Section 3.5.

Completely analogous normalization schemes can be used for point configurations, curve segments and any combination of points, curve segments and regions. It is easy to generalize Theorem 2 to general dimensions.

5 Projective or spherical normalization

Projective normalization will be discussed in this section. Some earlier work is first discussed in Section 5.1. A normalization scheme for point configurations will then be described in Section 5.2. This scheme is continuous and gives a unique representative from each equivalence class. A normalization scheme for compact regions of R^2 under physically realisable transformations is then presented in Section 5.3. It is continuous but does not give a unique representative from each equivalence class. As is seen later, this is not a shortcoming of the normalization scheme, but a consequence of mixing projective equivalence and the Hausdorff metric. Finally these limitations are explained in Sections 5.4 and 5.5, where it is shown that all smooth curves can be projectively transformed into a curve arbitrarily close to a circle in the Hausdorff metric. It is also shown that arbitrarily close to each of a finite number of closed planar curves there is one member of a set of projectively equivalent curves.

5.1 Earlier work

The cross ratio is the classical projective invariant on ordered sets of four points on a line. At first sight it may appear that such simple configurations are not very common in computer vision. However, it is actually used in a commercial navigation system for autonomous industrial vehicle, as described in [Ås1]. The vehicle has a laser scanner that measures relative angles to identical beacons. The system is planar, i.e. all beacons lie in a horizontal plane. The beacons are glued to walls which means that many of them are on a straight line. The measuring device can be considered as a one-dimensional camera. A beacon at a position $X \in R^2$ is projected onto an angle $\beta \in S^1$. The position of the vehicle is not known when the system is initialized. There is then a problem of establishing the correspondence between the signal responses and the beacons. The cross ratio can be used as an invariant index for recognizing four beacons on a line, typically four beacons on the same wall. A system based on this idea has been implemented and tested, cf. [Ås1, Ås2].

It is often important to analyze the discriminatory power of an invariant feature. To do this both the probabilistic behaviour of the invariant features and the background noise have to be investigated. Typical questions are:

- What is the distribution of the cross ratio given that the four points are noisy measurements of the projection of four points with a given cross ratio.
- What is the distribution of the cross ratio given four random measurements.

These questions have been studied in [My1, My2, ÅM1].

A generalization of cross ratios to areas of planar regions can be found in [NS1].

The extraction of projectively invariant features from planar curves has been investigated earlier using the following ideas. In each case the normal reference frame is chosen from point features or derivatives at distinguished points.

1. **Differential invariants**, cf. [We1, Will]. Smooth curves can be recognized using differential invariants. These invariant signatures can in principle be used to recognize curves even when a large part of the curve is occluded. The differential invariants under projective transformations do unfortunately require derivatives of fifth order. There are substantial numerical problems in estimating these high-order derivatives.
2. **Semi-differential invariants**, cf. [GM1]. If distinguished points can be found it is possible to reduce the orders of the derivatives needed.
3. **Distinguished points**. If many distinguished points can be found it is possible to use these directly to construct invariants without calculating derivatives.
4. **Projectively invariant fitting of algebraic curves**, cf. [Car1]. A method has been developed to find projective invariants by fitting ellipses to planar curves. These can be used to construct invariants directly, or in the transformation into a canonical reference frame. This construction can also be used to find distinguished points.
5. **Canonical frame**, cf. [RZ1]. Four distinguished points at a concavity can be used to transform a concavity into a canonical reference frame. Any feature in this reference frame are invariant and can be used to recognize the curve.

5.2 Spherical normalization of point configurations

Let Ω be the set of finite point configurations on the sphere S^2 , that could be the projection of points on a plane, i.e. there is a plane through the origin which divides the sphere into two halves such that all the points lie on one half. A generalization of the planar moments (23) can be defined on point configurations according to

$$\begin{aligned} pm_1(\omega) &\sim \sum_{x \in \omega} \frac{x}{\|x\|} dx \\ pm_2(\omega) &\sim \sum_{x \in \omega} \frac{xx^T}{\|x\|} dx \end{aligned} \tag{25}$$

Consider the group of linear transformations of the sphere, cf. Section 2.3

$$g(x) = gx / \|gx\|$$

Now consider the normalization scheme based on the following normal configuration

$$\Omega_P = \{\omega \mid pm_2(\omega) \sim I, pm_1(\omega) \sim (0, 0, 1)\}$$

It turns out that the normal representatives are unique up to rotation around $(0, 0, 1)$. Here it is crucial that S^2 is used. The first moment is well defined for point configurations that lie on one half of the sphere. This property is kept under the transformation group. It is not possible to define the first moment using P_R^2 . Notice that the normal reference frames do not depend on fiducial points or ordering of the points.

The same idea can be used in higher dimensions.

5.3 Projective normalization of regions

An example of a projective normalization scheme of compact regions is presented in this section. Let

$$\Omega = \{\omega \subset R^2 \mid \omega \text{ is compact with positive area}\}.$$

The results below generate corresponding results for the closed boundary curves of such regions. A typical example is an extracted concavity as in Figure 11.

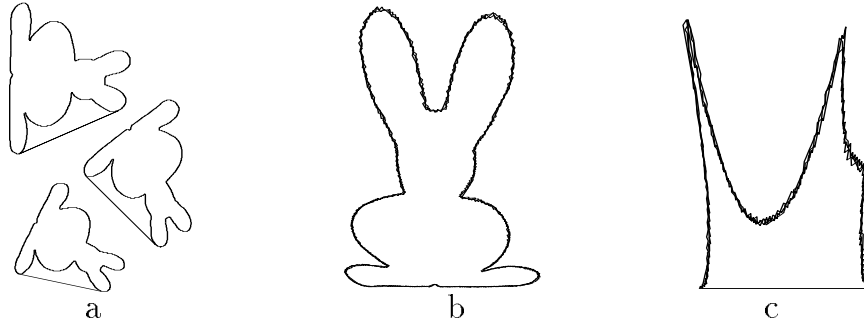


Figure 6: Three images of rabbits, cf. Figure 6a, are normalized into a reference frame in which their third moment tensor is zero and the second moment matrix is the identity, cf. Figure 6b. The same normalization is applied to the three regions enclosed by a bitangent and part of a contour, cf. Figure 6c.

