Robotica (1985) volume 3, pp 7–11



# Real-time cooperative interaction between structured-light and reflectance ranging for robot guidance

Ernest W. Kent, Thomas Wheatley and Marilyn Nashman

*Sensory-Interactive Robotics Group, National Bureau of Standards, Washington D.C. (USA)*

(Received: October 20, 1984)


## SUMMARY

When applied to rapidly moving objects with complex trajectories, the information-rate limitation imposed by video-camera frame rates impairs the effectiveness of structured-light techniques in real-time robot servoing. To improve the performance of such systems, the use of fast infra-red proximity detectors to augment visual guidance in the final phase of target acquisition was explored. It was found that this approach was limited by the necessity of employing a different range/intensity calibration curve for the proximity detectors for every object and for every angle of approach to complex objects. Consideration of the physics of the detector process suggested that a single log-linear parametric family could describe all such calibration curves, and this was confirmed by experiment. From this result, a technique was devised for cooperative interaction between modalities, in which the vision sense provided on-the-fly determination of calibration parameters for the proximity detectors, for every approach to a target, before passing control of the system to the other modality. This technique provided a three hundred percent increase in useful manipulator velocity, and improved performance during the transition of control from one modality to the other.


## INTRODUCTION

Structured-light techniques are commonly used for extracting limited amounts of visual information from a scene very quickly.[1-3] In particular, they are frequently employed in acquiring sparse range data. When this information is combined with ordinary two-dimensional images,[4] visual surfaces in the scene may be understood in terms of the six degrees of freedom of the camera with respect to the object. Thus, a robot with a camera and structured light projectors mounted on the end-effector may be visually servoed with respect to objects in arbitrary positions and orientations. Additionally, the ability to interpret structured-light range images at frame rates renders this technique useful for tracking and acquiring moving objects, as on conveyors and turntables.

The work described here was motivated by the results of a series of experiments designed to explore the limits of structured-light servoing when applied to rapidly moving objects with complex trajectories. Such situations occur, for example, when the robot must acquire objects swinging freely on hooks from overhead conveyors. An experimental demonstration was set up in which a single, simple structured light projection, a sheet of light projected parallel to the plane of the gripper, was used to obtain the range and azimuth of objects moving erratically on a table surface. Range and azimuth were obtained every frame time, and the object's velocity and acceleration were continuously updated by differencing successive frames. The details of the structured-light apparatus employed are presented in the appendix, and in Figure 1.

A robot arm was programmed to use this sensory information to pursue and acquire target objects. The program had no *a priori* information about the object's position, trajectory, velocity, or acceleration. Under these circumstances, a number of shortcomings of pure visual tracking were discovered. The most serious of these is that, as the robot gripper approaches closely to the object, a course change due to an acceleration of given magnitude on the part of the target requires progressively larger corrective changes in the angle of the robot's trajectory. As a result, more rapid response of the control system is required to track the object, and the system is quickly limited by the relatively slow (30 frames per second) servo rate obtainable from the camera. This frequently resulted in the gripper either attempting to overrun the object and colliding with it, or falling short in its grasp when the gripper closed. With a rapidly moving object close to the camera, it was common for a lateral acceleration to remove the object entirely from the field of view before the robot could react to the acceleration with an adequate correction in its own course. A second, less serious, problem was the deterioration of the camera image at very close ranges, due to lack of an automatic focus control.

A variety of measures were undertaken to improve performance, including optimization of weighting factors in the PID control algorithm, and the addition of "intelligent" rule-based strategies. While distinct improvement was obtained, these modifications failed to produce a system that could, for example, seriously challenge a human playing "keep away" with the robot, or catch erratically-moving mechanical toys. Further improvement would be anticipated if the system employed specific *a priori* models of the behavior of the physical systems being tracked (e.g., rolling, swinging, etc.), but this approach was not considered germane to the purpose of the present experiments. It was decided instead to attempt to improve performance by adding additional senses to assist the visual servoing. Ultimately, tactile

