Pattern Recognition Letters 12 (1991) 519-523
North-Holland

September 1991

A Lie group theoretic approach to the invariance problem in feature extraction and object recognition

James B. Cole*, Hiroshi Murase and Seiichiro Naito

NTT Basic Research Laboratory, Musashino-Shi, Tokyo 180, Japan

Received 5 April 1991

Abstract

Cole, J.B., H. Murase and S. Naito, A Lie group theoretic approach to the invariance problem in feature extraction and object recognition, Pattern Recognition Letters 12 (1991) 519-523.

We derive a formal solution to the invariance problem and construct it using Lie group generators. Representations of these generators with respect to image data are discussed. Group theoretical obstacles to three-dimensional invariant recognition and possible solutions are considered.

Keywords. Invariance problem, Lie group generators.

1. Introduction

A major goal of image analysis is to recognize objects and to extract features from their images independently of the positions, orientations and sizes of the objects.

Let x be a set of image data from some object, and let f be a function which operates on x to yield the information $f(x)$. If the image data change from x to Sx when the object 'moves' in some way (e.g., translation or rotation), then in general $f(Sx) \neq f(x)$. S is an operator that acts on x, the 'old' data to give the new image data, Sx. Since the original object remains physically unchanged, one would like to find an 'invariant' function g such that

$$g(Sx) = g(x). \tag{1}$$

This is a very difficult problem to solve in general.

If S is an element of a Lie group, however, it is possible to construct a formal solution of (1) (see for example (Belinfante and Kolman, 1972) or (Schultz, 1990) for a review of Lie groups). We seek an operator, T, that acts on f such that

$$Tf(Sx) = f(x). \tag{2}$$

T changes f in such a way as to 'compensate' for the change from x to Sx. It is straightforward, as we will show, to find T.

*Now at Naval Research Laboratory Washington, DC 20375-5000, USA.

2. Formal solution

We assume that the image data can be represented in the form $x=(x^1,x^2,\ldots)$, where x^i $(i=1,2,3,\ldots)$ are 'coordinates' that could represent the pixel brightnesses in the scene or which could be the coefficients in some kind of 'expansion' such as that described by (Roseborough and Murase, 1990). For simplicity, we take $f(x)$ to be a real scalar function. Let $S(\theta)$ be a transformation characterized by the parameter set $\theta=(\theta^1,\theta^2,\theta^3,\ldots)$. S can be any operation such as translation, rotation, or dilation that forms a Lie group. An infinitesimal transformation is given by

$$S(\mathrm{d}\theta) = \mathbf{1} + \mathrm{d}\theta^i G_i \tag{3}$$

where the G_i are the group generators and $\mathrm{d}\theta^i$ are infinitesimal. Throughout this paper, unless otherwise specified, we sum on upper and lower repeated indices in accord with the Einstein summation convention. A group generator is defined as

$$\begin{aligned} G_i &= \lim_{\theta\to 0} (S(\theta)-\mathbf{1})/\theta^i \\ &= \frac{\partial}{\partial\theta^i} S(\theta)\bigg|_{\theta=0}, \end{aligned} \tag{4}$$

where $S(0)=\mathbf{1}$ by definition. Applying an infinitesimal transformation, $S(\mathrm{d}\theta)$, to x, we obtain

$$S(\mathrm{d}\theta)\,x = x + \mathrm{d}\theta^i G_i x. \tag{5}$$

Now making the substitutions $x \to S(\mathrm{d}\theta)^{-1}x$ and $S \to S(\mathrm{d}\theta)$ in (2), where $S(\theta)^{-1}$ denotes the inverse of $S(\theta)$, we find

$$Tf(x) = f(S(\mathrm{d}\theta)^{-1}x). \tag{6}$$

For Lie groups

$$S(\mathrm{d}\theta)^{-1} = S(-\mathrm{d}\theta) = \mathbf{1} - \mathrm{d}\theta^i G_i,$$

thus

$$S(\mathrm{d}\theta)^{-1}x = x - \mathrm{d}\theta^i G_i x.$$

Substituting into (6) and expanding in a Taylor series up to the first order in $\mathrm{d}\theta$, we obtain