Let the moments of a region ω be defined according to

$$\begin{aligned}
 m_0(\omega) &= \int_{x \in \omega} dx_1 dx_2 \\
 m_1(\omega)_i &= \int_{x \in \omega} x_i dx_1 dx_2 \\
 m_2(\omega)_{ij} &= \int_{x \in \omega} x_i x_j dx_1 dx_2 \\
 m_3(\omega)_{ijk} &= \int_{x \in \omega} x_i x_j x_k dx_1 dx_2
 \end{aligned}$$

Consider the group of planar projective transformations G acting on Ω . In the sequel only physically realisable transformations will be considered, i.e. those which do not send any of the points in ω to infinity. Now base a normalization scheme on the following ‘normal’ reference frames

$$\Omega_P = \{\omega \mid m_0(\omega) = 1, m_2(\omega) = aI, m_3(\omega) = 0, a \in R\} \quad (26)$$

This gives a normalization scheme with several representatives from each equivalence class. This might seem like a drawback, since having a unique representative from each equivalence class is conceptually simpler. The number of solutions may vary. It is shown in Section 5.5 that no matter what method is used in normalizing a curve with respect to projective transformations in a continuous way, there are curves with more than one normal reference frame.

One way of locking the rotation is to demand that the maximum distance of a point in ω to the origin occurs at the x_1 -axis. This method can be used also with convex curves.

The normalization scheme has been implemented and an experimental session will be presented. In this experiment, gray-scale images of roughly planar objects are taken with a digital camera. Polygon approximations of contours in the image are obtained using a Canny-Deriche edge detector. These curves are then normalized according to the proposed method, see Figure 6. Notice the good performance in Figures 6b and 6c. The three normalized curves lie practically on top of each other in spite of possible nonlinearities in the camera, errors in segmentation and in edge detection. The rabbit covered roughly 150×200 pixels in the image.

The normalization scheme is quite general and it is possible to normalize convex curves as well. This is illustrated in Figure 7. Sixteen images are taken of a convex shape. These are transformed into the unique normal affine reference frame, given by (24) and into one normal projective reference frame, given by (26). The affine approximation holds reasonably well for

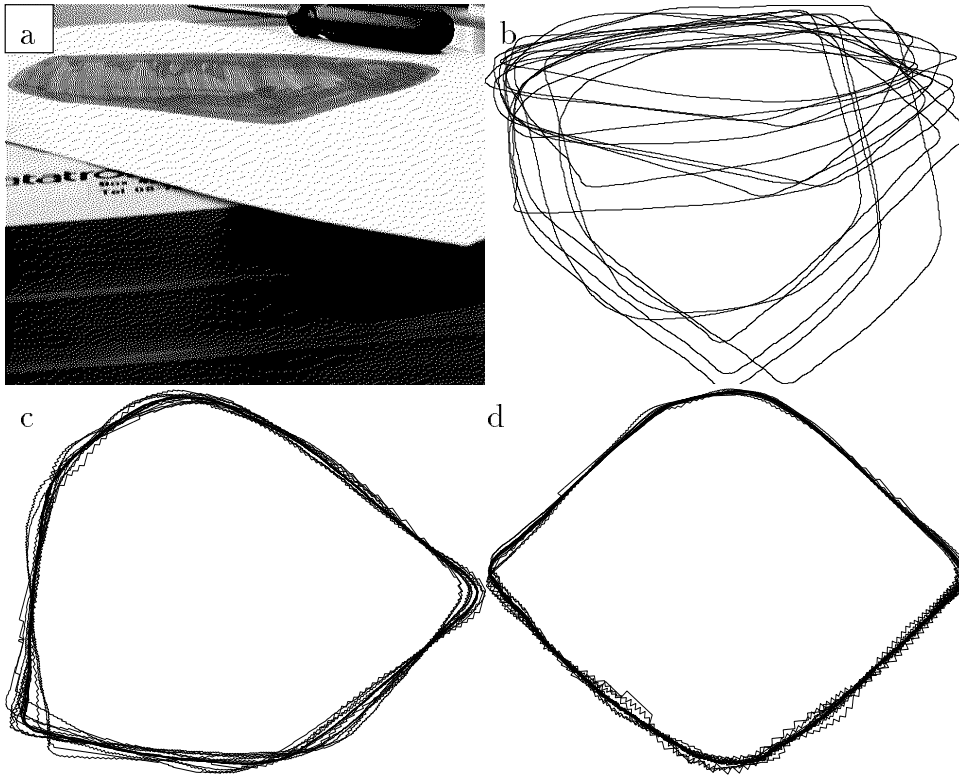


Figure 7: Figure 7a shows one of 16 images taken of a planar convex curve. All 16 extracted edges are shown in Figure 7b. In figure 7c the edges have been transformed into their unique affine normal reference frame, and in Figure 7d one of several possible projective normal reference frames have been used.

several of these images, but for the more extreme views, like Figure 7a, projective normalization is more adequate. Projective invariant R^2 -features are shown in Figure 8. These are extracted, according to Section 3.5, from one of several possible normal projective reference frames.

5.4 Inherent difficulties in maximizing compactness

It has been, proposed, cf. [BS1], to use maximal compactness over the physically realisable projective transformations to select the normal reference frames. One problem with this approach is that the optimization is done over a non-compact parameter space. It is therefore possible that the infimum is not attained at any point. In this section it is shown that this is indeed the case if the convex hull has a part that is smooth and curved. Since most curves in computer vision can be approximated by such curves this is a serious problem.