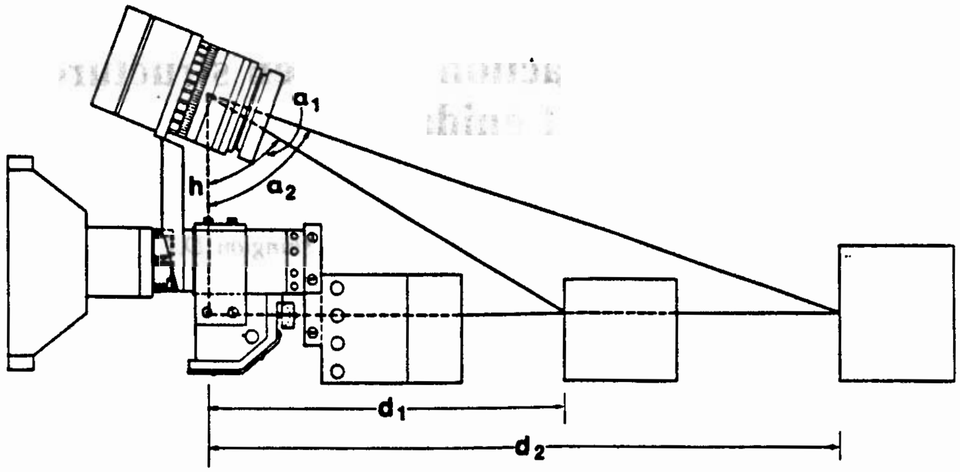


Fig. 1. Structured-light apparatus.

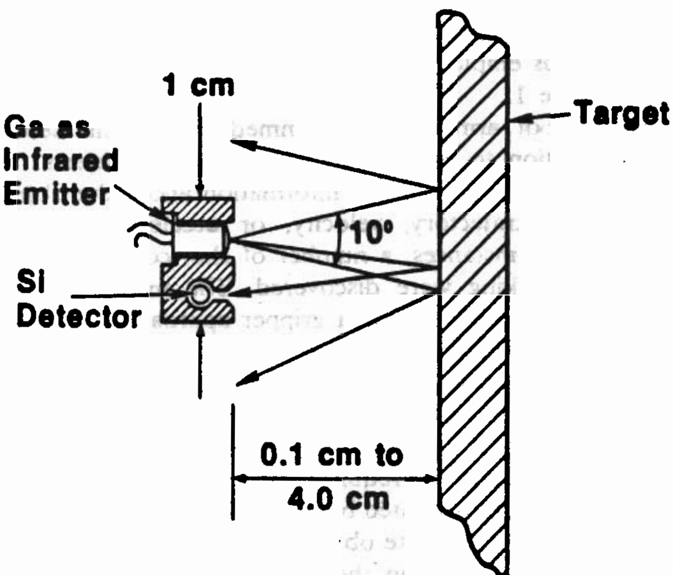


Fig. 2. Proximity sensor.

sensing and infra-red proximity sensing were integrated with structured-light vision. The details of the proximity sensor, and the arrangement of the sensors is detailed in the appendix, and in Figures 2 and 3.

The principal advantage of the infra-red proximity sensor is that it can be used to close the servo loop in one to two milliseconds at close range. By employing multiple sensors with different directions of view, and restricted (but overlapping) receptive fields, discrimination of azimuth as well as range can be obtained. Preliminary experiments indicated that the use of such sensors could markedly improve performance in the final phases of approach to a target, allowing the arm to run at much higher velocities. However, they were useless beyond about 15 centimeters, and gave very poor discrimination of objects. The strengths and weaknesses of the structured-light vision sensor and the infra-red proximity sensor nicely complemented one another, and the camera was used for target acquisition and distant tracking while

