International Journal of Computer Vision, 11:2, 187–203 (1993)

Active Vision for Reliable Ranging: Cooperating Focus, Stereo, and Vergence

ERIC KROTKOV
The Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891

RUZENA BAJCSY
Computer and Information Science, University of Pennsylvania, GRASP Laboratory, 3401 Walnut Street, Philadelphia, PA 19104-6228

Received September 30, 1992; Revised May 8, 1993.

Abstract
This article addresses the problem of measuring reliabily the absolute three-dimensional position of objects in an unknown and cluttered scene. It circumvents the limitations of a single sensor or single algorithm by using several range recovery techniques together, so that they cooperate in visual behaviors similar to those exhibited by the human visual system. Implemented visual behaviors include (i) aperture adjustment to vary depth of field and contrast, (ii) focus ranging followed by fixation, (iii) stereo ranging followed by focus ranging, and (iv) focus ranging followed by disparity prediction followed by focus ranging. The main contribution is a demonstration that two particular visual ranging processes—focusing and stereo—can cooperate to improve measurement reliability. The results of 75 experiments processing close to 3000 different object points lying at distances between 1 and 3 meters demonstrate that the computed range values are highly reliable.

1 Introduction

This article addresses the problem of measuring reliably the three-dimensional position of objects is an unknown and cluttered scene. Although relative distances suffice for many tasks, we seek to estimate absolute position, say, for motion planning or map making.

Computer vision research has established various three-dimensional recovery techniques, but the fundamental problem remains open. One reason is that any single sensor or single algorithm is necessarily limited. We propose to circumvent these limitations by using several range recovery techniques in conjunction, so that they cooperate in visual *behaviors* (actions and reactions in specific *circumstances*) similar to those exhibited by the human visual system. The approach fits within the framework of sensor fusion, but differs from traditional methods by addressing directly the data-acquisition process.

This article presents and analyzes four implemented visual behaviors:

1. Aperture adjustment to vary depth of field and improve contrast (section 4.1, 4.3).
2. Focus ranging followed by fixation (section 4.2).
3. Stereo ranging followed by focus ranging (section 4.3).
4. Focus ranging of one camera followed by prediction of binocular disparity followed by focus ranging of the other cameras (section 4.4).

The main contribution of this work is that it demonstrates that two particular visual ranging processes—focusing and stereo—can cooperate to improve measurement reliability. The advance is not in developing the individual ranging processes, but in enabling their behavioral cooperation. Benefits of cooperation include (i) providing statistically more effective data sets, (ii) enforcing data consistency via mutual constraint, and (iii) reducing mistakes generated by improper

Krotkov, E., Summers, J.F., and Fuma, F. 1988. An agile stereo camera system for flexible image acquisition, *IEEE J. Robot. Autom.* 4(1): 108-113.

Lesser, V. and Corkill, D. 1981. Functionally-accurate, cooperative distributed systems, *IEEE Trans. Syst. Man. Cybern.* 11(1):81-96.

Lesser, V. and Corkill, D. 1983. The distributed vehicle monitoring testbed: A tool for investigating distributed problem solving networks, *AI Magazine* 4(3):15-33.

Miles, F., Judge, S., and Optican, L. 1987. Optically induced changes in the couplings between vergence and accommodation, *J. Neuroscience* 7&8):2576-2589.

Nayar, S. and Nakagawa, Y. 1990. Shape from focus: An effective approach for rough surfaces, *Proc. IEEE Intern. Conf. Robot. Autom.*, pp. 218-225, Cincinnati, May.

Olson, T. and Coombs, D. 1991. Real-time vergence control for binocular robots, *Intern. J. Comput. Vi.* 7(1):67-89.

Pentland, A. 1987. A new sense for depth of field, *IEEE Trans. Patt. Anal. Mach. Intell.* 9(4):523-531.

Schor, C. 1979. The relationship between fusional vergence eye movements and fixation disparity, *V i Research* 19(12):1359-1367.