Let \mathcal{C} be the class of all closed curves, with finite arc-length, such that the boundary of the convex hull has at least one smooth, curved part. For such curves it is possible to calculate arc-length l and the area A enclosed by the curve. It is a well known fact in the calculus of variations that for curves in \mathcal{C} , then $l(C)^2/A(C) \geq 4\pi$ with equality if and only if C is a circle. For a specific curve $C \in \mathcal{C}$ let P_C be the set of *physically realisable projective transformations*, cf. Section 2.3. They do not send any of the points of C to infinity. The following **modified**

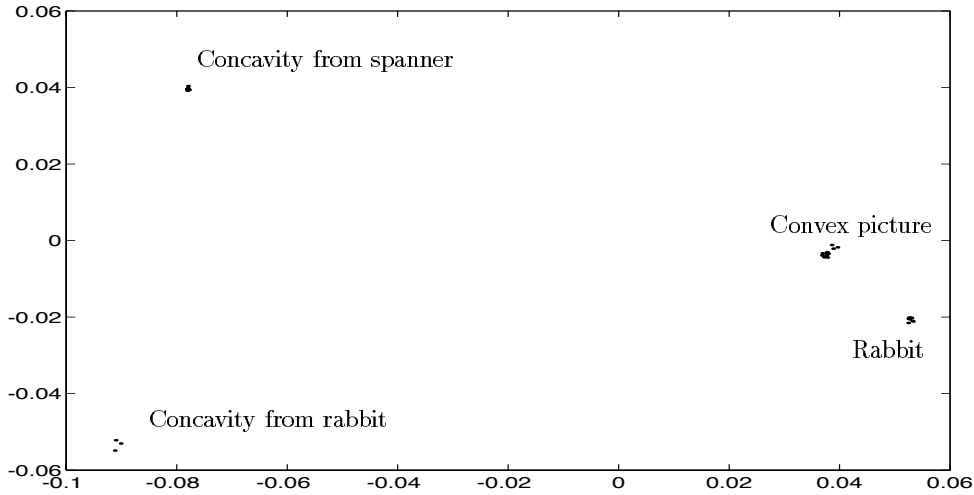


Figure 8: Invariant features extracted after projective normalization. Sixteen images of the convex shape in Figure 7, seven images of the whole rabbit in Figure 6, three images of one of the concavities from the rabbit in Figure 6 and nine images of one concavity from a spanner. Notice the high discriminatory power.

Hausdorff metric on the class of all curves with finite arc-length will also be needed.

$$\tilde{d}(\omega_1, \omega_2) = \max_{z_1 \in \omega_1} \min_{z_2 \in \omega_2} \|z_1 - z_2\| + \max_{z_2 \in \omega_2} \min_{z_1 \in \omega_1} \|z_1 - z_2\| + |l(\omega_1) - l(\omega_2)|, \quad (27)$$

where $\|z\|$ is the Euclidean norm.

This metric is a modification of the Hausdorff metric, (16), the modification being that also the arc-lengths should be compared. Being close to a given curve in this metric is a strong restriction. Two curves that are close really look alike, i.e. they would produce almost the same image on a CCD screen or on the retina of an eye. All points on one curve are close to some point of the other and their arc-lengths are almost equal. Theorems 3 and 5 are true using the ordinary Hausdorff metric. The modified metric is needed for Corollary 4.

Theorem 3 *Let Γ_0 be the unit circle and let \mathcal{C} and P_C be as above. Then*

$$C \in \mathcal{C} \Rightarrow \inf_{p \in P_C} \tilde{d}(p(C), \Gamma_0) = 0$$

One interpretation of this theorem is that comparing the images of C from some sequence of viewpoints, these images look more and more like a circle. The projective transformations involved when approaching the limit are, however, quite extreme. They will be given by

$$p_n((x, y)) = \left(\frac{2nx}{(n^2 - 1)y + 2}, \frac{2n^2y}{(n^2 - 1)y + 2} \right).$$

We will also use the ellipses

$$C_\epsilon = \{((1 + \epsilon) \cos t, \sin t + 1) | t \in \mathcal{R}\}, \quad \epsilon > -1$$

with center at the point $(0, 1)$, axes of length 1 in the x -direction and axis of length $1 + \epsilon$ in the y -direction. In particular, C_0 is the circle $x^2 + (y - 1)^2 = 1$. These ellipses intersect twice at $(0, 0)$ and twice at $(0, 2)$.

One can easily verify, e.g. using homogeneous coordinates, that the family p_n has the following properties

- $p_1 = I, \quad p_a \circ p_b = p_{ab}$
- $p_n(C_\epsilon) = C_\epsilon, \quad p_n((0, 0)) = (0, 0), \quad p_n((0, 2)) = (0, 2)$

By the latter item, the transformations p_n reparametrise the ellipses C_ϵ . It can be shown that if $n > 1$ a neighbourhood of $(0, 0)$ expands and a neighbourhood of a focus $(0, 2)$ contracts. The properties above indicate that $p_n^{-1} = p_{1/n}$. Another important fact that will be used in the proof of Theorem 5 is that

$$p_n^{-1} = q \circ p_n \circ q \quad (28)$$

where q is a change of coordinates $q((x, y)) = (-x, 2 - y)$, in which the two points $(0, 0)$ and $(0, 2)$ change places. The following lemma quantifies the contractive properties, and will be needed for the proofs of Theorems 3 and 5.

