ELSEVIER

Artificial Intelligence 77 (1995) 395–400

Artificial Intelligence

Forthcoming Papers

Special Volume on Vision
(N. Ahuja and R. Horaud, Guest Editors)

R. Basri and E. Rivlin, Localization and homing using combinations of model views

Navigation involves recognizing the environment, identifying the current position within the environment, and reaching particular positions. We present a method for *localization* (the act of recognizing the environment), *positioning* (the act of computing the exact coordinates of a robot in the environment), and *homing* (the act of returning to a previously visited position) from visual input. The method is based on representing the scene as a set of 2D views and predicting the appearances of novel views by linear combinations of the model views. The method accurately approximates the appearance of scenes under weak-perspective projection. Analysis of this projection as well as experimental results demonstrate that in many cases this approximation is sufficient to accurately describe the scene. When weak-perspective approximation is invalid, either a larger number of models can be acquired or an iterative solution to account for the perspective distortions can be employed.

The method has several advantages over other approaches. It uses relatively rich representations; the representations are 2D rather than 3D; and localization can be done from only a single 2D view without calibration. The same principal method is applied for both the localization and positioning problems, and a simple "qualitative" algorithm for homing is derived from this method.

Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry

This paper proposes a robust approach to image matching by exploiting the only available geometric constraint, namely, the epipolar constraint. The images are uncalibrated, namely the motion between them and the camera parameters are not known. Thus, the images can be taken by different cameras or a single camera at different time instants. If we make an exhaustive search for the epipolar geometry, the complexity is prohibitively high. The idea underlying our approach is to use classical techniques (correlation and relaxation methods in our particular implementation) to find an initial set of matches, and then use a robust technique—the Least Median of Squares (LMedS)—to discard false matches in this set. The epipolar geometry can then be accurately estimated using a meaningful image criterion. More matches are eventually found, as in stereo matching, by using the recovered epipolar geometry. A large number of experiments have been carried out, and very good results have been obtained.

Regarding the relaxation technique, we define a new measure of matching support, which allows a higher tolerance to deformation with respect to rigid transformations in the image plane and a smaller contribution for distant matches than for nearby ones. A new strategy for updating matches is developed, which only

selects those matches having both high matching support and low matching ambiguity. The update strategy is different from the classical "winner-take-all", which is easily stuck at a local minimum, and also from "loser-take-nothing", which is usually very slow. The proposed algorithm has been widely tested and works remarkably well in a scene with many repetitive patterns.

A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, N. Pillow, 3D object recognition using invariance

The systems and concepts described in this paper document the evolution of the geometric invariance approach to object recognition over the last five years. Invariance overcomes one of the fundamental difficulties in recognising objects from images: that the appearance of an object depends on viewpoint. This problem is entirely avoided if the geometric description is unaffected by the imaging transformation. Such invariant descriptions can be measured from images without any prior knowledge of the position, orientation and calibration of the camera. These invariant measurements can be used to index a library of object models for recognition and provide a principled basis for the other stages of the recognition process such as feature grouping and hypothesis verification. Object models can be acquired directly from images, allowing efficient construction of model libraries without manual intervention.

A significant part of the paper is a summary of recent results on the construction of invariants for 3D objects from a single perspective view. A proposed recognition architecture is described which enables the integration of multiple general object classes and provides a means for enforcing global scene consistency.

Various criticisms of the invariant approach are articulated and addressed.

J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis and F. Nuflo, Modeling visual attention via selective tuning

A model for aspects of visual attention based on the concept of selective tuning is presented. It provides for a solution to the problems of selection in an image, information routing through the visual processing hierarchy and task-specific attentional bias. The central thesis is that attention acts to optimize the search procedure inherent in a solution to vision. It does so by selectively tuning the visual processing network which is accomplished by a top-down hierarchy of winner-take-all processes embedded within the visual processing pyramid. Comparisons to other major computational models of attention and to the relevant neurobiology are included in detail throughout the paper. The model has been implemented; several examples of its performance are shown. This model is a hypothesis for primate visual attention, but it also outperforms existing computational solutions for attention in machine vision and is highly appropriate to solving the problem in a robot vision system.

