INVARIANTS AND OBJECT RECOGNITION

NIGEL BOSTON
University of Wisconsin, Madison

Face Recognition Group,

February 11, 2004

OUTLINE OF TALK

Methods of Recognition

Types of Invariance

Groups and Manifolds

Moving Frames and Invariants

Differential Invariants

Integral Invariants

Conclusions

METHODS OF RECOGNITION

Pawan Sinha's group at MIT

Three approaches to object recognition:

(a) Transformationist approach (popular in machine vision community)
- compute optimal transformation to bring image and model in register, then match.
Computationally expensive.

(b) View-based approach
- links viewing parameters encountered in learning phase with subsequent performance on recognition tasks.
Expensive on storage.

(c) Invariant-based approach
- encode object views into a compact description of their invariant attributes.
Inexpensive both in computation and storage. Simple and intuitively appealing.

But how DO humans carry out object recognition?!

SINHA'S RESULTS

In collaboration with Tomaso Poggio, Sinha obtained psychophysical evidence that
(i) humans do use a view-based approach;
(ii) humans do use an invariant-based approach;
(iii) invariants are useful for real vision tasks.

The strategy for e.g. (ii) was to employ "targets" and "distractors". Object manipulations that destroyed invariant attributes degraded recognition performance much more than other manipulations of similar magnitude that left the invariants unaffected.

As for (iii), they studied face recognition under variable illumination, seeking an easily computable image invariant immune to changes in lighting. It turned out that the local ordinal structure of the brightnesses at different parts of the face gave a robust invariant immune to lighting variation.

TYPES OF INVARIANCE

See e.g. Amnon Shashua's MIT Ph.D. thesis.

There are four general sources of variability:

(a) Geometric - changes in spatial location of image information.

(b) Photometric - changes in light intensity distribution because of illumination.

(c) Context - objects rarely appear in isolation and differing segmentations or groupings are possible.

(d) Nonrigid characteristics - objects can change shape or have moving parts.

Most work has gone into the first two variabilities. We focus on these. Extensive evidence suggests that humans have elaborate processes for performing (c). As for (d), it can be handled by defining a larger space of admissible transformations.

GROUPS AND MANIFOLDS

Manifold = object that locally looks like an open subset of Euclidean space

Maps between manifolds locally look like maps between Euclidean spaces.

Smooth = infinitely differentiable.

Lie Group = smooth manifold that is also a group.

Interested in Lie groups acting on manifolds.

Example. Let $SO(2)$ be the group of all rotations acting on $\mathbf{R}^2$.
So $SO(2) = \{\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} : 0 \leq \theta < 2\pi\}$, a 1-dimensional Lie group.

Let $SE(2)$ be the group generated by all translations and rotations, acting on $\mathbf{R}^2$.

$SE(2) = \{\begin{pmatrix} \cos(\theta) & -\sin(\theta) & a \\ \sin(\theta) & \cos(\theta) & b \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \theta < 2\pi, a, b \in \mathbf{R}\}$, a 3-dimensional Lie group

In general, a Lie group G acting on a manifold M will be given by a smooth map $\Phi : G \times M \to M, \Phi(g, x) = g.x$

INVARIANTS

The **isotropy subgroup** at $z \in M$ is $G_z := \{g \in G : g.z = z\}$.

The **orbit** of $z \in M$ is $O_z := \{g.z : g \in G\}$.

An **invariant** of G is a real-valued function $I : M \to \mathbf{R}$ such that $I(g.z) = I(z)$ for all $z \in M$ and for all $g \in G$.

Think of M as parametrizing various faces and G as a group of rotations and translations. We want to consider two faces as being the same if one is just the other moved in space, i.e. two faces will be the same iff they are in the same orbit. Invariant features are real functions constant on orbits. Call such faces **equivalent.**

Example Let $G = SO(2)$ acting on $\mathbf{R}^2$. Orbits are circles centered at the origin and for any $z \neq (0,0)$, G_z is trivial. Define invariant $I : \mathbf{R}^2 \to \mathbf{R}$ by $I(x,y) = \sqrt{x^2 + y^2}$.