Lemma 4 *For every compact region D in the upper half plane there exists a constant K , so that for every smooth curve $C \subset D$ and point $(x, y) \in D$ holds*

$$l(p_n(C)) < \frac{K}{n} l(C) \quad (29)$$

$$\|p_n((x, y)) - (0, 2)\| < \frac{K}{n} \|(x, y) - (0, 2)\|, \quad (x, y) \in C \quad (30)$$

Proof: Let

$$(u, v) = p_n(x, y) = \left(\frac{2nx}{(n^2 - 1)y + 2}, \frac{2n^2y}{(n^2 - 1)y + 2} \right).$$

For any parametrisation $x = x(t), y = y(t)$ holds

$$u' = \frac{2n}{(n^2 - 1)y + 2} x' - \frac{2nx}{((n^2 - 1)y + 2)^2} (n^2 - 1) y'$$

and

$$v' = \frac{2n^2}{(n^2 - 1)y + 2} y' - \frac{2n^2y}{((n^2 - 1)y + 2)^2} (n^2 - 1) y' = \frac{4n^2}{((n^2 - 1)y + 2)^2} y'.$$

Since both x and y are bounded and $y > y_{\max} > 0$, holds

$$|u'| < \frac{K_1}{n} |(x', y')| \quad \text{and} \quad |v'| < \frac{K_2}{n} |(x', y')|,$$

which implies that

$$|(u', v')| < \frac{K}{n} |(x', y')|$$

for some constants K_1, K_2 and K , independent of parametrisation, only depending on D . Both (29) and (30) follow from the above. ■

Proof: of Theorem 3.

Let $C \in \mathcal{C}$, and let a be a point in the interior of a smooth curved part of the boundary of the convex hull of C . Choose a coordinate system with origin at a with x -axis along the tangent

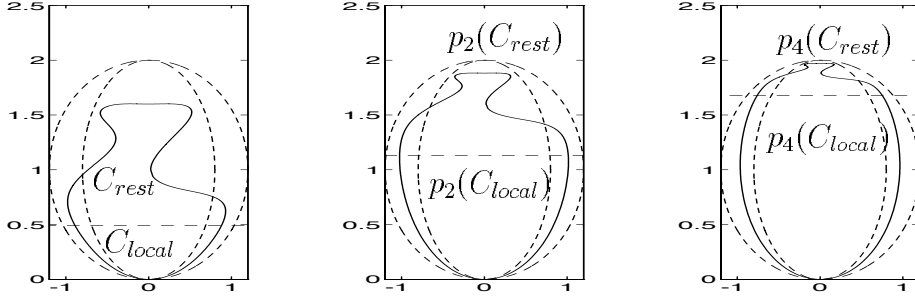


Figure 9: The curve is split into two parts. A local part C_{local} under the line is between the two ellipses C_ϵ and $C_{-\epsilon}$. The rest C_{rest} tend to the point $(0, 2)$ as n increases.

and so that the curvature at a is one. It will be shown that as $n \rightarrow \infty$, using the metric d , $p_n(C)$ will tend to the translated unit circle C_0 , from which the theorem follows. The image of the local part of the curve around a will form the main part of the boundary of the circle and the remaining part will be mapped into a neighbourhood of $(0, 2)$.

Take $\epsilon > 0$, and let C_{local} be the connected part of C in a neighbourhood of $(0, 0)$, that lies between the ellipses C_ϵ and $C_{-\epsilon}$. See Figure 9. The ellipses are invariant under p_n for all n , so

$$1 - \epsilon < |(u, v) - (0, 1)| < 1 + \epsilon, \quad \forall (u, v) \in p_n(C_{local}), \quad \forall n.$$

The rest of the curve $C_{rest} = C \setminus C_{local}$ is compact and belongs to the upper half plane. According to Lemma 4, all points (u, v) of $p_n(C_{rest})$ tend uniformly to $(0, 2)$, as $n \rightarrow \infty$. This means that for each $\epsilon > 0$ we can choose n so that all points of $p_n(C)$ lie within the distance ϵ from C_0 . See Figure 9. Hence

$$\lim_{n \rightarrow \infty} \left(\max_{z_1 \in p_n(C)} \min_{z_2 \in C_0} \|z_1 - z_2\| + \max_{z_1 \in C_0} \min_{z_2 \in p_n(C)} \|z_1 - z_2\| \right) = 0$$

By this, one has control of the first two terms in the definition of \tilde{d} . A consequence, that will be used below is that $\lim_{n \rightarrow \infty} A(p_n(C)) = \pi$.

It remains to consider the third term in \tilde{d} . The curve C is smooth around $(0, 0)$, so it is possible to choose C_{local} so small that it, together with the lines L_1 and L_2 from the endpoints of C_{local} to $(0, 2)$, forms the boundary of a convex region R . See Figure 10. For all n the transformed region $p_n(R)$ is convex and is contained in the ellipse C_ϵ . Since the shortest path circumventing a bounded region is the boundary of its convex hull, and since $p_n(C_{local})$ is part of the boundary of the convex region $p_n(R)$, we can deduce that $l(p_n(C_{local})) < l(p(C_\epsilon))$ for all n . By comparison with a circle of radius $1 + \epsilon$ we get $l(p_n(C_\epsilon)) < 2\pi(1 + \epsilon)$. Again, using the compactness of C_{rest} and Lemma 4, holds for each $\epsilon > 0$

$$\lim_{n \rightarrow \infty} l(p_n(C)) = \lim_{n \rightarrow \infty} l(p_n(C_{local})) + \lim_{n \rightarrow \infty} l(p_n(C_{rest})) \leq 2\pi(1 + \epsilon) + 0.$$