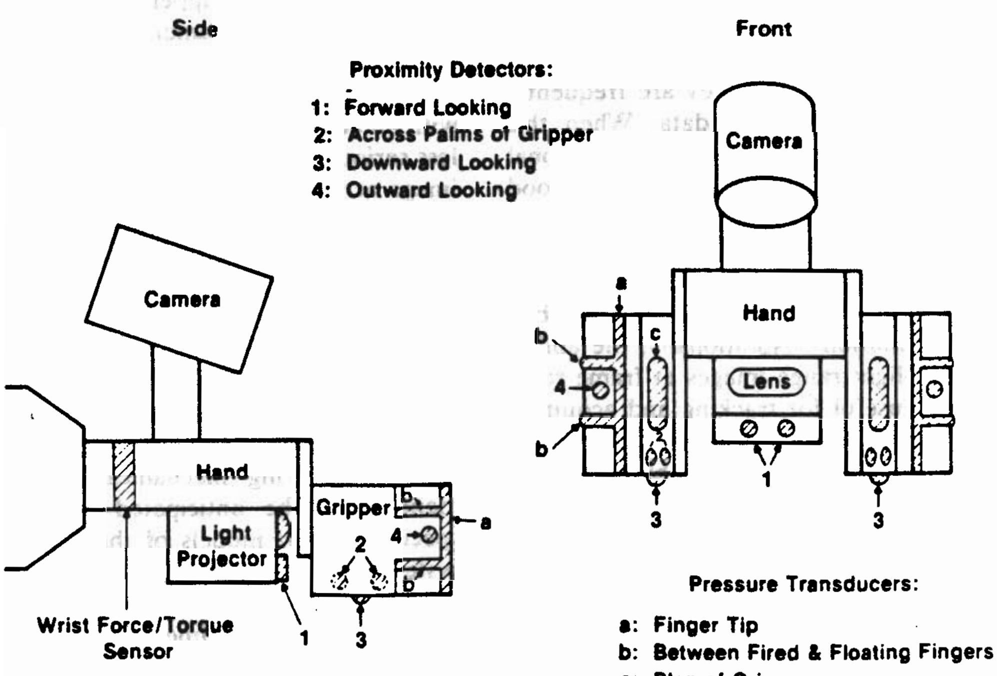


Fig. 3. Arrangement of sensors.

the proximity sensors were employed for rapid servoing of the final grasp.

A significant problem exists with this approach. The camera and structured-light projector, once calibrated, give accurate position measurements under essentially all relevant conditions. In the case of the proximity detector, however, the relation between detected intensity and object range contains a great many variables. These include the color of the object, its reflectivity, its shape, its orientation, and its surface texture to name only a few. A different formula describes the intensity/range relation under each condition of approach to the target. Such formulae do not have to be computed; a calibration table can be experimentally produced so that the distance computation becomes merely a matter of lookup and interpolation. However, the table is good for only one kind of object, approached from one orientation, under one set of ambient light conditions. This may suffice for some sorts of restricted applications, but it is certainly too limited for a general purpose sensory-interactive robot.

A second concern in such a multi-modal system is the problem of smoothly passing objects from one sensory modality to another. If the system is servoing the approach to an object with the vision sense, and at some point switches over to the proximity sensor, any discrepancy between the two would at least result in an apparent jump in position by the object, leading to a false perception of rapid target acceleration. It might even introduce confusion about the orientation or identity of the object. Both of these problems are fundamentally calibration issues. One refers to the absolute calibration of the proximity sensor, and the other to its calibration relative to the camera.

A possible solution to this problem in multi-modal systems is to recalibrate one modality against another for each instance of an approach to an object. This type of cooperative action is particularly appealing when, as in the present case, one of the sensors has robust absolute calibration. We sought a means of doing this, in real-time, during the visually-servoed phase of the approach. We began by examining the response of the infra-red proximity system to a variety of objects.

## METHOD

From the physics of optics, we would expect that the light intercepted by the receiver would be some power function of the distance to the object. In general, letting $D$ be the distance to the object, and $I$ be the intensity read by the detector, we would expect it to be of the form:

$$\ln(D) = B + A \ln(I - R); \tag{1}$$

where $R$ is the "residual" reading at infinite distance attributable to detector noise and background illumination, and $A$ and $B$ are constants collecting all of the parameters peculiar to the particular features of the object, the angle of approach, etc.

