Nikos Paragios, Yunmei Chen and Olivier Faugeras
Editors

Handbook of Mathematical Models in Computer Vision

Springer

HANDBOOK OF MATHEMATICAL MODELS IN COMPUTER VISION

Edited by

Nikos Paragios
Ecole Nationale des Ponts et Chaussees

Yunmei Chen
University of Florida

Olivier Faugeras
INRIA

Springer

Chapter11

Invariant Processing and Occlusion Resistant Recognition of Planar Shapes

A. Bruckstein

Abstract

This short paper surveys methods for planar shape smoothing and processing and planar shape recognition invariant under viewing distortions and even partial occlusions. It is argued that all the results available in the literature on these problems implicitly follow from successfully addressing two basic problems: invariant location of points with respect to a given shape (a given set of points in the plane) and invariant displacement of points with regard to the given shape.

11.1 Introduction

Vision is an extremely complex process aimed at extracting useful information from images: recognizing three-dimensional shapes from their two-dimensional projections, evaluating distances and depths and spatial relationships between objects are tantamount to what we commonly mean by seeing. In spite of some irresponsible promises, made by computer scientists in the early 60's, that within a decade computers will be able "to see", we are not even close today to having machines that can recognize objects in images the way even the youngest of children are capable to do. As a scientific and technological challenge, the process of vision has taught us a lesson in modesty: we are indeed quite limited in what we can accomplish in this domain, even if we call to arms deep mathematical results and deploy amazingly fast and powerful electronic computing devices. In order to address some practical technological image analysis questions and in order to appreciate the complexity of the issues involved in "seeing" it helps to consider simplified vision problems such as "character recognition" and other "model-based planar shape recognition" problems and see how far our theories (i.e. our "brain-power") and experiments (or our "number-crunching power") can take us toward working systems that accomplish useful image analysis tasks. As a result of such scientific and commercial efforts we do have a few vision systems that work and there is a vast literature in the "hot" field of computer vision dealing with representation, approximation, completion, enhancement, smoothing, exaggeration, characterization and recognition of planar shapes. This paper surveys methods for planar shape recognition and processing (smoothing, enhancement, exaggeration etc.) invariant under distortions that occur when looking at planar shapes from various points of view. These distortions are the Euclidean, Similarity, Affine and Projective maps of the plane to itself and model the possible viewing projections of the plane where a shape is assumed to reside into the image plane of a pinhole camera, that captures the shape from arbitrary locations. A further problem one must often deal with when looking at shapes is occlusion. If several planar shapes are superimposed in the plane or are floating in 3D-space they can and will (fully, or partially) occlude each other. Under full occlusion there is of course no hope for recognition, but how about partial occlusion? Can we recognize a planar shape from a partial glimpse of its contour? Is there enough information in a portion of the projection of a planar shape to enable its recognition? We shall here address such questions too. The main goal of this paper will be to point out that all methods proposed to address the above mentioned topics implicitly require the solution of two fundamental problems: distortion-invariant location of points with respect to given planar shape (which for our purposes can be a planar region with curved or polygonal boundaries or in fact an arbitrary set of points) and invariant displacement, motion or relocation of points with respect to the given shape.

11.2 Invariant Point Locations and Displacements

A planar shape S, for our purpose, will be a set of points in R^2 points that most often specify a connected a planar region with a boundary that is either smooth or polygonal. The viewing distortions are classes of transformations $V_\phi : R^2 \longrightarrow R^2$ parameterized by a set of values ϕ, and, while the class of transformations is assumed to be known to us, the exact values of the parameters are not. The classes of transformations considered are continuous groups of transformations modeling various imaging modalities, the important examples being:

- The Euclidean motions (parameterized by a rotation angle θ and a two-dimensional translation vector (t_x, t_y), i.e. ϕ has 3 parameters).

$$\mathrm{V}_\phi^E : \ (x, y) \rightarrow (x, y) \begin{bmatrix} \cos\theta & sin\theta \\ -sin\theta & \cos\theta \end{bmatrix} + (t_x, t_y)$$

- Similarity transformations (Euclidean motions complemented by uniform scaling transformations, i.e. $|\phi| = 4$ parameters).

$$\mathrm{V}_{\phi}^{S}: \ (x,y) \rightarrow (x,y) \begin{bmatrix} \cos\theta & sin\theta \\ -sin\theta & \cos\theta \end{bmatrix} \alpha + (t_x, t_y)$$