Hence

$$\lim_{n \rightarrow \infty} l(p_n(C)) \leq 2\pi.$$

On the other hand, since $l(p_n(C))^2/A(p_n(C)) \geq 4\pi$, it follows that

$$\lim_{n \rightarrow \infty} l(p_n(C)) \geq 2\pi.$$

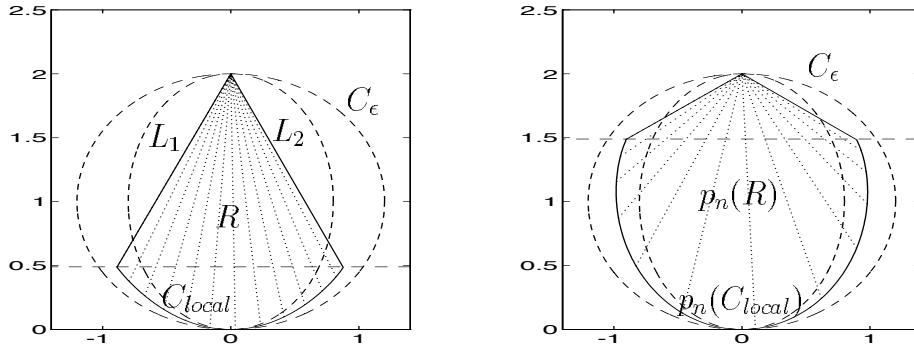


Figure 10: The local part C_{local} together with two line segments form the boundary of a convex region R . The ellipse C_ϵ circumvents the convex region $p_n(R)$ for all n .

This proves the theorem. ■

An immediate corollary is

Corollary 4

$$C \in \mathcal{C} \Rightarrow \inf_{p \in P_C} \frac{l(p(C))^2}{A(p(C))} = 4\pi$$

Here we need the modified metric (27) to control the arc-length of the curve. The projective transformations involved when approaching the limit are, however, quite extreme.

Since most curves can be approximated by smooth curves, Theorem 3 has some serious consequences. If a curve is modelled by a cubic spline the infimum is not attained. If a curve is modelled as a piecewise linear curve then the minimum is attained but the normalized reference frame will depend crucially on how the boundary is approximated. This is illustrated in Figure 11. In this figure the same shape is approximated with polygons to different accuracy. The reference frame in which they have maximal compactness is indeed different.

5.5 On the limitations on projective normalization schemes

In the proof of Theorem 3 one notices that the main part of the curve is squeezed into a neighbourhood of a point. The curve $p_n(C)$ looks like a circle but it has a small ripple that corresponds to the main part of the curve C . It turns out that if we slightly perturb the curve $p_n(C)$ outside this ripple and then do the inverse projective transformation, the new curve is almost identical to the original one. A consequence is the following somewhat surprising theorem.

Theorem 5 *Given $\Gamma_1, \dots, \Gamma_m$, closed continuous curves with finite arc-length. To every $\epsilon > 0$, there exist a curve C and projective transformations q_1, \dots, q_m so that*

$$d(q_i(C), \Gamma_i) < \epsilon, \quad i = 1, \dots, m$$

Note that the curves Γ_i do not have to be smooth.

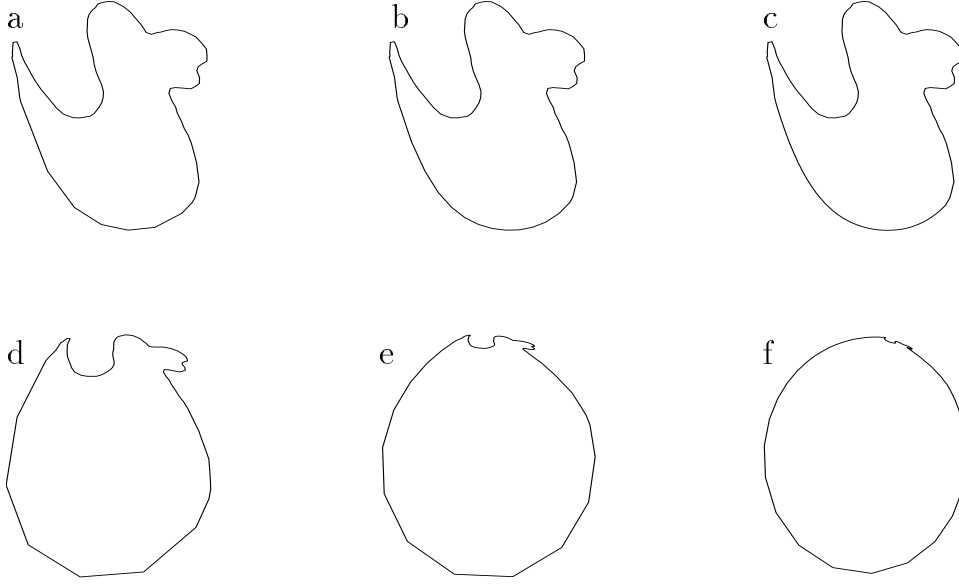


Figure 11: Three different approximations of a planar curve (a, b and c) and their normal projective reference frames with respect to maximal compactness (d, e and f). Notice how much the normal reference frame depends on the representation of the boundary.

Proof: Since there is a smooth curve arbitrarily close to every curve fulfilling the assumptions above, it is no restriction to assume that the curves $\Gamma_1, \dots, \Gamma_m$ are smooth and therefore in \mathcal{C} . Let C_0 be the unit circle, $x^2 + (y - 1)^2 = 1$. To each Γ_j associate a point P_j on C_0 , a compact region D_j and a sector S_j , according to Figure 12, where the sectors S_j are supposed to be pairwise disjoint.

By the construction of the proof of Theorem 3 it is possible to find projective transformations π_j such that $\pi_j(\Gamma_j)$ is close to C_0 and

$$\begin{aligned} \pi_j(\Gamma_{j,local}) &\subset D_j \\ \pi_j(\Gamma_{j,rest}) &\subset (\cap_{i \neq j} D_i) \cap S_j. \end{aligned}$$

By Lemma 4, for each j , π_j can be chosen so that the inverse projection $q_j = \pi_j^{-1}$ shrinks all curves in D_j of arc-length less than, say $3\pi m$ into a curve with arc-length less than $\epsilon/2$, and whose points are within $\epsilon/4$ of a point of Γ_j .