The constant $R$ is easily measured, but the constants $A$ and $B$ must be re-determined for every direction of approach to every new object. From equation (1), we would expect that at close ranges, where $R$ is small compared to $I$, $D$ would be a log-linear function of $I$, with slope $A$ and intercept $B$. In general, we would expect that $B$ would vary with the reflectance of the surface, while $A$ would be a complex function of the type of scattering, the dispersion angle of the source of illumination, etc.

To test this hypothesis, we performed measurements on a variety of objects at many ranges and orientations, under several conditions of ambient lighting. All of the expected relationships were confirmed. For surfaces varying from mirrors to matte paper to sandblasted aluminium, the resultant calibration curves were of the form described by equation (1). The observed sensitivities of $A$ and $B$ to experimental manipulations were qualitatively in agreement with our assumptions about the underlying physical model. Figure 4 compares several examples of these results for different types of objects.

Given this, it becomes a simple matter to produce a rapid recalibration of the proximity sensor for every approach to every object. The only requirement is that the useable ranges of the vision and proximity senses overlap. Then, as the arm approaches the object, the distance is measured at two points using the vision sensor, and simultaneously intensity readings are taken from the proximity sensor.

Let $D_1$ and $D_2$ be the two visually measured distances, and $I_1$ and $I_2$ the corresponding intensities obtained from the proximity sensor. We then compute:

$$A = (\ln D_2 - \ln D_1)/(\ln I_1 - \ln I_2) \tag{2}$$

$$B = \ln D_2 - AI_2 \tag{3}$$

(If the residual detector noise at infinity, $R$, is appreciably large compared to the intensity readings, it may be subtracted from $I_1$ and $I_2$ first.)

With the parameters $A$ and $B$ now known for the current object and angle of approach, a particular member of the family of calibration curves described by equation (1) has been defined. The distance $D$ can now be computed directly from $I$ for the remainder of the

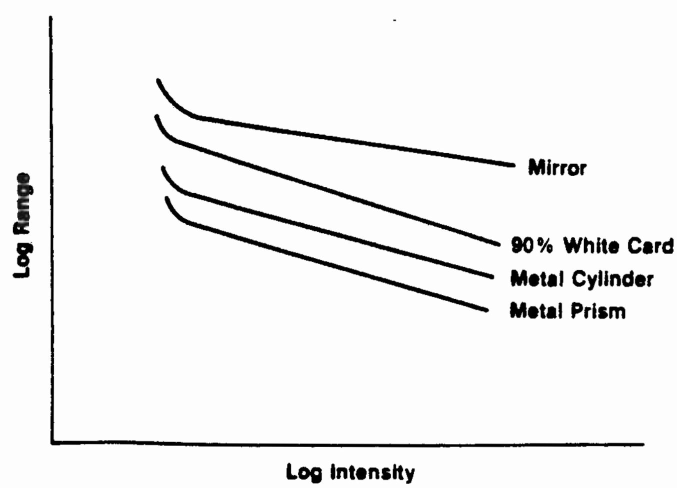


Fig. 4. Results for different types of objects.

approach trajectory by:

$$D = e^{(B + A \ln I)} \tag{4}$$

By applying this procedure to two or more appropriately oriented detectors, the azimuth of the object may be inferred as well. If the target performs a successful evasion, as indicated by an interval of increasing range, or failure to capture after a significant change of direction, the calibration may be recomputed for the new situation.

This procedure provides a rapid means of calibrating the proximity sensor to any particular direction of approach to a complex three-dimensional surface, as well as to all possible conditions of ambient lighting, surface reflectance and color, and variation among individuals of an object class (as for example by dirt or grease). In addition, since the initial $D_1$ and $D_2$ measurements used in the calibration are determined from the structured-light vision sense, the two senses are always calibrated together. This ensures that there will be no apparent discontinuities and jerks as control is passed from one sense to the other. Another significant advantage which is obtained is that the method permits us to define the calibration distance relative to any particular feature on the approaching surface of the object which can be discriminated by the camera, for example, a corner feature, or the nearest surface. This would be impossible for the low-resolution proximity sensor itself, but once calibrated, it continues to provide correct range relative to the desired feature so long as the conditions of approach angle do not change significantly.

