Measurement of Motion-Induced Image Deformations: Spatio-Temporal Operators for Translation, Divergence, Curl, and Shear

Roy Eagleson

Centre for Cognitive Science
The University of Western Ontario,
(rm SS7332) London, Ontario, Canada, N6A 5C2
elroy@uwo.ca, elroy@julian.uucp, elroy@uwovax.bitnet

KEYWORDS: Motion Analysis, Lie Transformation Group, Spatio-Temporal Operators, Group Theoretic Approach.

ABSTRACT

Many approaches to motion understanding in computer vision are based first on computing the 'optical flow' of a set of time-varying images. Ideally, this vector field corresponds to the 2D translations of local regions of the intensity profile which shift as a result of the 3D motions in the visible world. If this vector field could be computed, it is argued that is would provide an intermediate representation that could be used as input data to higher-level motion analysis algorithms. However; this is an ill-posed problem, due to the underconstrained nature of the projection from 3D to 2D. Many researchers have shown (cf. Koenderink and van Doorn[1], Longuet-Higgins and Prazdny[2], Kanatani[3],[4]) that it is *in principle* possible to estimate the local motion and low-level structure of a visible surface if additional information is added by obtaining higher-order spatio-temporal derivatives of the changing image signal. This information must be combined with additional knowledge about rigid bodies and the projection from 3D to 2D.

The six degrees-of-freedom of the Euclidean rigid-body motion group [Given by the semi-direct product of 3D translations and rotations, (cf. Carlton and Shepard[5]), $E^+ = \Re^3 \circledS SO(3)$] project perspectively to a unique six-dimensional vector field group. (cf. Blicher and Omohundro[6]). It is proposed that the structure and motion of a visible surface can be computed by integrating these infinitesimal deformations over space and time; incorporating the objects' spatio-temporal continuity properties. However; this is not a "structure-from-motion" paper. Rather; it deals with the estimation of a six-element group of local coordinate transformations in the image. The group is proposed as a set of measures for computing structure and relative motion. Despite this formulation, the problem would remain ill-posed when implemented using spatio-temporal derivatives, due to the instability of gradient operators with respect to added noise. Because of this, *six* orthogonal spatio-temporal *filters* are proposed. They extract the translation, divergence, curl, and 'shear', which are independent deformations that can be measured locally in the image. More importantly, they cover the six degrees-of-freedom of motion in the 3D scene.

[1]This research was supported in part by NSERC operating grant number A2600 awarded to Dr. Zenon Pylyshyn.

1 Introduction

"Motion Understanding" in the field of Computer Vision requires that sequences of 2D images be analysed for changes that result from the relative 3D motions of objects with respect to the observer. In order to implement this process on a machine vision system, it must be described with mathematical and computational rigour. Due to the ill-posed nature of this 'inverse optics' problem, there are two major hurdles; the first being the measurement of these motion-induced image deformations, and the second being the interpretation of this vector field.

Many approaches to motion understanding in computer vision are based first on computing the 'optical flow' of a set of time-varying images. Classically, this vector field corresponds to the 2D translations of local regions of the intensity profile which shift as a result of the 3D motions in the visible world. If this vector field could be computed, it is argued that is would provide an intermediate representation that could be used as input data to higher-level motion analysis algorithms. This paper departs from this approach essentially by proposing that the 'optical flow' field can be regarded as a 2D field of six-dimensional vectors, rather than a 2D field of two-dimensional vectors.

1.1 Ill-posed Nature of Inverse Optics

Perceiving the 3D structure of the world from a 2D image can be described as solving an inverse problem. Since the projection removes one dimension (depth), solving for the inverse is inherently underconstrained. Additional information must be added to constrain the missing degree of freedom. The inverse problem of finding the structure of an object given only its projected motion in the image is an ill-posed problem unless other constraints or information is added. The requirements for a problem to be well-posed are that; (1) A solution exists, (2) The solution is unique, and (3) There are no discontinuities in the relationship between the solution result and the amount of added input noise. ("Graceful degradation")

There are many cues to structure and environment layout that can be used by a vision system. A system which uses two spatially-displaced images of the same scene can recover depth from stereopsis. In addition, there are monocular cues to the shape of an object from its shading in the image. In fact, Cavanagh[7] shows evidence