Let C be constructed by gluing the patches $\pi_j(\Gamma_{j,rest})$, the line segments obtained by radially connecting the endpoints of $\pi_j(\Gamma_{j,rest})$ with C_0 , for all j , and the intermediate arcs of C_0 . Both $C \setminus \pi_j(\Gamma_{j,rest})$ and $\pi_j(\Gamma_{j,local})$ are in D_j and their arc-length is for small ϵ certainly less than $3\pi m$ so,

$$d(q_j(C \setminus \pi_j(\Gamma_{j,rest}), \Gamma_{j,local})) < \epsilon.$$

The remaining part of C is $\pi_j(\Gamma_{j,rest})$, and is thus mapped identically into $\Gamma_{j,rest}$ by q_j . Hence $d(q_j(C), \Gamma_j) < \epsilon$. ■

The theorem is illustrated by Figure 13. The curve C was constructed in MATLAB as in the proof. Notice that in Figure 13 the lower four curves are projectively equivalent. The only errors are in the plotting, printing, copying and viewing of the curves. The theorem is in itself

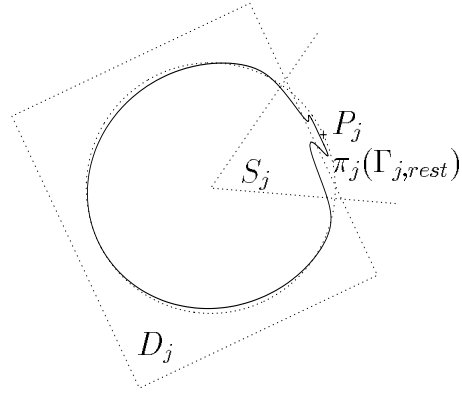


Figure 12: After applying a projective transformation π_j to curve Γ_j , the part in sector S_j will be glued to other curve fragments to form a curve that looks like all the curves $\Gamma_1, \dots, \Gamma_m$.

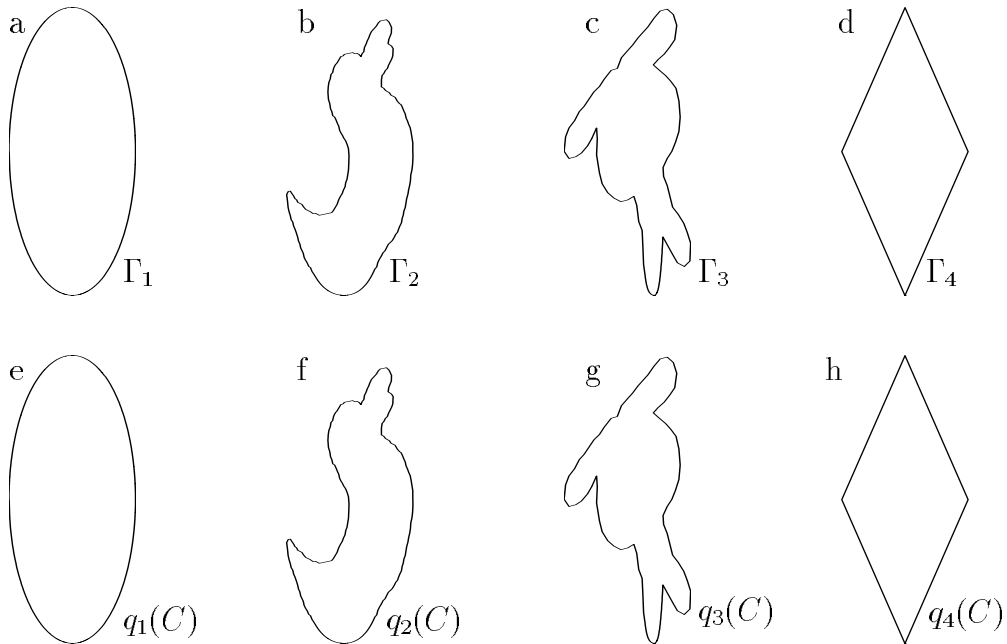


Figure 13: 13a-d: Four planar curves with different projective shapes. 13e-h: Four planar curves that are projectively equivalent, i.e. have the same shape. Errors in the plotting device, printing, copying and viewing conditions make them look pairwise equal (a and e, b and f, c and g, d and h).

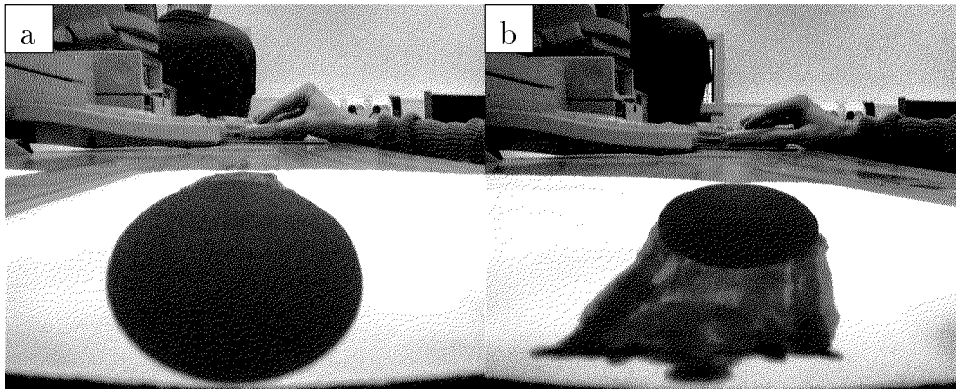


Figure 14: It is difficult to talk about scale when dealing with projective equivalence of planar curves. What seems like a small scale perturbation from one viewpoint might look like large scale shape from another.

somewhat surprising and unintuitive at first, but it is a simple trick of hiding a shape along the convex hull of another shape. The reason it works is the use of extreme, but non-singular, projective transformations. The consequences are perhaps more important.

