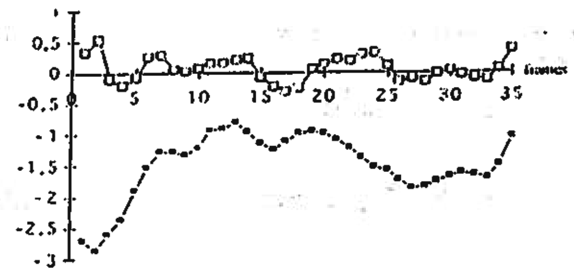

Figure 4: V_x and V_y components of the tip of the pen during the action

Figure 5: Frame-to-impact during the motor act

- a further observation, even if based on rather qualitative grounds, is the fact that a deceleration in the X direction is often synchronous with a relative high value of V_y.

Finally, the plot of time-to-impact, actually measured in terms of *frame-to-impact* (FTI), (see fig. 5) shows that in the initial phase (up to frame 19 in this example) the the value of FTI is rather oscillatory, while in the final phase (during the more crucial phase of the approach) the FTI is asymptotically decreasing in a much smoother way.

5 Discussion and Conclusions

This paper has shown two things:

- Humans perform two-arms manipulation in a way which is compatible with visual measures not requiring explicit 3D computation.
- The computation of such measures is possible from real images using state-of-the art optical flow computation.

Many point are still unexplored and worth studying in more details. Among them, the most relevant are:

- the role of tactile (and more generally proprioceptive) sensory information;
- the role of the stereo system to account for depth (or disparity measures)
- the structure and theoretical aspects of the controller
- the role of kinematics transforming visual measures into commands to the joints of the arms.

[1]

References

[1] D.J. Coombs and C.M. Brown. Intelligent gaze control in binocular vision. In *Proc. of the Fifth IEEE International Symposium on Intelligent Control*, Philadelphia, PA, September 1990.

[2] R. K. Bajcsy. Active perception vs passive perception. In *Proc. Third IEEE Workshop on Computer Vision: Representation and Control*, pages 13–16, Bellaire (MI), 1985.

[3] K. Brunnstrom, T. Lindeberg, and J-O. Eklundh. Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch. In *Proc. ECCV92 - European Conference of Computer Vision*, volume LNCS-588, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[4] Eric Krotkov, K. Henriksen, and R. Kories. Stereo ranging from verging cameras. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12 No.12:1200–1205, 1990.

[5] A. Conkie and P. Chongstitvatana. An Uncalibrated Stereo Visual Servo System. In *BMVC() Proceedings of the British Machine Vision Conference*, pages 277–280. University of Oxford, September 1990.

[6] Russell L. Andersson. *A Robot Ping-Pong Player*. MIT-Press, 1988.

[7] D.H. Ballard, R.C. Nelson, and B. Yamauchi. Animate vision. *Optics News*, 15(5):17–25, 1989.

[8] J. Aloimonos. Purposive and qualitative active vision. In *Proc. 10th Internat. Conf. on Pattern Recognition*, pages 346–360, Atlantic City, U.S.A., June 1990.

[1]This research has been supported by the ESPRIT Project SECOND and by the Special Projects on Robotics of the C.N.R.

THE ROLE OF VISION IN TWO-ARMS MANIPULATION

G. Sandini, F. Gandolfo, G. Metta, and A. Oddera

1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition

New York City
June 15-18, 1993

Sponsored by the
IEEE Computer Society
Technical Committee on Computer Graphics
in cooperation with
ACM/SIGGRAPH

IEEE Computer Society
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1264

Washington, DC • Los Alamitos • Brussels • Tokyo

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC
IEEE

IEEE COMPUTER SOCIETY

The Role of Vision in Two-Arms Manipulation

G. Sandini, F. Gandolfo, G. Metta, A. Oddera

DIST - University of Genoa
Laboratory for Integrated Advanced Robotics (LIRA - Lab)
e-mail:giulio@vision.dist.unige.it

Abstract

The role of vision in two-arms manipulation is investigated to suggest a qualitative approach for the solution of coordinated manipulative actions. The visuo-motor behavior proposed, although partially supported by experimental results, differs from more "geometrical" approaches, based on explicit 4D measures and trajectory estimation, in that it is based on direct visual information, such as stereo disparity or the 2D image motion field. The experimental part presented in the current paper is limited to the analysis of the optical flow information extracted from manipulative actions performed by a human operator. These experiments demonstrate that the qualitative approach proposed is not in contrast with human behavior and that the two-arm manipulation task proposed can be seen as a testbed for qualitative and geometrical approaches.

1 Introduction

In the last years the emphasis of computer vision research has been directed toward the extraction of *metric information* in different forms. As far as the use of vision for the continuos control of motor actions many interesting experiments have been performed in gaze control [1, 2, 3, 4] and to program robot arms interacting with external objects moving independently (see for example [6]).

What we would like to argue in this paper is that the full, explicit knowledge of the 4D trajectory of the object to be handled is not strictly necessary to perform even complex two-arm manipulations.