$$Tf(x) = f(x) - \mathrm{d}\theta^i (G_i x)^j\, \partial_j f(x), \tag{7}$$

where $(G_i x)^j$ denotes the j-th component of $(G_i x)$, and $\partial_j f(x)$ denotes the partial derivative of f with respect to x^j. T thus depends on θ and its infinitesimal form, $T(\mathrm{d}\theta)$, is given by

$$T(\mathrm{d}\theta) = \mathbf{1} - \mathrm{d}\theta^i (G_i x)^j\, \partial_j. \tag{8}$$

The generators of T, H_i, are thus

$$H_i = -(G_i x)^j\, \partial_j. \tag{9}$$

From the theory of Lie groups, a finite Lie group transformation can be expressed in terms of its generators by

$$T(\theta) = \mathrm{e}^{\theta^i H_i}. \tag{10}$$

See (Lenz, 1990) or (Kanatani, 1990) for more details of this derivation.

Defining $f^*(x,\theta) = T(\theta)\,f(x)$ and expanding (10) we obtain

$$\begin{aligned} f^*(x,\theta) &= T(\theta)\,f(x) = \mathrm{e}^{-\theta^i (G_i x)^j \partial_j} f(x) \\ &= f(x) - \theta^i (G_i x)^j\, \partial_j f(x) \\ &\quad + \frac{1}{2!}\,\theta^i\theta^j (G_i x)^l\, \partial_l ((G_j x)^k\, \partial_k f(x)) \\ &\quad + \cdots. \end{aligned} \tag{11}$$

$f^*(x,\theta)$ satisfies (1) in the sense that

$$f^*(x,\theta) = f^*(S(\theta)^{-1}x, 0).$$

Although f is arbitrary, in order to actually construct $f^*(x,\theta)$, its derivatives must be sufficiently 'well behaved' that the series in (11) converges. Although $f^*(x,\theta)$ is an invariant function, it contains θ, whose value is usually not accessible from raw image data.

But knowing T, another approach is now possible. We can attempt to solve the differential equation

$$T(\theta)f = f \tag{12}$$

for f. This is usually a difficult problem but it is relatively easy to find a class of 'first-order' solutions, as we shall later see.

3. Constructing an invariant image function

We now discuss how to actually construct $f^*(x,\theta)$ with respect to some transformation. For simplicity, we take $S(\theta)$ to be a one-parameter Lie group.

Making the replacements $G_i \to G$ and $\theta^i \to \theta$, (11) becomes

$$f^*(x,\theta) = f(x) - \theta (Gx)^i \partial_i f(x) + \frac{1}{2!}\theta^2 (Gx)^j \partial_j ((Gx)^i \partial_i f(s)) - \frac{1}{3!}\theta^3 (Gx)^k \partial_k ((Gx)^j \partial_j ((Gx)^i \partial_i f(x))) + \cdots. \tag{13}$$

Calculating the derivatives, we obtain

$$\begin{aligned} f^*(x,\theta) &= f(x) - \theta (Gx)^i \partial_i f(x) \\ &+ \frac{1}{2!}\theta^2 \{(Gx)^i (Gx)^j \partial_j \partial_i f(x) + (G^2x)^i \partial_i f(x)\} \\ &- \frac{1}{3!}\theta^3 \{(Gx)^i (Gx)^j (Gx)^k \partial_k \partial_j \partial_i f(x) \\ &\quad + 2(G^2x)^i (Gx)^j \partial_i \partial_j f(x) + (G^3x)^i \partial_i f(x)\} \\ &+ \cdots. \end{aligned} \tag{14}$$

Now if we choose $f = f^{(1)}$, where $f^{(1)}$ satisfies

$$(Gx)^i \partial_i f^{(1)} = 0, \tag{15}$$

the second term on the right in (14) vanishes along with parts of the higher order terms. By postulating a form for $f^{(1)}$, a class of 'first-order' solutions to (12) can be determined. Such solutions may be useful in practical problems.