## RESULTS

When the on-the-fly recalibration technique was added to the system, performance was significantly enhanced. It was found possible to increase the approach velocity of the arm (and hence the maximum target velocity) by three hundred per cent while maintaining good capture performance. In addition, the repeatability of the positioning of the captured object in the gripper was enhanced.

In practice, we found that the particular system described in the appendix gave optimal results when the calibration parameters were obtained at 15 and at 8 centimeters from the target (as measured from the tips of the gripper fingers), and control was transferred to the proximity sensors at 8 centimeters. Some inaccuracy can result from perspective-induced changes in the surface patch seen by the proximity sensor as the target is approached. However, this was not significant when the sensors used were well-back from the gripper finger tips, at the base of the hand. The fingertip position is, however, a good location from the standpoint of azimuth determination and centering of the approach trajectory. A division of labor is suggested with absolute range determined from sensors in the palm, and centering determined from relative ranges of sensors in the fingertips. Additional improvement can be obtained by employing tactile sensing to abort the grasp and enter a new calibration and acquisition phase when a capture failure does occur.

## APPENDIX

The structured-light apparatus employed in the present study is diagrammed in Figure 1. A flat sheet of illumination is projected from the robot's hand, parallel to the plane of the grippers, by a cylindrical lens. The region of space into which this light projects is viewed by an offset digital video camera, inclined at an angle to the plane of light. When an object is in the path of the plane of light, the light makes a bright stripe across it which is imaged by the camera. Because the camera's position is offset from the plane of light, it will "see" the light stripe lower down in the field of view if the surface illuminated is closer, and higher in the field of view if it is further away. Thus, the vertical position of the light stripe in the frame can be used to obtain the distance to any point on the object which the light illuminates. The range can be computed by simple trigonometry, or obtained directly from a table. The lateral position, or azimuth, of the illuminated object points can also be immediately computed from the lateral position of the point in the field of view, and a knowledge of the focal length of the lens.

Figure 3 shows the hand of the NBS robot, with the arrangement of the various sensors. In addition to the camera and structured light projector, there are infra-red proximity detectors looking in several directions, including into the gripper itself. There are also tactile sensors on the palm and the fingers. The middle finger is fixed, while the outer fingers are floating. Pressure transducers between the fingers allow the detection of pressures which tend to move the floating fingers relative to the fixed ones. In addition, in the wrist, a force/torque sensor can detect forces which oppose the arm's attempt to move the hand. Such forces include the weight of the object and its inertia. They may also include forces such as those generated by improper alignments in mating parts.

The design of a simple proximity sensor based on infra-red reflectively is illustrated in Figure 2. A LED infra-red emitter is mounted to illuminate objects approaching from the direction of interest. A phototransistor is arranged with a lens or mask to receive light from the same direction, but with a somewhat smaller angular coverage than that of the source. An infra-red filter, matched to the emitter frequency, was placed in front of the detector to minimize the ambient light effects. As an object approaches, it reflects light from the source back to the detector. The intensity of the light received will be a function of, among other factors, the distance of the object.

### References

1. G.J. Agin, "Real time control of a robot with a mobile camera", *Technical Note 179*, SRI International (February, 1979).

2. R.C. Bolles and M.A. Fischler. "A ransac-based approach to model fitting and its application to finding cylinders in range data" *Proc. 7th IJCAI* 633–637 (August, 1981).
3. R.N. Nagel, G.J. Vanderburg, J.S. Albus and E. Lowenfeld, "Experiments in part acquisition using robot vision" *Proc. Autofact II, Robots IV Conference, Detroit MI* (October, 1979).
4. J.S. Albus, E.W. Kent, M. Nashman, P. Mansbach, L. Palombo and M. Shneir, "Six-dimensional vision system" *Proc. SPIE* **336**, Robot Vision, 142–153 (1982).