- Equi-Affine and Affine Mappings (parameterized by 2×2 matrix - 4 parameters - or 3, if the matrix has determinant 1 - and a translation vector, i.e. $|\phi| = 6$ or 5 parameters).

$$\mathrm{V}_{\phi}^{A}: \ (x,y) \rightarrow (x,y) \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + (t_x, t_y)$$

- Projective Transformations (modeling the perspective projection with $|\phi| =$ 8 parameters).

$$\mathrm{V}_{\phi}^{P}: \ (x,y) \rightarrow \frac{1}{a_{31}x + a_{32}y + 1}(x,y,1) \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

Given a planar shape $S \subset R^2$ and a class of viewing distortions $V_\phi : R^2 \longrightarrow R^2$ we consider the following problem:

> Two observers A and B look at $S_A = V_{\phi_A}(S)$ and at $S_B = V_{\phi_B}(S)$ respectively without knowing ϕ_A and ϕ_B. In other words A and B look at S from different points of view and the details of their camera location orientation and settings are unknown to them (See Figure 11.1). Observer A chooses a point P_A in its image plane R^2, and wants to describe its location w.r.t. $V_{\phi_A}(S)$ to observer B, in order to enable him to locate the corresponding point $P_B = V_{\phi_B}(V_{\phi_A}^{-1}(P_A))$. A knows that B looks at $S_B = V_{\phi_B}(S) = V_{\phi_B}(V_{\phi_A}^{-1}(S_A))$, but this is all the information available to A and B. How should A describe the location of P_A w.r.t. S_A to B?
> Solving this problem raises the issue of characterizing a position (P_A) in the plane of S_A in a way that is invariant to the class of transformations V_Φ.

Let us consider a very simple example: take S be a set of indistinguishable points in the plane $\{P_1, P_2, \ldots, P_N\}$ and V_ϕ be the class of Euclidean motions. A new point $\tilde{P}$ should be described to observers of this point constellation, under arbitrary viewing distortions V_ϕ, i.e. observers of

$$V_\phi\{P_1, \ldots, P_N\} = \{V_\phi(P_1), V_\phi(P_2) \ldots V_\phi(P_N)\}$$

so that they will be able to locate $V_\phi(\tilde{P})$ in their respective "images". How should we do this? Well, we shall have to describe $\tilde{P}$'s location w.r.t. $\{P_1, P_2, \ldots, P_N\}$ in an Euclidean-invariant way. We know from elementary geometry that Euclidean motions preserve lengths and angles between line segments so there are several ways to provide invariant coordinates in the plane w.r.t. the shape S. The origin of an invariant coordinate system could be the Euclidean-invariant (in fact even

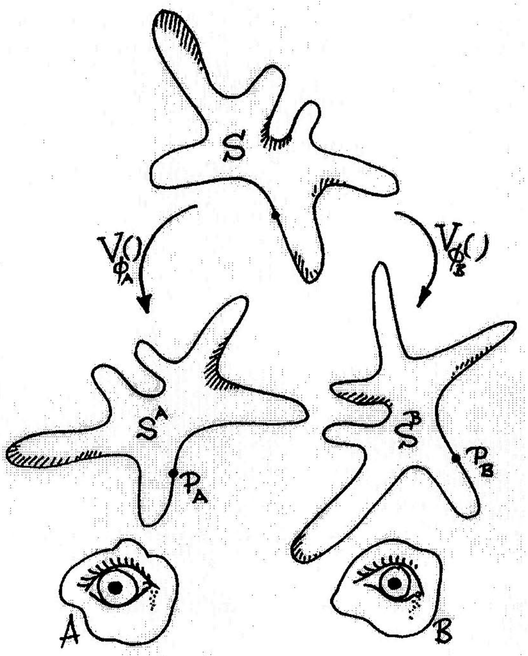

Figure 11.1.