R.P.N. Rao and D.H. Ballard, An active vision architecture based on iconic representations

Active vision systems have the capability of continuously interacting with the environment. The rapidly changing environment of such systems means that it is attractive to replace static representations with visual routines that compute information on demand. Such routines place a premium on image data structures that are easily computed and used.

The purpose of this paper is to propose a general active vision architecture based on efficiently computable iconic representations. This architecture employs two primary visual routines, one for identifying the visual image near the fovea (object identification), and another for locating a stored prototype on the retina (object location). This design allows complex visual behaviors to be obtained by composing these two routines with different parameters.

The iconic representations are comprised of high-dimensional feature vectors obtained from the responses of an ensemble of Gaussian derivative spatial filters at a number of orientations and scales. These representations are stored in two separate memories. One memory is indexed by image coordinates while the other is indexed by object coordinates. Object location matches a localized set of model features with image features at all possible retinal locations. Object identification matches a foveal set of image features with all possible model features. We present experimental results for a near real-time implementation of these routines on a pipeline image processor and suggest relatively simple strategies for tackling the problems of occlusions and scale variations. We also discuss two additional visual routines, one for top-down foveal targeting using log-polar sensors and another for looming detection, which are facilitated by the proposed architecture.

K.N. Kutulakos and C.R. Dyer, Global surface reconstruction by purposive control of observer motion

What viewpoint-control strategies are important for performing *global* visual exploration tasks such as searching for specific surface markings, building a global model of an arbitrary object, or recognizing an object? In this paper we consider the task of *purposefully* controlling the motion of an active, monocular observer in order to recover a global description of a smooth, arbitrarily-shaped object. We formulate global surface reconstruction as the task of controlling the motion of the observer so that the visible rim slides over the maximal, connected, reconstructible surface regions intersecting the visible rim at the initial viewpoint. We show that these regions are bounded by a subset of the visual event curves defined on the surface.

By studying the epipolar parameterization, we develop two basic strategies that allow reconstruction of a surface region around any point in a reconstructible surface region. These strategies control viewpoint to achieve and maintain a well-defined geometric relationship with the object's surface, rely only on information extracted directly from images (e.g., tangents to the occluding contour), and are simple enough to be performed in real time. We then show how global surface reconstruction can be provably achieved by (1) appropriately integrating these strategies to iteratively "grow" the reconstructed regions, and (2) obeying four simple rules.

Y. Yang and A.L. Yuille, Multilevel enhancement and detection of stereo disparity surfaces

The problem of stereo vision has been of increasing interest to the computer vision community over the past decade. This paper presents a new computational framework for matching a pair of stereo images arising from viewing the same object from two different positions. In contrast to previous work, this approach formulates the matching problem as detection of a "bright", coherent disparity surface in a 3D image called the *spatio-disparity space* (SDS) image. The SDS images represents the goodness of each and every possible match.

A nonlinear filter is proposed for enhancing the disparity surface in the SDS image and for suppressing the noise. This filter is used to construct a hyperpyramid representation of the SDS image. Then the disparity surface is detected using a coarse-to-fine control structure. The proposed method is robust to photometric and geometric distortions in the stereo images, and has a number of computational advantages. It produces good results for complex scenes.

D.A. Reece and S.A. Shafer, Control of perceptual attention in robot driving

Computer vision research aimed at performing general scene understanding has proven to be conceptually difficult and computationally complex. *Active vision* is a promising approach to solving this problem. Active vision systems use optimized sensor settings, reduced fields of view, and relatively simple algorithms to efficiently extract specific information from a scene. This approach is only appropriate in the context of a *task* that motivates the selection of the information to extract. While there has been a fair amount of research that describes the extraction processes, there has been little work that investigates how active vision could be

used for a realistic task in a dynamic domain. We are studying such a task: driving an autonomous vehicle in traffic.

