Visual Observation Under Uncertainty as a Discrete Event Process

Tarek M. Sobh and Ruzena Bajcsy*

General Robotics and Active Sensory Perception Laboratory
University of Pennsylvania
Philadelphia, PA 19104-6228, U.S.A.

Abstract

We address the problem of development and implementation of a discrete event dynamic system (DEDS) observer for a moving agent. We advocate a modeling approach for the visual system and its observer, where the "events" are defined as ranges on parameter subsets. In particular, the proposed system is used for observing a manipulation process, where a robot hand manipulates an object. We recognize the hand/object interaction over time and a stabilizing observer is constructed. The resulting robot arm behavior is constructed as a hybrid intelligent mechanism. The work examines closely the possibilities for errors, mistakes and uncertainties in the manipulation system, observer construction process and event identification mechanisms. Some results from a sequence of a peg-in-hole operation are documented.

1 Introduction

We discuss a new framework and representation for the general problem of observation. The system being studied can be considered as a "hybrid" one, due to the fact that we need to report on *distinct* and *discrete* visual states that occur in the *continuous*, *asynchronous* and three-dimensional world, from two-dimensional observations that are sampled periodically. In other word, the system being observed and reported on consists of a number of continuous, discrete and symbolic parameters that vary over time in a manner that might not be "smooth" enough for the observer, due to visual obscurities and other perceptual uncertainties.

The problem of observing a moving agent was addressed in the literature extensively. It was discussed in the work addressing tracking of targets and, determination of the optic flow [2,7,10,17], recovering 3-D parameters of different kinds of surfaces [6,12,15,16], and also in the context of other problems [1,3,8,9]. However, the need to *recognize*, *understand* and *report* on different visual steps within a dynamic task was not sufficiently addressed. In particular, there is a need for high-level symbolic interpretations of the actions of an agent that attaches meaning to the 3-D world events, as opposed to simple recovery of 3-D parameters and the consequent tracking movements to compensate their variation over time. In this work we establish a framework for the general problem of observation, recognition and understanding of dynamic visual systems, which may be applied to different kinds of visual tasks.

*Professor Sobh is now with the Department of Computer Science at the University of Utah, Salt Lake City, Utah 84112, U.S.A. **Acknowledgments** : This research was supported in part by Air Force AFOSR Grants 88-0244, 88-0296; Army/DAAL Grant 03-89-C-0031PRI; NSF Grants CISE/CDA 88-22719, IRI 89-06770; DARPA Grant N0014-88-0630 and DuPont Corporation.

2 Discrete Event Dynamic Systems

Discrete event dynamic systems (DEDS) are dynamic systems (typically asynchronous) in which state transitions are triggered by the occurrence of discrete events in the system. DEDS are usually modeled by finite state automata with partially observable events together with a mechanism for enabling and disabling a subset of state transitions [11,13,14]. We can represent a DEDS by the following quadruple :

$$G = (X, \Sigma, U, \Gamma)$$

where X is the finite set of states, Σ is the finite set of possible events, U is the set of admissible control inputs consisting of a specified collection of subsets of Σ, corresponding to the choices of sets of controllable events that can be enabled and $\Gamma \subseteq \Sigma$ is the set of observable events.

Stability can be defined with respect to the *states* of a DEDS automaton. Assuming that we have identified the set of "good" states, E, that we would like our DEDS to "stay within" or do not stay outside for an infinite time, then stabilizability can be formally defined as follows :

Given a live system A and some $E \subset X$, $x \in X$ is stabilizable with respect to E (or E-stabilizable) if there exists a state feedback K such that x is alive and E-stable in A_K. A set of states, Q, is a stabilizable set if there exists a feedback law $K(s)$ (a control pattern) so that every $x \in Q$ is alive and stable in A_K, and A is a stabilizable system if X is a stabilizable set.

A DEDS is termed *observable* if we can use the observation sequence to determine the current state exactly at intermittent points in time separated by a *bounded* number of events. The basic idea behind strong output stabilizability is that we will know that the system is in state E iff the observer state is a subset of E. The compensator should then *force* the observer to a state corresponding to a subset of E at intervals of at most a finite integer i observable transitions.

0-8186-2910-X/92 $3.00 © 1992 IEEE

3 Modeling and Observer Construction

Manipulation actions can be modeled efficiently within a discrete event dynamic system framework. We use the DEDS model as a high level structuring technique to preserve and make use of the information we know about the way in which each manipulation task should be performed. The goal will be to make the observer system a strongly output stabilizable one.