Note that if $I_1, ..., I_k$ are invariants and $H(y_1, ..., y_k)$ any real-valued function, then $I(x) = H(I_1(x), ..., I_k(x))$ is also an invariant. We seek a complete set of fundamental invariants, meaning that any other invariant can be written as a function of these "fundamental invariants".

The idea is that in face recognition, given a face, represented by $x \in M$, we store $I_1(x), ..., I_k(x)$ and then to check if a face y is equivalent to x, simply compute $I_1(y), ..., I_k(y)$ and compare.

Example, contd. For the $SO(2)$ example above, I is the fundamental invariant in that two points P, Q in $\mathbf{R}^2$ lie in the same orbit if and only if $I(P) = I(Q)$.

MOVING FRAMES

We want some systematic way to produce invariants for a Lie group G acting on a manifold M. One idea, due to Cartan, is moving frames.

A **moving frame** is a smooth G-equivariant map $\rho : M \to G$, i.e. $\rho(g.z) = g\rho(z)$ for all $g \in G, z \in M$.

Theorem A moving frame exists in a neighborhood of a point $z \in M$ if and only if G acts freely and regularly near z.

Proof Suppose a moving frame exists. If $g \in G_z$, then $\rho(z) = \rho(g.z) = g\rho(z)$ implying that $g = 1$, so all G_z are trivial. This is what acting "freely" means.

If $z_n = g_n.z \to z$ as $n \to \infty$, then $\rho(z_n) = \rho(g_n.z) = g_n\rho(z) \to \rho(z)$ and so $g_n \to 1$ as $n \to \infty$. This is what acting "regularly" means. The same as saying that all orbits have the same dimension and the intersection of O_z with neighborhoods of z are connected.

The converse is by a construction.

Example, contd. The $SO(2)$ example has a free and regular action (away from $z = (0, 0)$) and so a moving frame exists.

Invariants arise by taking a set $K \subset M$ such that $K \cap O_z$ has 1 element, e.g. $K = \{(x, y) : y = 0, x > 0\}$ in our example. Elements of K are called canonical forms.

Let G be r-dimensional, M m-dimensional, $r < m$. Pick coordinates of M such that $K = \{(z_1, ..., z_m) : z_1 = c_1, ..., z_r = c_r\}$ Say $g^{-1}.z = (w_1(g, z), ..., w_m(g, z))$. Then the fundamental invariants are $w_{r+1}(\rho(z), z), ..., w_m(\rho(z), z)$.

Example, contd. For $SO(2)$, $r = 1$, $M = \mathbf{R}^2$ and so $m = 2$. Thus there is one fundamental invariant as above. $\rho : M \to G$ takes (x, y) to rotation by θ, where θ is the argument of (x, y).

DIFFERENTIAL INVARIANTS

In general, actions are not free. For example, $SE(2)$ acting on $\mathbf{R}^2$. What can we do?

Idea:- prolong the action to derivatives (jet spaces) leading to differential invariants, or to Cartesian product spaces leading to joint invariants, or to multi-space (both).

Suppose $1 \leq p \leq m-1$. Jet space, $J^n(M,p)$ consists of p-dimensional submanifolds where two are considered the same if they agree to nth order contact. Equivalently, it can be regarded as a set of nth order Taylor polynomials.

This gives, for instance, the **signature curve of a curve** $C \subset \mathbf{R}^2$**.** This curve $S \subset \mathbf{R}^2$ is parametrized by the first two differential invariants κ and κ_s of C.

Theorem. Two curves are equivalent under the action of $SE(2)$ on $\mathbf{R}^2$ if and only if their signature curves are equal.

The big problem with differential invariants is that they depend on high order derivatives and so are very sensitive to noise and of little practical use. What's the remedy?

