A 'Centre of Disparity' based Robotic Gaze and Grasp Reflex Mechanism

Ray Jarvis
Intelligent Robotics Research Centre
Monash University
Wellington Road
Clayton, Victoria, 3800
Australia

0. Abstract

The behaviour based robotics paradigm promoted by Brooks [Brooks, 1990] and others has sensitised the robotics research community to the possibility of building agile and reliable stimulus/response systems unencumbered by planning requirements. Whilst it can be argued that reasoned responses to stimuli are often appropriate, few can deny the value of reactive modes when they can be effectively applied. Discovering appropriate reactive modes often requires ingenuity but this effort can be well rewarded through their delivery of fast and reliable control at low complexity.

In the context of humanoid robotics [Inoue, 1998], there are many implementations of reactive behaviour. One particularly rich domain where there is plentiful evidence of reactive mode behaviour is in hand/eye coordination. The ability (acquired through experimental learning) of very young children to gaze at and then grasp objects within their reach even prior to the recognition of those objects is one such inspiration.

This paper addresses the opportunities for of a simple gaze and grasp reflex using a passive stereopsis vision system mounted on a fast pan/tilt head and a small six degree of freedom anthropomorphic robot. The notion of 'centre of disparity' is invented as the basis of the gaze control mechanism. Such a computation is totally non-specific to particular domains, ie. entirely generalised, a most valuable attribute. The computation is simple, direct, fast and robust. The pan/tilt head is directed to keep the 'centre of disparity' point at the centre of its gaze.

The location of the target object is extracted from the disparity data, enabling the robotic manipulator to approach and grasp it. The success or otherwise of the grasp action can also be monitored by the system and grasp attempts repeated if necessary.

This paper will detail the formulation of the concept of 'centre of disparity' and provide experimental evidence of is successful application to gaze and grasp reflex implementation. Video clips will be presented at the conference showing the behaviour of the implemented system.

1. 'Centre of Disparity' Formulation

In image binary processing [Jarvis and Patrick, 1970] the idea of centre of area of blobs is well established; individual blobs can first be isolated using connectively analysis [Rosenfield and Pfaltz, 1966]. If such a blob is thought of as a cut out a uniformly thick and dense sheet, the centre of area is that point at which it would balance on a pin. With reference to a Cartesian coordinate system and an origin the computation is as follows for a NXN cell laminate where M [i, j] is the cell mass at [i, j], which in this case is either 0 or 1: -

$$\bar{y} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} M[i,j] \times j$$

$$\bar{x} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} M[i,j] \times i$$

Where density variation is permitted, the above calculation determines the centre of mass, again the balancing point.

Only one small additional concept is needed to define the 'centre of disparity'. If, in the above computation, the cell mass, is replaced by the stereo disparity value (assumed non-negative) the results are the 'centre of disparity'.

It is not difficult to visualize a disparity image and to physically interpret the 'centre of disparity' calculation. Simply consider a rectangularly tesselated 'image' in which each cell contains a scalar value indicating the inverse of its range from the observer; the larger the value the closer to the observer. If each cell is replaced by a physical mass proportional to its disparity value, the 'centre of disparity' is the location at which the 'image' would balance on a pin.

If the cell disparity figure is raised to some power, the range selectively of the calculation is sharpened as may be preferred for some applications (including the one described here).

The 'centre of disparity' sort of favours the closet largest object, just what is wanted for directing the gaze of a system to support a grasp reflex. If two large close objects are present the result is ambiguous but seemingly consistent with observed biological system behaviour.

Some shape, colour or size based prior segmentation could be used to resolve this ambiguity but in the simple, one dominant close object case, this is not necessary.

2. Implementation of Grasp Reflex

A Triclops (Point Grey) three camera stereopsis based passive ranging system generates disparity maps up to the camera frame rates depending on the resolution, the correlation mask dimensions and the computational power available. Typically, using a Pentium III 550 mhz processor, disparity maps of 64x64 resolution can be generated at 15 maps/second. Including the 'centre of disparity' (using disparity raised to some power, typically 2 to 6) in the loop slows down the cycle rate a little. In our experiment the 'centre of disparity' parameters were transferred to a Silicon Graphics workstation via ethernet, there to be used to drive a Directed Perception Inc. pan/tilt head carrying the Triclops system to centre itself on the 'centre of disparity'. The response of the systems is quite good (several adjustments/second), but would be improved if the pan/tilt head were controlled directly by the same computer used for the disparity calculations. A small colour camera mounted at the centre of the Triclops three camera head verifies what is being gazed at. The results were quite impressive for object and human hand or head tracking. In our case, the transfer to Silicon Graphics system was one of convenience as this system was already programmed to control the pan/tilt head and Unimate 250 robot which is used for the grasp action.

The average disparity in the vicinity of the 'centre of disparity' provides the third dimension of the location of the target object for grasping purposes; the current gaze vector provides the other two.

3. Robot Approach and Grasp Control

To control the Unimate 250, the Silicon graphics workstation sends streams of Val II instructions along a RS232 serial line to the robot control unit, thus emulating ordinary keyboard terminal exchanges; this process is slow but simple. The resting pose of the Unimate 250 prior to grasping action trajectories approximates the location and pose of a human arm relative to the head, in our case the head being the Triclops vision system [See Figure 1].