Corollary 6 *Let T be a projectively invariant mapping from the set of closed continuous curves with finite arc-length to a Hausdorff topological feature space, e.g. the real line, then T maps all curves at which it is continuous onto the same value.*

Proof: Assume to the contrary that $r_1 = T(\Gamma_1) \neq r_2 = T(\Gamma_2)$. Since the feature space is Hausdorff it is possible to find disjoint open sets $O_1 \ni r_1$ and $O_2 \ni r_2$. According to Theorem 5 the inverse images $T^{-1}(O_1)$ and $T^{-1}(O_2)$, which are open sets around Γ_1 and Γ_2 , contain a projectively equivalent pair of curves, contradicting the assumption. ■

The problems that cause these kind of anomalies can be experienced in almost normal viewing conditions. It is the simple fact that Euclidean scale is changed under projective transformations so that what seems like a small scale perturbation from one viewpoint is large scale shape from another. This is illustrated in Figure 14.

Suppose that we have a continuous projective normalization scheme that gives a unique representative from each equivalence class. It would then be possible to construct a continuous and non-constant projective invariant mapping from the set of curves with the Hausdorff metric to the real line. This is impossible according to Corollary 6. The conclusion is that *projective normalization schemes on the set of planar curves cannot both be continuous and give a unique representative from each equivalence class.* Either continuity or uniqueness has to be sacrificed.

6 Conclusions

Recognition and pose determination of planar features in a single gray-scale image have been discussed in this paper. It has been shown how geometrical configurations, like points, curve segments and regions can be found in an image using edge detection and segmentation. These segments are relatively robust to variations in lighting and occlusion.

These configurations are then recognized using affine and projective invariant features. Normalization is used as a conceptually simple method to construct such invariants. The normalization schemes should be continuous, and preferably give a unique representative from each equivalence class. They can be used to induce a metric on the equivalence classes of curves. They can also be used to construct invariant indices. The resulting indices have good discriminatory properties and are robust to image distortions and digitization errors.

In the affine case two methods have been investigated. The first uses maximal compactness of a curve segment or equivalently weak isotropy to define a normal reference frame which is unique up to similarity transformations. The scheme is relatively robust to perturbations of the curve. However, it is not continuous with respect to the Hausdorff metric. Another method uses moments of the region enclosed by a curve to define a normal reference frame which is unique up to rotations and continuous with respect to the Hausdorff metric. In both of these methods the canonical reference frame does not depend on how the points are ordered. Fiducial points are not needed. It is easy to generalize the normalization scheme based on moments to general configurations of points, curves and regions. It can also be extended to higher dimensions.

It is more difficult to find invariants under projective transformations. Moments can be used to define a unique normal reference frame for point configurations. A normalization scheme for regions has been presented. It is also based on moments. The normal reference frame is generally not unique, but individual reference frames are in many cases continuous with respect to the Hausdorff metric. Normal reference frames exist for many types of regions, but there are curves which are inherently difficult to normalize. When investigating the limitations of projective normalization schemes of planar closed curves, it was found that maximal compactness is not suitable. Such schemes are not even well-defined for smooth curves. Finally it has been shown that it is impossible to achieve both uniqueness and continuity for projective normalization of general non-algebraic curves.

The work has focused on simple algebraic and topological properties and can be extended in several directions. First of all there are several questions of how the invariants should be used in a recognition system. There seems to be curves that are generically difficult to normalize with respect to projective transformations. These critical sets should be investigated as well as the continuity of the proposed invariants. It would also be interesting to incorporate probabilistic models for image distortions. This could give valuable insight into the effectiveness and optimality of the normalization schemes and their use in recognition.

Acknowledgements

I would like to thank my supervisor Gunnar Sparr for inspiration and guidance. I would also like to thank my fellow students Anders Heyden and Carl-Gustav Werner for their help.

References

- [AL1] Alvarez, L., Guichard, F., Lions, P.-L., Morel, J.-M.: Axioms and Fundamental Equations of Image Processing. Technical Report 9231, CEREMADE (1992).
- [AL1] Alvarez, L., Guichard, F., Lions, P.-L., Morel, J.-M.: Axiomes et équations fondamentales du traitement d'images. (Analyses multiéchelle et E.D.P.). Comptes Rendus, Acad. Sci. Paris, t. 315, Série I (1992), 135-138.
- [Ar1] Arbter, K.: Affineinvariante Fourierdeskriptoren ebener Kurven. Ph.D diss. Hamburg, Germany (1990).