INTEGRAL INVARIANTS

Following Hann and Hickman "Projective curvature and integral invariants", we extend actions by using potentials rather than derivatives. Consider e.g. Lie group G acting on $\mathbf{R}^2$, written $g.(x,u) = (\overline{x}, \overline{u})$. The jet space approach would be to prolong the action to one on $(x, u, u_x, u_{xx}, ...)$ with more terms than the dimension of G.

The **monomial potential** $V^{i,j}$ of order k $(j \neq 0, i+j=k)$ is given by:
$V_x^{i,j} = x^i u^j$.
Set $z = V^{0,1}, v = V^{1,1}, w = V^{0,2}$. So $z = \int_{x_0}^x u\ dx, v = \int_{x_0}^x xu\ dx, w = \int_{x_0}^x u^2\ dx$.

Hann and Hickman use this to define **potential jet space** with local coordinates $(x, u, x_0, u_0, V^{0,1}, V^{1,1}, V^{0,2}, ...)$, and prove that the action of G prolongs to an action on potential jet space.

Say $(x, u, x_0, u_0, z, v, w, ...) \mapsto (\overline{x}, \overline{u}, \overline{x_0}, \overline{u_0}, \overline{z}, \overline{v}, \overline{w}, ...)$.

Let $G = SE(2) = \{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) & a \\ \sin(\theta) & \cos(\theta) & b \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \theta < 2\pi, a, b \in \mathbf{R}\}$. Then e.g.

$\overline{z} = z + (\cos(\theta)\sin(\theta)/2)(x^2 - x_0^2) - \sin^2(\theta)(xu - x_0u_0) - (\cos(\theta)\sin(\theta)/2)(u^2 - u_0^2) + b\cos(\theta)(x - x_0) - b\sin(\theta)(u - u_0)$

and there are longer explicit expressions for $\overline{v}, \overline{w}$ etc. From this they get a moving frame, and so the group action is free and regular, and so a signature curve can be defined which characterizes all curves up to equivalence as before. The coordinates involve combinations of point values and areas under the graphs of u, xu, u^2.

Note that these integral invariants are different from the traditional moment invariants, since the moment invariants are global invariants whereas these invariants are semi-local since by varying (x_0, u_0) it can be defined on any segment of the curve.

FURTHER DEVELOPMENTS

Soatto's group at UCLA have indicated that they are working on integral invariants for shape description. They have yet to publish anything in this direction (except a webpage). They want to use integral invariants to define a "shape distance". One method is to use the invariant values to match points on the curves, and with additional constraints such as that the points should match in order, the distance and correspondence between points on a curve become the optimization of a functional. (Cf. Faugeras' approach to correspondence problem.) Their webpage focuses on 2D recognition, in particular hands.

Others, e.g. Sean He, have proposed using Spiral Architecture, a data structure representing an image as a collection of hexagons. They propose integral invariants should be the object representation. Not much published and has focused on 2D recognition. They note that Spiral Architecture is inspired by anatomical considerations of a primate's vision system, in that the geometrical arrangement of cones on the primate's retina can be described in terms of a hexagonal grid. Computationally powerful.

CONCLUSIONS

1. Integral invariants are more robust than the traditional differential invariants, but have only been used in 2D recognition so far. We should implement the 3D case.

2. To that end, viewing faces as surfaces in $\mathbf{R}^3$, acted on by rotations and translations, traditionally a 2-dimensional signature manifold in $\mathbf{R}^6$ is associated to each face, with the property that two faces are equivalent iff they have the same signature manifold. Integral invariants should produce a more robust signature manifold in this case.

3. As noted by Sinha, humans really do use invariant-based recognition. It's computationally and storage most efficient. We store the signature manifolds in our database and to compare a face we compute its signature manifold and compare.

4. Sinha also notes that variations in lighting can be handled by local ordinal relationships. These can be viewed as comparing semi-local integrals over intensity.

5. As noted by Shashua, geometric and photometric variations should be studied in tandem, so combining work on semi-local integral invariants both in the geometric and photometric sense could link Sinha and Hann-Hickman for greatest success.