We use the image motion to estimate the hand movement. This task can be accomplished by either feature tracking or by computing the full optic flow. The image flow detection technique we use is based on the sum-of-squared-differences optic flow. The sensor acquisition procedure (grabbing images) and uncertainty in image processing mechanisms for determining features are factors that should be taken into consideration when we compute the uncertainty in the optic flow.

One can model an arbitrary 3-D motion in terms of stationary-scene/moving-viewer. The optical flow at the image plane can be related to the 3-D world using a pair of non-linear equations for each point (x, y) in the image plane [12] In this system of equations, the only knowns are the 2-D image flow vectors v_x and v_y, if we use the formulation with uncertainty then basically the 2-D vectors are random variables with a known probability distribution. A number of techniques can be used to linearize the system of equations and to solve for the motion and structure parameters as random variables [4,5,15].

4 Modeling and Recovering 3-D Uncertainties

The uncertainty in the recovered image flow values results from sensor uncertainties and noise and from the image processing techniques used to extract and track features. We use a static camera calibration technique to model the uncertainty in 3-D to 2-D feature locations. The strategy used to find the 2-D uncertainty in the features 2-D representation is to utilize the recovered camera parameters and the 3-D world coordinates (x_w, y_w, z_w) of a known set of points and compute the corresponding pixel coordinates, for points distributed throughout the image plane a number of times, find the actual feature pixel coordinates and construct 2-D histograms for the displacements from the recovered coordinates for the experiments performed. The number of the experiments giving a certain displacement error would be the z axis of this histogram, while the x and y axis are the displacement error. The three dimensional histogram functions are then normalized such that the volume under the histogram is equal to 1 unit volume and the resulting normalized function is used as the distribution of pixel displacement error.

The spatial uncertainty in the image processing technique can be modeled by using synthesized images and corrupting them, then applying the feature extraction mechanism to both images and computing the resulting spatial histogram for the error in finding features. The probability density function for the error in finding the flow vectors can thus be computed as a spatial convolution of the sensor and strategy uncertainties. We then eliminate the unrealistic motion estimates by using the physical (geometric and mechanical) limitations of the manipulating hand. Assuming that feature points lie on a planar surface on the hand, then we can develop bounds on the coefficients of the motion equations, which are second degree functions in x and y in three dimensions, $v_x = f_1(x, y)$ and $v_y = f_2(x, y)$. The 2-D uncertainties are then used to recover the 3-D uncertainties in the motion and structure parameters. The system is linearized by either dividing the parameter space into three subspaces for the translational, rotational and structure parameters and solving iteratively or using other linearization techniques and/or assumptions to solve a linear system of random variables [4,5,6,15,16,18].

5 Conclusions and Results

State transitions are asserted within the DEDS observer model according to the probability value of the occurrence of an event. Events are thus defined as ranges for the different parameters. The problem then reduces to computing the corresponding areas under the refined distribution curves. An obvious way of using those probability values is to establish some threshold values and assert transitions according to those thresholds. It might be the case that none of the obtained probability values exceeds the set threshold value and/or all values are very low. In that case, there is a good chance that we are at either the wrong automata state. The remedy to such problems can be implemented through time proximity, that is, wait for a while (which is to be preset) till a strong probability value is registered and/or *backtrack* in the automaton model for the observer till a high enough probability value is asserted, a fail state is reached or the initial ambiguity is asserted. The backtracking strategy can be implemented using a stack-like structure associated with each state that has already been traversed, which includes a sorted list of the computed event probabilities and a father-state variable.

Experiments were performed to observe the robot hand. The low level visual feature acquisition is performed on the Datacube MaxVideo pipelined video processor at frame rate. The observer and manipulating robots are both PUMA 560's and the Lord experimental gripper is used as the manipulating hand. A peg-in-hole task using the Lord gripper, as seen by the observer, is shown in the figure. The observer tends to get closer to the peg when it approaches the hole, in order to focus on the insertion process. The demonstration sequence screen structure is divided into three sections, the right half is the observer and the agent environments, the upper left corner is the view from the observer camera, while the lower left corner is a graphical representation of the state in the DEDS automaton.