- [ASBH1] Arbter, K., Snyder, W. E., Burkhardt, H., Hirzinger, G.: Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects. *IEEE Trans. PAMI* Vol. 12, 7 (1990).
- [BM1] Blake, A., Marinos, C.: Shape from Texture: Estimation, Isotropy and Moments. *Artificial Intelligence* 45 (1990), 332-380.
- [BS1] Blake, A., Sinclair, D.: On the projective normalisation of planar shape. Technical Report OUEL Oxford, U.K. (1992).
- [BY1] Brady, M., Yuille, A.: An Extremum Principle for Shape from Contour. *IEEE Trans. PAMI* Vol. 6, 3 (1984), 288-301.
- [BW1] Burns J. B., Weiss R. S., Riseman E. M.: The Non-existence of General-case View-Invariants. *Geometrical Invariance in Computer Vision*, Mundy, J. L. and Zisserman, A. editors, MIT Press (1992).
- [Can1] Canny J. F.: A Computational Approach to Edge Detection. *IEEE Trans. PAMI*, Vol. 8. 6 (1986), 679-698.
- [Car1] Carlsson S.: Projectively Invariant Decomposition and Recognition of Planar Shapes, Proc. ICCV4, Berlin, Germany (1993), 471-475.
- [Cox1] Coxeter, H. S. M.: *The Real Projective Plane*. Springer-Verlag, New York, NY, USA (1993).
- [De1] Deriche R.: Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision* 1 (1987), 167-187.
- [DH1] Duda, R. O. and Hart, P. E.: *Pattern Classification and Scene Analysis*. Wiley-Interscience (1973).
- [Fr1] Fransson, P.: Using Wavelets to Detect Corner Points in Images. Masters Thesis, Dept. of Mathematics, Lund Institute of Technology, Lund, Sweden (1992).
- [GM1] Van Gool, L., Moons, T., Pauwels, E. and Oosterlinck, A.: Semi-differential Invariants. in Mundy, J. L., and Zisserman A. (editors): *Geometric invariance in Computer Vision*. MIT Press, Cambridge Ma, USA (1990) 157-192.
- [GW1] Gonzalez, R. C., Wintz, P.: *Digital Image Processing*. Addison-Wesley, Reading, Mass (1987).
- [Gå1] Gårding, J.: Shape from Surface Markings. Ph. D. thesis, Dept. of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden (1991).
- [Gr1] Gros, P., and Quan L.: Projective Invariants for Vision. Technical Report RT 90 IMAG - 15 LIFIA, LIFIA-IRIMAG, Grenoble, France (1992).
- [HW1] Hubel, D. H., Wiesel, T.N.: *Eye, brain and vision*. Freeman, New York (1990).
- [HM1] Hildreth, E. Marr, D.: Theory of Edge Detection. *Proceedings of Royal Society of London*, 207 (1980), 187-217.
- [Ho1] Horn B. K. P.: *Robot Vision*. MIT Press, Cambridge, Mass, USA (1986).

- [LS1] Lamdan, Y., Schwartz, J. T., and Wolfson, H. J.: Affine Invariant Model-based Object Recognition. *IEEE Journal of Robotics and Automation* **6** (1990), 578-589.
- [Ma1] Marr D.: *Vision*. W. H. Freeman New York, USA (1982).
- [My1] Maybank, S. J.: Probabilistic analysis of the application of the cross ratio to model based vision. Technical Report, GEC, U. K. (1993).
- [My2] Maybank, S. J.: Classification Based on the Cross Ratio. Proc. Second ARPA/NSF-ESPRIT Workshop on Invariance, Ponta Delgada, Azores, Portugal (1993), 113-132.
- [MZ1] Mundy, J. L., and Zisserman A. (editors): *Geometric invariance in Computer Vision*. MIT Press, Cambridge Ma, USA (1990).
- [NS1] Nielsen, L., Sparr, G., Projective Area-invariants as an extension of the Cross-Ratio. *CVGIP:Image Understanding*, Vol. 54., **1**, July (1991), 145-159.
- [PM1] Perona, P., and Malik, J., Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. PAMI*, Vol. 12, (1990), 629-639.
- [RZ1] Rothwell, C. A., Zisserman, A., Forsyth, D. A. and Mundy J. L.: Canonical Frames for Planar Object Recognition. Proc. ECCV2 Genova Italy (1992), 757-772.
- [Ro1] Rothwell, C. A.: Hierarchical Object Description Using Invariants. Proc. Second ARPA/NSF-ESPRIT Workshop on Invariance, Ponta Delgada, Azores (1993), 287-302.
- [ST1] Sapiro, G., and Tannenbaum A., On invariant curve evolution and image analysis. *to appear in Indiana University Journal of Mathematics* (1993).
- [ST2] Sapiro, G., and Tannenbaum A., On affine invariant scale-space. *to appear in Journal of Functional analysis* (1993).
- [Sch1] Schur, I., *Vorlesungen über Invariantentheorie*. Springer-Verlag, Berlin (1968).
- [WB1] Wang, H., Brady, M.: Corner Detection and its Application in 3D reconstruction from binocular images. Proc. 1991 Stockholm Workshop on Computational Vision, Ed. Jan-Olof Eklundh, CVAP 102, KTH, Stockholm, Sweden (1991).
- [We1] Weiss, I.: Noise-resistant Invariants of Curves. in Mundy, J. L., and Zisserman A. (editors): *Geometric invariance in Computer Vision*. MIT Press, Cambridge Ma, USA (1990) 135-156.
- [Wit1] Witkin, A. P.: Recovering Surface Shape and Orientation from Texture. *J. of Artificial Intelligence* **17** (1981), 17-45.
- [Wil1] Wilczynski, E. J.: *Projective Differential Geometry of Curves and Ruled Surfaces*. B. G. Teubners Lehrbücher, Leipzig, (1906).
- [ÅM1] Åström, K., and Morin, L.: Random Cross Ratios. Technical Report RT 88 IMAG - 14 LIFIA, LIFIA-IRIMAG, Grenoble, France (1992).
- [Ås1] Åström, K.: Where am I and what am I seeing? Algorithms for a laser guided vehicle. Masters Thesis, Dept. of Mathematics, Lund Institute of Technology, Lund, Sweden (1991).

- [Ås2] Åström, K.: A Correspondence Problem in Laser-Guided Navigation. Proc. Swedish Society for Automated Image Analysis, Uppsala, Sweden (1992), 141-144.
- [Ås4] Åström, K.: Automatic Mapmaking. Proc. First IFAC International Conference on Intelligent Autonomous Vehicles, Southampton, UK (1993).
- [Ås5] Åström, K.: Affine Invariants of Planar Sets. Proc. SCIA8, Tromsø, Norway (1993), 769-776.
- [Ås6] Åström, K.: Fundamental Difficulties with Projective Normalization of Planar Curves. Proc. Second ARPA/NSF-ESPRIT Workshop on Invariance, Ponta Delgada, Azores, Portugal (1993), 377-389.