Now, to proceed further, we must determine the group generator, G.

4. Lie group generators for image data

Let us represent x as an $n \times 1$ column vector, and let $\{e_i\}$ $(i = 1, 2, \ldots, n)$ be any complete set of orthonormal basis vectors that could be used to represent the image data, such as the Walsh patterns. Expanding x in the form $x = \sum_i x^i e_i$, a representation for G in terms of an $n \times n$ square matrix in terms of the $\{e_i\}$ can be found. Rearranging (5), making the substitutions $x \to e_j$, $G_i \to G$ and multiplying on the left by e_i^{T}, we find, using the definition of G,

$$\begin{aligned} e_i^{\mathrm{T}}(S(\mathrm{d}\theta)e_j - e_j) &= \mathrm{d}\theta\, e_i^{\mathrm{T}} G e_j \\ &= \mathrm{d}\theta\, G_j^i. \end{aligned} \tag{16}$$

To determine G, we apply an infinitesimal transformation, $S(\mathrm{d}\theta)$, to e_j and expand $S(\mathrm{d}\theta)e_j - e_j$ in the form

$$S(\mathrm{d}\theta)e_j - e_j = \sum_i (\mathrm{d}e_j)^i e_i,$$

where the $(\mathrm{d}e_j)^i$ are the expansion coefficients. The elements of G are thus

$$G_j^i = (\mathrm{d}e_j)^i / \mathrm{d}\theta.$$

Defining $S(\theta)e_j = e_j(\theta)$, we can write

$$G_j^i = e_i^{\mathrm{T}} \frac{\mathrm{d}}{\mathrm{d}\theta} e_j(0), \tag{17}$$

where $(\mathrm{d}/\mathrm{d}\theta)e_j(0)$ denotes the derivative at $\theta = 0$. G is thus determined in the form of an $n \times n$ matrix. This procedure can be carried out either 'theoretically' or 'empirically' using (16) or (17).

From (16), $G_j^i = e_i^{\mathrm{T}} G e_j$. Inserting a basis vector expansion of x into $(Gx)^i$, it is easy to show that $(Gx)^i = x^j G_j^i$, where repeated upper and lower indices are summed as usual. T, as given in (14), can then be expressed as

$$f^*(x,\theta) = f(x) - \theta x^j G_j^i \partial_i f(x) + \frac{1}{2!}\theta^2 \{x^k x^l G_k^i G_l^j \partial_j \partial_i f(x) + x^j (G^2)_j^i \partial_i f(x)\} - \cdots. \tag{18}$$

The foregoing developments have been carried out without reference to any specific set of basis vectors. To actually construct f^*, however, a 'suitable' set must be selected, and their transformation properties defined. These choices then determine the generator representations. These choices are not completely arbitrary. Generators of such physical operations as translation and rotation must reflect their corresponding physical properties. Translation generators in x- and y-directions, G_X and G_Y, respectively, must obey $[G_X, G_Y] = 0$ (where $[a,b] = ab - ba$), since translating first in the x-direction and then in the y-direction should give the same result as performing these operations in the reverse order.

In a future paper we shall discuss representations of image data that are suitable for the analysis outline above, but we illustrate our approach with a simple example.

6. Example

Let $\boldsymbol{v}$ be an 'ordinary' 2-dimensional column vector, given by

$$\boldsymbol{v}^{\mathrm{T}} = x\,\boldsymbol{e}_1 + y\,\boldsymbol{e}_2,$$

where

$$\boldsymbol{e}_1^{\mathrm{T}} = (1\ \ 0) \quad \text{and} \quad \boldsymbol{e}_2^{\mathrm{T}} = (0\ \ 1)$$

are basis vectors. Let $\boldsymbol{e}_j(\theta) = S(\theta)\boldsymbol{e}_j$ be a basis vector that has been rotated by angle θ. We can determine 'experimentally' that

$$\boldsymbol{e}_1(\theta) = (\cos\theta)\boldsymbol{e}_1 + (\sin\theta)\boldsymbol{e}_2$$

and

$$\boldsymbol{e}_2(\theta) = (-\sin\theta)\boldsymbol{e}_1 + (\cos\theta)\boldsymbol{e}_2.$$