Affine-invariant) centroid of the points S, i.e. $O_S = \left(\sum_{i=i}^{N} P_i\right)/N$. As one of axes (say the x-axis) of a "shape-adapted invariant" coordinate system, one may choose the longest or shortest (or closest in length to the "average" length) vector among $\{O\overline{P}_i\}$ for $i = 1, 2, \ldots, N$. This being settled, the y-axis can be defined as a 90°- rotation counter-clockwise and all one has to do is to specify $\tilde{P}$ in this adapted and Euclidean-invariant coordinate system with origin at O_s and orthogonal axes chosen as described above. Note that many other solutions are possible. We here assumed that the points of S are indistinguishable, otherwise the problem would be even simpler. Note also that ambiguous situations can and do arise. In case all the points of S form a regular N-gon, there are N equal length vectors $\{OP_i\}$ $i = 1, 2, \ldots, N$ and we can not specify uniquely an x-axis. But, a moment of thought will reveal that in this case the location of any point in the plane is inherently ambiguous up to rotations of $2\pi/N$.

Contemplating the above-presented simple example one realizes that solving the problem of invariant point location is heavily based on the invariants of the continuous group of transformations V_ϕ. The centroid of the point constellation (S), O_S, an invariant under V_ϕ, enabled the description of $\tilde{P}$ using a distance

$d(O_S, \tilde{P})$, the length of vector $O_S\tilde{P}$ (again a V_ϕ-invariant), up to a further parameter that locates $\tilde{P}$ on the circle centered at O_S with radius $d(O_SP)$, and then the "variability" or inherent "richness" of the geometry of S enabled the reduction of the remaining ambiguity.

Suppose next that we want not only to locate points in ways that are invariant under V_ϕ but we also want to perform invariant motions. This problem is already completely addressed in the above presented example, once an "S-shape-adapted" coordinate system became available. Any motion can be defined with respect to this coordinate system and hence invariantly reproduced by all viewers of S. In fact, when we establish an adapted frame of references we implicitly determine the transformation parameters, ϕ, and can effectively undo the action of V_ϕ.

To complicate the matters further consider the possibility that the shape S will be partially occluded in some of its views. Can we, in this case, establish the location of $\tilde{P}$ invariantly and perform some invariant motions as before? Clearly, in the example when S is a point constellation made of N indistinguishable points, if we assume that occlusion can remove arbitrarily some of the points, the situation may become rather hopeless. However, if the occlusion is restricted to wiping out only points covered by a disk of radius limited to some R_{max}, or alternatively, we can assume that we shall always see all the points within a certain radius around an (unknown) center point in the plane, the prospects of being able to solve the problem, at least in certain lucky instances, are much better. Indeed, returning to our simple example, assume that we have many indistinguishable landmark points (forming a "reference" shape S in the plane), and that a mobile robot navigates in the plane, and has a radius of sensing or visibility of R_{max}. At each location of the robot in the plane, $\tilde{P}$, it will see all points of S whose distance from $\tilde{P}$ is less than R_{max}, up to an arbitrary rotation. Hence, the question of being able to specify $\tilde{P}$ from this data becomes the problem of robotic self location w.r.t the landmarks. So given a reference map (showing the "landmark"points of S in some "absolute" coordinate system), we want the robot to be able to determine its location on this map from what it sees (i.e. a portion of the points of S translated by $\tilde{P}$ and seen in an arbitrary rotated coordinate system). Clearly, to locate itself the robot can do the following:

> Using the arbitrarily rotated constellation of points of S within its radius of sensing, i.e.
>
> $$\tilde{S}(\tilde{P}, R) = \{\Omega_\theta(P_i - \tilde{P})/P_i \in S,\ d((P_i\tilde{P}) \leq R\}$$
>
> when Ω_θ is a rotation matrix 2×2 about $\tilde{P}$, "search" in S for a similar constellation by checking various center points (2 parameters: $x_{\tilde{p}}, y_{\tilde{p}}$) and rotations (1 parameter: θ). As stated this solution involves a horrendous 3-dimensional search and it must be avoided by using various available tricks like invariant geometric signatures and (geometric) hashing based on "distances" from $\tilde{P}$ to $\Omega_\theta(P_i - \tilde{P})$

> and distances and angles between the P_i's seen from $\tilde{P}$. This leads to much more efficient, Hough-Transform like solutions, for the self location problem. By the way, this is exactly how satelites determine their orientation in space with respect to the constellations of distant stars acting as landmark points!

In the above discussed problem it would help if the points of S would be ordered on a curve, forming, say, a polygonal boundary of a planar region, or would be discrete landmarks on a continuous but clearly visible and definable boundary curve in the plane. Fortunately for those addressing planar shape analysis problems this is most often the case.