In this paper we present a method for controlling visual attention as part of the reasoning process for driving, and analyze the efficiency gained in doing so. We first describe a model of driving and the driving environment, and estimate the complexity of performing the required sensing with a general driving-scene understanding system. We then introduce three programs that use increasingly sophisticated perceptual control techniques to select perceptual actions. The first program, called Ulysses-1, uses *perceptual routines*, which use known reference objects to guide the search for new objects. The second program, Ulysses-2, creates an inference tree to infer the effect of uncertain input data on action choices, and searches this tree to decide which data to sense. Finally, Ulysses-3 uses domain knowledge to reason about how dynamic objects will move or change over time; objects that do not move enough to affect the robot's decisions are not selected as perceptual targets. For each technique we have run experiments in simulation to measure the cost savings realized by using selective perception. We estimate that the techniques included in Ulysses-3 reduce the computational cost of perception by 9 to 12 orders of magnitude when compared to a general perception system.

Il-Pyung Park and J.R. Kender, Topological direction-giving and visual navigation in large environments

In this paper, we propose and investigate a new model for robot navigation in large unstructured environments. Current models, which depend on metric information, have to deal with inherent mechanical and sensory errors. Instead we supply the navigator with qualitative information. Our model consists of two parts, a map-maker and a navigator. Given a source and a goal, the map-maker derives a navigational path based on the topological relationships between landmarks. A navigational path is generated as a combination of "parkway" and "trajectory" paths, both of which are abstractions of the real world into topological data structures. Traversing within a parkway enables the navigator to follow landmarks that are continuously visible. Traversing on a trajectory enables the navigator to move reliably into featureless space, based on local headings formed by visible landmarks that are robust to positional and orientational errors. Reliability measures of parkway and trajectory traversals are defined by appropriate error models that account for the sensory errors of the navigator, the population of neighboring objects, and the rotational and transitional errors of the navigator. The optimal path is further abstracted into a "custom map", which consists of a list of symbolic directional instructions, the vocabulary of which is defined by our environmental description language. Based on the custom map generated by the map-maker, the navigating robot looks for events that are characterized by spatial properties of the environment. The map-maker and the navigator are implemented using two cameras, an IBM 7575 robot arm, and a PIPE (Pipelined Image Processing Engine).

N. Gupta and L. Kanal, 3-D motion estimation from motion field

Several experiments suggest that the first stage of motion perception is the measurement of visual motion. The result of this stage is called the *motion field*, which assigns a velocity vector to each point in the image plane. The second stage involves interpreting the motion field in terms of objects and motion in the three-dimensional world. Recovering 3-D motion of the object from the motion field has been difficult owing to the nonlinear system of equations involved, and the sensitivity of the system to noise. The need for the stability of the system is essential as only the optical flow field can be recovered from a sequence of images, which is at best a crude approximation to the motion field.

We define two sets of "basic" parameters, which can be recovered from the motion field by solving a linear system of equations. The relationship between the basic parameters and the motion parameter being one-to-one and linear, we obtain a closed form solution for the 3-D motion parameter by solving a system of linear equations only. We prove the correctness, completeness and robustness of the approach and in that sense the problem of recovering the motion parameter from the motion field may be said to be "solved". We present the results of extensive experimentation with real and simulated image sequences.

A. Blake, M. Isard and D. Reynard, Learning to track the visual motion of contours

A development of a method for tracking visual contours is described. Given an "untrained" tracker, a training motion of an object can be observed over some extended time and stored as an image sequence. The image sequence is used to learn parameters in a stochastic differential equation model. These are used, in turn, to build a tracker whose predictor imitates the motion in the training set. Tests show that the resulting trackers can be markedly tuned to desired curve shapes and classes of motions.

M. Otte and H.-H. Nagel, Estimation of optical flow based on higher order spatiotemporal derivatives in interlaced and non-interlaced image sequences

This contribution investigates local differential techniques for estimating optical flow and its derivatives based on the brightness change constraint. By using the tensor calculus representation we build the Taylor expansion of the gray-value derivatives as well as of the optical flow in a spatiotemporal neighborhood. Such a formulation provides a unifying framework for all existing local differential approaches and allows to derive new systems of equations for the estimation of the optical flow and of its derivatives.

We also tested various optical flow estimation approaches on real image sequences recorded by a calibrated camera which was fixed on the arm of a robot. By moving the arm of the robot along a precisely defined trajectory, we can determine the true displacement rate of scene surface elements projected into the image plane and compare it quantitatively with the results of different optical flow estimators.