The approach proposed in this paper is that of using iconic information directly extracted from the images, such as stereo disparity and the 2D image motion field. Based upon the continuous extraction of these measures, purposive motor behaviors can be designed to solve the task [7, 8] without relying on gray-level-based (or frame-difference-based) segmentation of the object to be tracked (it is nor clear how to extend these, apparently simpler, approaches to general environments and to moving cameras).

2 The paradigm

The motor task selected to outline the approach is that of capping a pen. To perform this task the goal of the hand-eye system is to control the motion of one arm such as, at the end of the action, the tip of the pen correctly "docks" with the cap. If the action is performed in closed-loop using vision as a feedback signal, it is necessary to constantly monitor the trajectory of the moving hand (and the tip of the pen) with respect to the cap. A possible solution is to measure the position in a 4D space (x,y,z,t) of the tip of the pen with respect to the cap. The alternative hypothesis proposed in this paper is the use of a direct solution based on the measure of optical flow field and disparity without the need of explicit 4D measures.

Considering the capping action, the goal of the visually-guided controller is to keep the tip of the pen on an ideal linear trajectory connecting, at each instant of time, the cap with the tip itself. The projection of this 3D trajectory on the image plane represents the 2D trajectory that must be followed by the *image* of the moving hand to dock with the cap (see fig. 1). This is the first control variable that may be used: keep the image of the moving hand along this ideal 2D trajectory. In terms of optical flow, this can be achieved by minimizing the component of the flow field perpendicular to this ideal trajectory.

Of course, depth must be also considered. However if the fixation point is either on the cap or the tip of the pen, a successful trajectory lies on the zero-disparity plane: the control action is, in this situation,

1063-6919/93 $03.00 © 1993 IEEE

Figure 1: Ideal 2D trajectory connecting the tip of the pen and the cap. The controller should keep the moving hand along this trajectory

to minimize the absolute value of disparity or, in other words, to drive the moving hand toward (or along) the zero-disparity plane.

Moreover, it is worth stressing that, if the fixation point is on the cap (or on the tip of the pen), the velocity profile can be controlled through the measure of the time-to-crash computed from the component of the optical flow *along* the 2D trajectory of the approaching hand.

3 Experimental set-up

The experiment described here has been performed processing a sequence of 60 images acquired during a capping action performed by a human operator (see fig. 2).

The optical flow was computed from the image sequence and the following measures were extracted from each frame of the sequence:

- Segmentation of the moving hand from the static background;
- Computation of the position of the tip of the pen;
- Computation of the average components of the 2D velocity field along the direction connecting the starting position with the goal point (the cap), and along the direction perpendicular to it.
- Frame-to-crash (i.e. time-to-crash expressed in terms of number of frames)

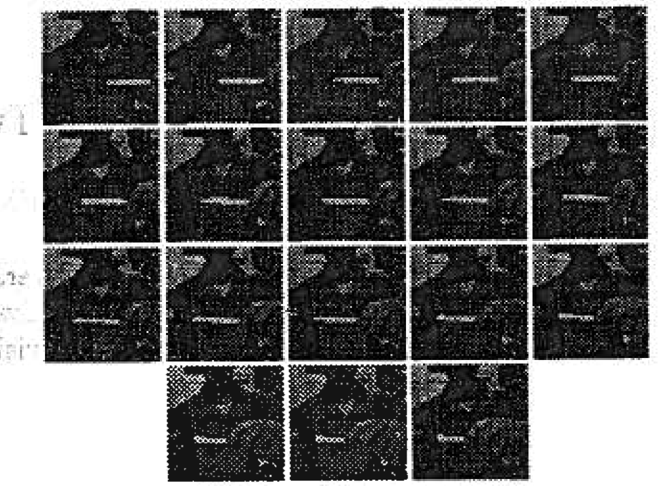

Figure 2: 18 of the 35 frames used in the experiment.

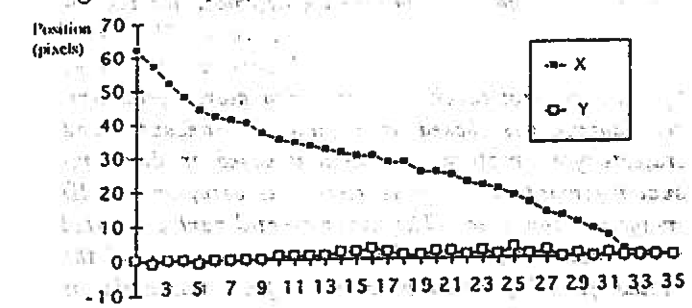

Figure 3: X and Y position of the tip of the pen during the action.

4 Results

The results of the experiment are shown in figures 3, 4, 5. In this particular experiment the capping action was performed in about 1.4 sec. (or 35 frames). Figure 4 represent the x and y components of velocity (V_x and V_y respectively). Few observations are worth mentioning:

- V_y is small compared to V_x, assuming values oscillating about zero. This demonstrates the plausibility of the hypothesis made that the controller is actually minimizing V_y in order to keep the tip of the pen "on tracks".
- V_x is not constant (as well as V_y) showing that constancy of velocity is not a goal of the controller and that the controller is constantly tuning the velocity to keep the tip in the direction of the cap.