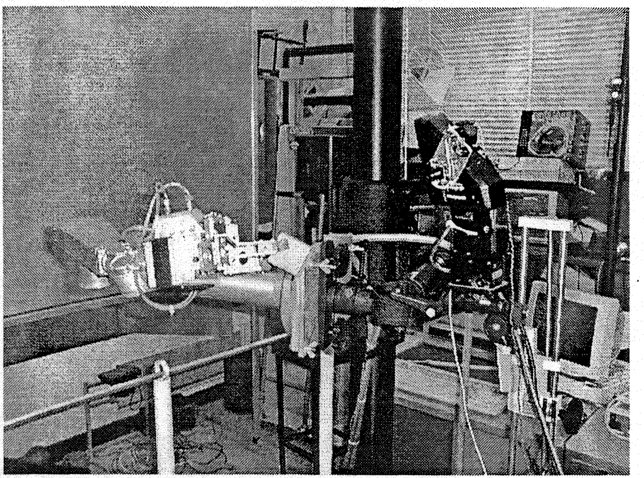

Figure 1. Triclops Vision System

In our simple experiments to demonstrate the 'centre of disparity' based gaze/grasp reflex idea, the Unimate 250 is instructed to make its trajectory and grasp movements only when the gaze and distance parameters have stabilised for an appropriate time period (say 0.5 seconds).

To minimise the serial line traffic in Val II code from the Silicon Graphics workstation to the Unimate 250 and thus speed up the robot's response to instructions, the following two Val II provisions are used.

(a) Specific precomputed position/pose points (in configuration space) can be loaded prior to a gaze and grasp experiment. These locations can be logged earlier by moving the robot by hand with the motor power turned off and indicating when positions/poses are to be recorded. Thus a number of 'canonical' robot positions can be predefined. Moving amongst these points requires a simple move instruction naming the point; both 'all motors flat out at once' or straight line trajectories can be specified. These points can define strategic points from where to begin approach and grasp actions. The location of the target as defined by the gaze operation can be used to choose which start point to adopt - a type of table look up approach would suffice.

(b) Once a suitable move from a relaxed, out of the way home position to a selected start point (and pose) has been completed, simple draw commands which are specified by X, Y, Z increments can complete the grasp trajectory. Closing the fingers should then grasp the target object. A grasp failure can be monitored by a video camera or simply by putting microswitches on the inside of the manipulator's fingers.

Most components of the above strategy are complete but the integrated experiment can not yet be demonstrated. It is anticipated that video clips illustrating the entire gaze and grasp reflex will be able to be shown at the conference.

4. Discussion and Future Developments

Whilst the fast response and generality of the 'centre of disparity' concept is its strength, a number of developments built on top of this simple procedure need considering. Firstly, the question of ambiguity when several close objects draw gaze attention needs resolution. A simple approach would be to apply binary blob connectivity analysis to range disparity) thresholded segments and to delete components of the 'centre of disparity' calculations which relate to less attractive targets (in terms of size, shape average range etc.). This would essentially filter out unwanted contributions without imposing much domain specificity. Including intensity texture and colour as extra filter parameters, whilst providing greater refinement would also narrow the generality of the approach. A way of preserving tracking speed may be to apply extra filtering calculations only after some stability has been established, the gaze focus perhaps no longer shifting too rapidly.

Another consideration is that relating to the robotic hand itself distorting the gaze focus in the very act of reaching out to grasp the target object. A simple remedy is to suspend gaze tracking momentarily to allow the grasp reflex to proceed without disturbing readjustments.

A more interesting consideration arises from the fact that the anthromorphically inspired separation of the 'eye' and the 'hand' permits a different kind of visual servoing to the grasp site than would be afforded by a 'eye-in-the-hand' approach. This latter alternative is enjoyed by animals who use their jaws for grasping. Some studies into these alternative modes may be worth the effort. However, in our case, the humanoid robotics project we have recently embarked on definitely favours the eye and hand separation alternative.

5. Preliminary Conclusions

This paper formulates and will demonstrate by experiment a gaze and grasp reflex mechanism which is robust, direct, reasonably fast and entirely generalised. The definition and computation of the 'centre of disparity' from near frame rate three camera passive stereopsis is central to this study. Some future developments are also discussed.

Bibliography

[Brooks, 1990] R. A. Brooks. Elephants don't Play Chess, *Journal of Robotics and Autonomous Systems*, Vol. 6, 1990, pp. 25-31.

[Inoue, 1998] H. A Inoue. Platform-based Humanoid Robotics Project, IARP, *First International Workshop on Humanoid and Human Friendly Robotics*, Oct. 26 - 27, 1998, Tsukuba Science City, Japan, pp.I-1.

[Jarvis and Patrick, 1970] R.A. Jarvis and E.A. Patrick. Interactive Binary Picture Manipulation, *UMR-Mervin J. Kelly Communications Conference*, Oct. 5 - 7, 1970, Rolla, Missouri, pp. 20-1-1 - 20-1-9.

[Rosenfield and Pfaltz, 1966] A. Rosenfield and J.L.,Pfaltz. Sequential Operations in Digital Picture Processing, *Journal of the Association for Computing Machinery,* Vol. 13, No. 4, Oct. 1966, pp. 471-494.