11.3 Invariant Boundary Signatures for Recognition under Partial Occlusions

If the shape S is a region of R^2 with a boundary curve $\partial S = C$ that is either smooth or polygonal, we shall have to address the problem of recognizing the shape S from V_ϕ-distorted portions of its boundary. Portions of the boundary, and not the entire boundary because, we must remember, we are dealing with a scenario of possible occlusions. Our claim is that if we can effectively solve the problem of locating a point $\tilde{P}$ on the curve C in a V_ϕ-invariant way based on the "local behavior" of C in a neighborhood of $\tilde{P}$, then we also have a way to detect the possible presence of the shape S from a portion of its boundary. How can we locate $\tilde{P}$ based on the local behavior of C in V_ϕ-invariant ways? We shall have to associate to $\tilde{P}$ a set of numbers ("co-ordinates" or "signature" values) that are invariant under the class of V_ϕ-transformations. To do so, one again has to rely on known geometric invariants of the group of viewing transformation assumed to act on S to produce its image. The fact that we live on a curve C makes our life quite a bit easier.

> As an example, consider first the case where C is a polygonal curve and V_ϕ is the group of Affine-transformations. Since all the viewing transformations map lines into lines and hence the vertices of the poly-line C into vertices of a transformed poly-line $V_\phi(C)$ we can define the local neighborhood of each vertex $C(i)$ of C, as the "ordered" constellation of $2n + 1$ points $\{C(i - n), \ldots, C(i - 1), C(i), C(i + 1), \ldots, C(i + n)\}$ and associate to $C(i)$ invariants of V_ϕ based on this constellation of points. Affine transformations are known to scale areas by the determinant of their associated 2×2 matrix, $\mathbf{A}$, of "shear and scale" parameters, hence we know that ratios of corresponding areas will be affine invariant. Therefore we could consider the areas of the triangles $\Delta_1 = [(C(i-1)C(i)C(i+1)], \Delta_2 = [C(i - 2)C(i)C(i + 2)] \cdots \Delta_n = [C(i - n), C(i), C(i + n)]$ and associate to $C(i)$ a vector of ratios of the type $\{\Delta_k/\Delta_l | k, l =$

$\{1, 2, \ldots, n\}, k \neq l\}$ (See Figure 11.2). This vector will be invariant under the affine group of viewing transformation and will (hopefully) uniquely characterize the point $C(i)$ in an affine-invariant way.

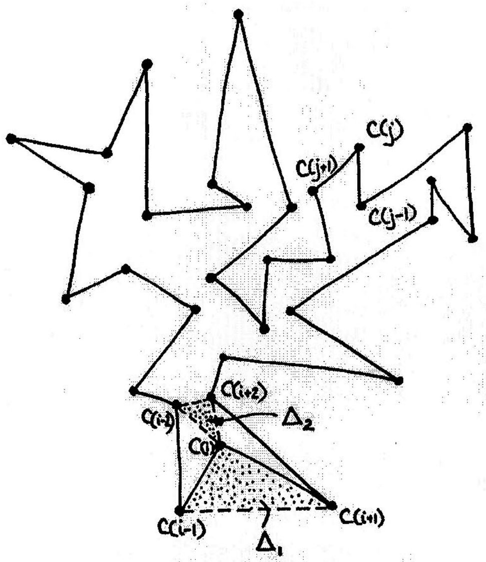

Figure 11.2.