Since the optical flow estimators are based on gray-value derivatives of up to fourth-order, we were forced to develop modified Gaussian derivative filters to obtain acceptable estimates for the derivatives. Further, we show quantitatively that these filters contribute to a much more robust optical flow estimation. In addition, successive lines of TV-cameras have an offset in time due to the interlace technique. We demonstrate the adaptation of filter kernels for estimating higher-order spatiotemporal derivatives in interlaced image sequences.

R. Mohr, B, Boufama and P. Brand, Understanding positioning from multiple images

It is possible to recover the three-dimensional structure of a scene using only correspondences between images taken with uncalibrated cameras. The reconstruction obtained this way is only defined up to a projective transformation of the 3D space. However, this kind of structure allows some spatial reasoning such as finding a path. In order to perform more specific reasoning, or to perform work with a robot moving in Euclidean space, Euclidean or affine constraints have to be added to the camera observations. Such constraints arise from the knowledge of the scene: location of points, geometrical constraints on lines, etc. First, this paper presents a reconstruction method for the scene, then it discusses how the framework of projective geometry allows symbolic or numerical information about positions to be derived, and how knowledge about the scene can be used for computing symbolic or numerical relationships. Implementation issues and experimental results are discussed.

H. Buxton and S. Gong, Visual surveillance in a dynamic and uncertain world

Advanced visual surveillance systems not only need to track moving objects but also interpret their patterns of behaviour. This means that solving the information integration problem becomes very important. We use conceptual knowledge of both the scene and the visual task to provide constraints. We also control the system using dynamic attention and selective processing. Bayesian belief networks support this and allow us to model dynamic dependencies between parameters involved in visual interpretation. We illustrate these arguments using experimental results from a traffic surveillance application. In particular, we demonstrate that using expectations of object trajectory, size and speed for the particular scene improves robustness and sensitivity in dynamic tracking and segmentation. We also demonstrate behavioral evaluation under attentional control

using a combination of a static BBN TASKNET and dynamic network. The causal structure of these networks provides a framework for the design and integration of advanced vision systems.

I.D. Reid and J.M. Brady, Recognition of object classes from range data

We develop techniques for recognizing instances of 3D object classes (which may consist of multiple and/or repeated sub-parts with internal degrees of freedom, linked by parameterized transformations), from sets of 3D feature observations. Recognition of a class instance is structured as a search of an interpretation tree in which geometric constraints on pairs of sensed features not only prune the tree, but are used to determine upper and lower bounds on the model parameter values of the instance. A real-valued constraint propagation network unifies the representations of the model parameters, model constraints and feature constraints, and provides a simple and effective mechanism for accessing and updating parameter values.

Recognition of objects with multiple internal degrees of freedom, including non-uniform scaling and stretching, articulations, and sub-part repetitions, is demonstrated and analysed for two different types of real range data: 3D edge fragments from a stereo vision system, and position/surface normal data derived from planar patches extracted from a range image.

W.W. Cohen, Pac-learning non-recursive Prolog clauses

G. Schwarz, In search of a "true" logic of knowledge: the nonmonotonic perspective

M. Stefik and S. Smoliar, *The Creative Mind*: *Myths and Mechanisms*: six reviews and a response

K.B. Haase, Too many ideas, just one word: a review of Margaret Boden's *The Creative Mind*: *Myths and Mechanisms*

R. Lustig, Book Review of *The Creative Mind*: *Myths and Mechanisms* (Margaret Boden)

D. Perkins, An unfair review of Margaret Boden's *The Creative Mind* from the perspective of creative systems

A. Ram, L. Wills, E. Domeshek, N. Nersessian and J. Kolodner, Understanding the creative mind: a review of Margaret Boden's *The Creative Mind*

R.C. Schank and D.A. Foster, The engineering of creativity: a review of Boden's *The Creative Mind*

S.R. Turner, Book Review of *The Creative Mind*: *Myths and Mechanisms* (Margaret Boden)

M. Boden, Modelling creativity: a reply to the reviewers

S.W. Smoliar, Book Review of *Creative Cognition* (Ronald Finke, Thomas Ward, Steven Smith)