Using (17) we calculate the $G_j^i = \boldsymbol{e}_i \cdot (\mathrm{d}/\mathrm{d}\theta)\boldsymbol{e}_j(0)$ and find that

$$G_2^1 = -1, \quad G_1^2 = 1, \quad \text{and} \quad G_1^1 = G_2^2 = 0.$$

Now let $f = f(x, y)$ be some arbitrary function. Substituting f and G into (18), we construct a function, f^*, that is invariant under rotations according to

$$f^*(x, y, \theta) = f(x, y) - \theta x\,\partial_y f(x, y) + \theta y\,\partial_x f(x, y) + \cdots.$$

A first-order invariant function with respect to G must satisfy

$$x\,\partial_y f - y\,\partial_x f = 0.$$

It is not difficult to show that one such function is

$$f(x, y) = x^2 + y^2.$$

7. Discussion

Most image data are obtained by gathering the light reflected from the surface of a three-dimensional object onto a two-dimensional plane. In this case, S can be expressed in the form $S = PQ$, where P is a 'projection' operator, and Q is some operation (such as rotation) on a physical object of which we have a two-dimensional image. Unfortunately $S = PQ$ is not necessarily an element of a Lie group even when Q is. With respect to pixel brightness data alone, PQ can be guaranteed to form a Lie group only in certain special cases, such as when Q is restricted to rotations about an axis perpendicular to the plane of projection or when the object has special symmetries. The requirement that PQ form a group thus constitutes a severe constraint on Q with respect to pixel brightness data.

If we use a set of basis vectors that contain 'extra' information about an object (for example, the appearance of occluded surfaces or texture information), such as the eigenvectors described by (Roseborough and Murase, 1990), S may have Lie group properties even though it is not a Lie group element with respect to a pixel brightness representation. This topic will be explored in a later paper.

The human perception systsm is able to solve the invariance problem in a wide variety of situations even when S is not a Lie group element with respect to pixel brightness data. The perception system utilizes many cues such as shading and curvature as well as prior knowledge along with a priori assumptions to extract information from an image. We might say that this process amounts to a modification of P to a new operator P^* such that $S = P^*Q$ is an element of a Lie group for a wide class of operations, Q. An interesting topic for future work is to study P^*.

We wander from a firm mathematical foundation, but even if $S = PQ$ only 'approximately' satisfies the properties of a Lie group, $f^*(x, \theta)$ as given by (11) may be 'approximately' invariant and 'reasonable' first-order solutions of (12) may still be possible. Such questions can only be addressed empirically.

8. Summary

If $S(\theta)$ is a Lie group transformation on a set of image data, x, an invariant function $f^*(x, \theta)$ can be constructed according to (11) that formally satisfies (1). To find an invariant function that is independent of θ, one must solve (12). General solutions are difficult, but it is relatively easy to determine a set of 'first-order' solutions by solving (15). Although it is difficult to actually solve (12), eqs. (14) and (18) yield criteria for constructing image invariants. If we want invariance with respect to more than one image transformation, the general development of Section 2 must be applied.

The requirement that S be a member of a Lie group constitutes a severe constraint with respect to pixel brightness data alone, but may be less of a constraint with respect to other image representations.

References

Belinfante, J.G.F. and B. Kolman (1972). *A Survey of Lie Groups and Lie Algebras with Applications to Computational Methods.* SIAM, Philadelphia, PA.

Kanatani, K. (1990). *Group Theoretical Methods in Image Understanding.* Springer, Berlin.

Lenz, R. (1990). *Group Theoretical Methods in Image Processing.* Springer, Berlin.

Roseborough, J. and H. Murase (1990). Partial eigenvalue decomposition for very large matrices using run-length encoding. NTT Basic Research Laboratory, Internal Report, Tokyo. Submitted to *IEEE Trans. Pattern Anal. Machine Intell.*

Schutz, B.F. (1985). *Geometrical Methods of Mathematical Physics.* Cambridge Univ. Press, London.