The ideas outlined above provide us a procedure for invariantly characterizing the vertices of a poly-line, however, we can use similar ideas to also locate intermediate points situated on the line segments connecting them. Note that the number n in the example above is a locality-parameter : smaller n's imply more local characterization in terms of the size of neighborhoods on the curve C. Contemplating the foregoing example we may ask how to adapt this method to smooth curves where there are no vertices to enable us to count "landmark" points to the left and to the right of the chosen vertex in view-invariant ways. There is a beautiful body of mathematical work on invariant differential geometry providing differential invariants associated to smooth curves and surfaces, work that essentially carried out Klein's famous Erlangen program for differential geometry, and is reported on in books and papers that appeared many years ago, see ([97], [890], [378], [501] and [136]). Differential invariants enable us to determine a V_ϕ-invariant metric, i.e. a way to measure "length" on the curve C invariantly with respect to the viewing distortion, similar to the way one has, in a straightforward manner, the Euclidean-invariant arclength on smooth curves. If we have an invariant metric, we claim that our problem of invariant point characterizations on C can be readily put in the same framework as in the example of a poly-line. Indeed we can now use the invariant metric to locate to the left and right of $\tilde{P}$ on C (if we define $\tilde{P} \triangleq C(0)$, and describe C as $C(\mu)$ where μ is the invariant metric parameterization of C about $C(0) = \tilde{P}$) the points $\{C(0-n\Delta), \ldots, C(0-\Delta), C(0+\Delta), \ldots, C(0+n\Delta\}$, and these $2n+1$ points now form an invariant constellation of landmarks anchored at $\tilde{P} = C(0)$ (See Figure 11.3). Here Δ is arbitrarily chosen as a small "invariant" distance in terms of the invariant metric. It is very nice to see that letting $\Delta \searrow 0$ one often recovers, from global invariant quantities that were defined on the constellation of points about $C(0) = \tilde{P}$, differential invariant quantities that correspond to known "generalized invariant curvatures" (generalizing the classical curvature obtained if V_ϕ is the simplest, Euclidean viewing distortion). Therefore to invariantly locate a point $\tilde{P}$ on C, we can use the existing V_ϕ invariant metrics on C (note that if C is a polygon - the ordering of vertices is an immediate invariant metric!) to determine about $\tilde{P}$ an invariant constellation of "landmark" points on the boundary curve and use global invariants of V_ϕ to associate to $\tilde{P}$ an "invariant signature vector" $I_{\tilde{P}}(\Delta)$. If $\Delta \searrow 0$ this vector yields, for quite a variety of "good" choices of invariant quantities "generalized invariant curvatures" for the various viewing groups of transformations V_ϕ.

We however do not propose to let $\Delta \searrow 0$. Δ is a locality parameter (as was n before) and we could use several small, but finite, values for Δ to produce (what we can call) a "scale-space" of invariant signature vectors $\{I_{\tilde{P}}\}_{\Delta_i \in Range(of \Delta's)}$.

This freedom allows us to associate to a curve $C(\mu)$, parameterized in terms of its "invariant metric or arclength", a vector valued scale space of signature functions $\{I_P(\mu)\}_{\Delta_i \in Range}$, that will characterize it in both a localized and view-invariant ways. This characterization being local (its locality being in fact under our control via Δ and n) is useful to recognize portions of boundaries in scenes where planar shapes appear both distorted and partially occluded. The recognition process becomes, in terms of the vector-valued signature function, a partial matching algorithm, see [129].

11.4 Invariant Processing of Planar Shapes

Smoothing and other processes of modifying and enhancing planar shapes involves moving their points to new locations. Here we are naturally led to define planar shape deformations or evolutions, by motions of points on the shape boundaries that are small and based on the local geometry, i.e. the geometry of the constellation of other boundary points in the neighborhood. In the spirit of the discussion above, we want to do this in "viewing-distortion-invariant" ways. To do so we have to locate the points of a shape S (or of its boundary $C = \delta S$) and then invariantly move them to new locations in the plane. The discussions of the

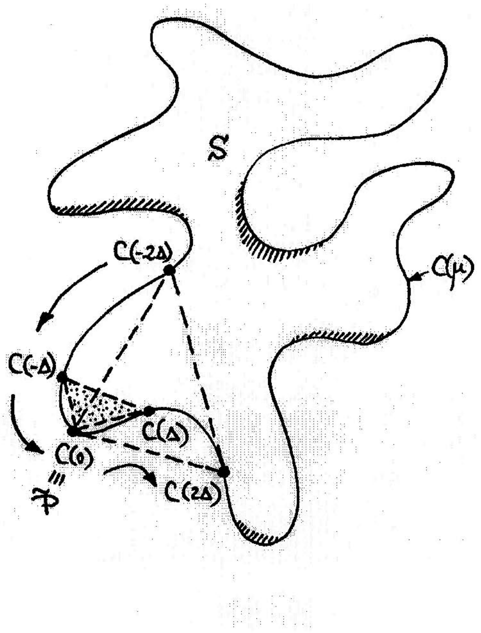

Figure 11.3.