Figure 1 : A Peg-in-Hole Sequence

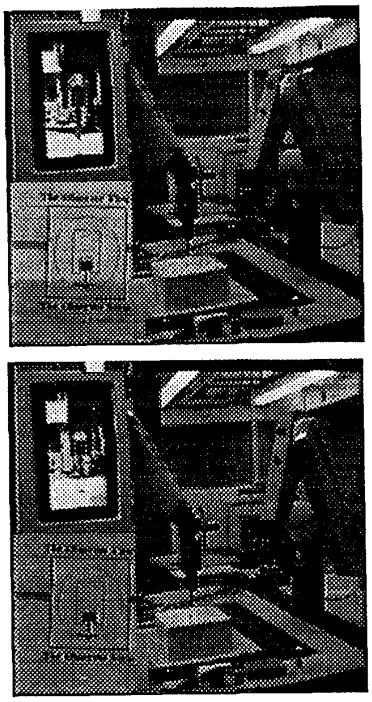

Figure 1 : A Peg-in-Hole Sequence

References

[1] J. Aloimonos and A. Bandyopadhyay, "Active Vision". In *Proceedings of the* 1^{st} *International Conference on Computer Vision*, 1987.

[2] P. Anandan, "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion". In *Proceedings of the* 1^{st} *International Conference on Computer Vision*, 1987.

[3] R. Bajcsy, "Active Perception", *Proceedings of the IEEE*, Vol. 76, No. 8, August 1988.

[4] R. Bajcsy and T. M. Sobh, *A Framework for Observing a Manipulation Process.* Technical Report MS-CIS-90-34 and GRASP Lab. TR 216, University of Pennsylvania, June 1990.

[5] R. Bajcsy and T. M. Sobh, *Observing a Moving Agent.* Technical Report MS-CIS-91-01 and GRASP Lab. TR 247, Computer Science Dept., School of Engineering and Applied Science, University of Pennsylvania, January 1991.

[6] J. L. Barron, A. D. Jepson and J. K. Tsotsos, "The Feasibility of Motion and Structure from Noisy Time-Varying Image Velocity Information", *International Journal of Computer Vision*, December 1990.

[7] P. J. Burt, et al., "Object Tracking with a Moving Camera", *IEEE Workshop on Visual Motion*, March 1989.

[8] F. Chaumette and P. Rives, "Vision-Based-Control for Robotic Tasks", In *Proceedings of the IEEE International Workshop on Intelligent Motion Control*, Vol. 2, pp. 395-400, August 1990.

[9] J. Hervé, P. Cucka and R. Sharma, "Qualitative Visual Control of a Robot Manipulator". In *Proceedings of the DARPA Image Understanding Workshop*, September 1990.

[10] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", *Artificial Intelligence*, vol. 17, 1981, pp. 185-203.

[11] Y. Li and W. M. Wonham, "Controllability and Observability in the State-Feedback Control of Discrete-Event Systems", *Proc.* 27^{th} *Conf. on Decision and Control*, 1988.

[12] H. C. Longuet-Higgins and K. Prazdny, *The interpretation of a moving Retinal Image*, Proc. Royal Society of London B, 208, 385-397.

[13] C. M. Özveren, *Analysis and Control of Discrete Event Dynamic Systems : A State Space Approach*, Ph.D. Thesis, Massachusetts Institute of Technology, August 1989.

[14] P. J. Ramadge and W. M. Wonham, "Modular Feedback Logic for Discrete Event Systems", *SIAM Journal of Control and Optimization*, September 1987.

[15] T. M. Sobh and K. Wohn, "Recovery of 3-D Motion and Structure by Temporal Fusion". In *Proceedings of the* 2^{nd} *SPIE Conference on Sensor Fusion*, November 1989.

[16] M. Subbarao and A. M. Waxman, *On The Uniqueness of Image Flow Solutions for Planar Surfaces in Motion*, CAR–TR–113, Center for Automation Research, University of Maryland, April 1985.

[17] S. Ullman, "Analysis of Visual Motion by Biological and Computer Systems", *IEEE Computer*, August 1981.

[18] S. Ullman, *Maximizing Rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion*, AI Memo 721, MIT AI lab. 1983.

[19] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Teaching by showing : Generating robot programs by visual observation of human performance", 20^{th} ISIR, 1989.

[20] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Design and implementation of a system that generates assembly programs from visual recognition of human action sequences", IROS, 1990.