previous sections showed us various ways to invariantly locate points on S or in the plane of S. Moving points around is not much more difficult. We shall have to associate to each point (of S, or in the plane of S) a vector M whose direction and length have been defined so as to take us to another point, in a way that is V_ϕ-invariant. In the example of S being a constellation of points, with a robot using the points of S to locate itself at $\tilde{P}$, we may also want it to determine a new place to go, i.e. to determine a point $\tilde{P}_{new} = \tilde{P}+M$, so as to have the property that from $V_\phi(\tilde{P})$ a robot using the points $\{V_\phi(P_1)\ldots V_\phi(P_N)\}$ will be able to both locate itself and move to $V_\phi(\tilde{P}_{new})$. Of course, on shapes we shall have to do motions that achieve certain goals like smoothing the shape or enhancing it in desirable ways as discussed in [761] . To design view-distortion invariant motions, we can (and indeed must) rely on invariant point characterizations. Suppose we are at a point $\tilde{P}$ on the boundary $C = \delta S$ of a shape S, and we have established a constellation of landmark points about $\tilde{P}$. We can use the invariant point constellation about $\tilde{P}$ to define a V_ϕ-invariant motion from $\tilde{P}$ to $\tilde{P}_{new}$ (See Figure 11.4).

Let us first consider again a very simple example: if V_ϕ is the Affine group of viewing transformations, the centroid of the point constellations about $\tilde{P}$ is an invariantly defined candidate for $\tilde{P}_{new}$. Indeed it is an average of points around $\tilde{P}$ and the process of moving $\tilde{P}$ to such a $\tilde{P}_{new}$ or, differentially, toward such a new position can (relatively easily) be proved to provide an affine invariant shape smoothing operation. If S is a polygonal shape, i.e. $\partial S = C$ is a poly-line, then moving the vertices according to such a smoothing operation can be shown to shrink any shape into a polygonal ellipse, "the affine image of a regular polygon" with the same number of vertices as the original shape. This beautiful result is a generalization of the very early work of Darboux [244], see also [130] and [717], on "a problem in elementary geometry" that addresses the evolutions of planar polygons under an iterative process which replaces the vertices of a polygon by the (ordered) midpoints of its edges. In fact ellipses and polygonal ellipses are the results of many reasonably defined invariant averaging processes [703].

If we are dealing with a smooth continuous boundary curve C and we move the points infinitesimally according to a local "velocity" vector invariantly defined we are in the realm of "geometric" curve evolution processes described by nonlinear partial differential equations. A very prominent recent example of such an evolution process is the Euclidean invariant curve evolution moving the smooth boundary points in the direction of the local normal vector N_c proportionally to the local Euclidean curvature k_c. The temporal evolution of simple closed curves (boundaries of planar shapes) under this rule, described by

$$\frac{\partial}{\partial t} C_t = k_{C_t} N_{C_t} \qquad C_0 = original\ boundary.$$

was thoroughly analyzed and it was proved to smoothly deform and shrink any original curve into an infinitesimal circle, see e.g. [334] and [361]. This nice mathematical result, together with the fact that other Euclidean invariant motions like Blum's prairie fire evolution model [100], [142] which postulates constant velocity motion in the direction of the local normal and leads to "shocks" or "wavefront" collisions that were found useful for shape descriptions since they produce the so-called "shape skeletons", generated a lot of interest and activity in the computer vision community. This activity culminated with the realization that a variety of geometric and viewpoint invariant shape evolutions exist and may be useful in invariant shape analysis and classification. Note that $k_c N_c = \frac{\partial^2 C}{\partial s^2}$, i.e. that in the Euclidean case the invariant vector $k_c N_c$ associated to a point C on a smooth curve is the second derivative of the curve with respect to the Euclidean invariant arclength. This observation yields a very nice interpretation of this invariant evolution: a point on the curve is simply replaced by weighted average of points of the curve of $C(s)$ in the neighborhood with averaging weight dependent on their distance measured in the view invariant metric, from the anchor part C. If the averaging kernel is Gaussian then if is readily seen that a "geometric" diffusion process results, but recall that a variety of local processing and averaging operations are readily available and implementable and should be considered as viable alternatives in generating useful shape evolutions. The process of deriving the invariant evolutions is, of course, readily generalized to more complex viewing transformations and distortions [131], [703].

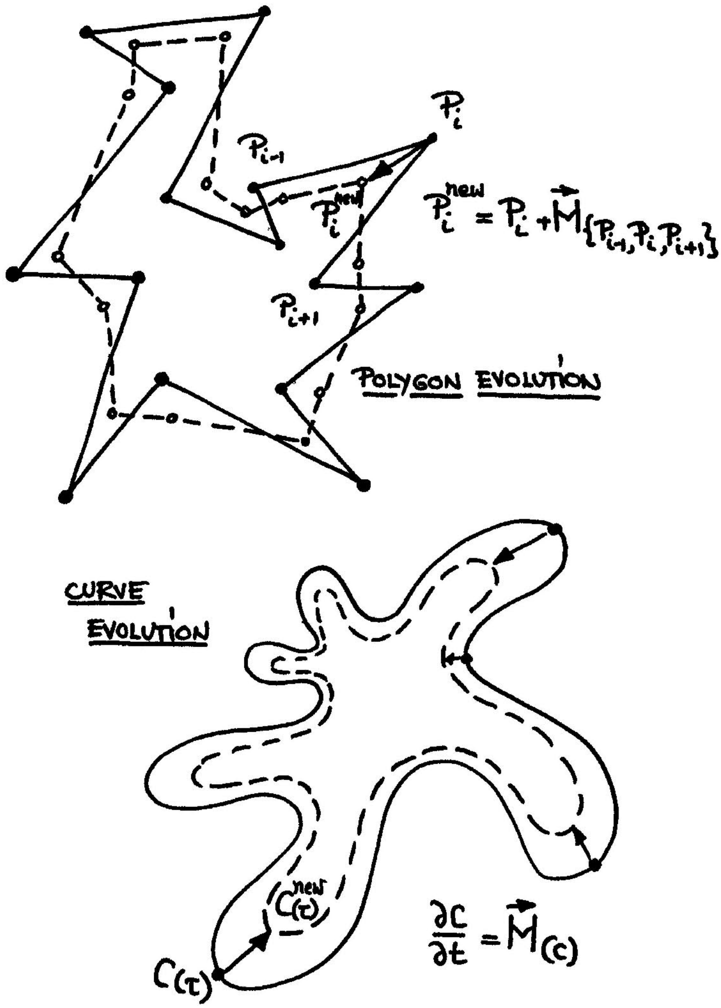

Figure 11.4.

11.5 Concluding Remarks

The main point of this paper is the thesis that in doing "practical" view-point invariant shape recognition or shape processing for smoothing or enhancement, one has to rely on the interplay between global and local (and preferably not differential) invariants of the group of viewing transformations.

Invariant reparameterization of curves based on "adapted metrics" enables us to design generalized and local but not necessarily differential signatures for partially occluded recognition. These signatures have many incarnations, they can be scalars, vectors or even a scale-space of values associated to each point on shape boundaries. They are sometimes quite easy to derive, and generalize the differential concept of "invariant curvature" in meaningful ways. A study of the interplay between local and global invariances of viewing transformations is also very useful for invariant shape smoothing, generating invariant scale-space shape representations, and also leads to various useful invariant shape enhancement an exaggeration operations.

The point of view that geometry is the study of invariances under groups of transformations is, of course, the famous Erlangen program of Felix Klein. Several books appeared over the years that carry out parts of this program for Euclidean affine and projective geometry, see for example Guggenheimer [370], Buchin [136], Blaschke [97], Lane [501]. These theories found their way into the computer vision literature rather late, for example though the works of Weiss [875],[875] [876], [877], [677], Cygansky [242], [833], Abter and Burkhardt [1] and others. The point of view exposed in this paper developed through a series of papers written over many years. These papers, with the details of what is exposed herein, are [127], [126], [128], [130], [131], [129], [132], [761]. Other researchers have made significant contributions to the field and I'll mention the important contributions of Peter Olver [141], [610], Jean-Michel Morel and T. Cohignac [211], [212], Luc Van Gool and his team [829], [830], M. Brill [55], [54], Z. Pizlo and A. Rosenfeld [650], L. Moisan [582], J. Sato and R. Cippola [707], [708] and O. Faugeras [307].

Students, collaborators and academic colleagues and friends have helped me develop the point of view exposed in this paper. I am grateful to all of them for the many hours of discussions and debates on these topics, for agreeing and disagreeing with me, for sometimes fighting and competing, and often joining me on my personal journey into the field of applied invariance theory.