Shmuel, A. and Werman, M. 1990. Active vision: 3D from an image sequence, *Proc. 10th Intern. Conf. Patt. Recog.*, pp. 48-54, Atlantic City, June.

Smith, R. and Davis R. 1981. Frameworks for cooperation in distributed problem solving, *IEEE Trans. Syst. Man, Cybern.*, 11:61-70.

Sperling, G. 1970. Binocular vision: A physical and a neural theory, *Amer. J. Psych.* 83:461-534.

Subbarao, M. 1988. Parallel depth recovery by changing camera parameters, *Proc. IEEE Intern. Conf. Comput. Vir.*, pp. 149-155, Tarpon springs, FL.

Swain, J. and Stricker, M. eds. 1991. *Promising Directions in Active Vision*, available as University of Chicago Tech. Rep. CS 91-27, November.

Tenenbaum, J. 1970. *Accommodation in Computer Vision.* Ph.D. thesis, Stanford University, November.

Westheimer, G. 1976. Oculomotor control: The vergence system. In R. Monty and J. Senders, eds., *Eye Movements and psychological Processes*, pp. 55-64. Erlbaum: Hillsdale, NJ.

Whaite, P. and Ferrie, F. 1991. From uncertainty to visual exploration, *IEEE Trans. Patt. Anal. Mach. Intell.* 13(10):1038-1049.

Zhang, C. 1992. Cooperation under uncertainty in distributed expert systems, *Artificial Intelligence* 56:21-69.

The human visual system couples accommodation and convergence. One aspect of this coupling is *convergence accommodation:* as the eyes converge, they accommodate as if to focus objects nearer and nearer. Studies of this phenomenon show that convergence alone, in the absence of blur, *can* drive accommodation (Fincham & Walton 1957; Kersten & Stark 1977). The converse aspect of this coupling is *accommodative convergence:* when one eye accommodates to a target, the visual axes converge to fixate that target. Studies of this phenomenon reveal that a subject, when accommodating to a monocularly presented near target, will exhibit convergence (as well as accommodation in the other eye) even though the monocular presentation eliminates the need for convergence (Westheimer 1976). In short, convergence in the absence of disparity *can* drive accommodation, and accommodation in the absence of disparity *can* drive convergence and also accommodation of the other eye.

A third aspect of this coupling is variability: the control parameters vary with the optical stimulus. Miles et al. (1987) studied human subjects before and after wearing various optical devices (periscopes and prisms), and confirmed the existence of adaptive elements that regulate the bias in the vergence and/or accommodation systems.

The general form of cooperation exhibited by the human visual system-one cue triggering the other—inspires our approach. The four implemented visual behaviors have biological analogues:

1. The aperture adjustment behavior is analogous to pupil constriction.
2. The focus-fixate behavior is analogous to accommodative convergence.
3. The stereo-focus behavior is analogous to convergence accommodation, where disparity estimation serves as virtual convergence.
4. The focus-predict-focus behavior is analogous to accommodative convergence, where disparity prediction acts as virtual convergence.

We take the biological examples of cooperation as inspiration, but not more; we do not attempt to implement proposed models—e.g., (Krishnan & Stark 1977; Schor 1979; Sperling 1970)—of the human visual system, nor do we seek to synthesize its mechanisms. We note that our approach exhibits a looser and more sequential coupling between convergence and accommodation than the approach taken by nature for the human visual system.

2.2 Machine Accommodation and Convergence

Computer vision researchers have devoted significant effort to understanding the individual depth cues. For accommodation, pioneering efforts include papers by Horn (1968), Tenenbaum (1970), and Jarvis (1976). Pentland (1987) initiated an effort to recover range from blur that precedes contributions from Grossman (1987). Subbarao (1988), and Ens (1990). Other contributions to the literature come from Krotkov (1987). Nayar and Nakagawa (1990), and Cardillo and Sid-Ahmed (1991). None of these efforts seriously addresses the role of convergence.

For convergence, Grimson (1981) investigated vergence movements as an adjunct stereo process responsible for aligning the images on the retina so as to facilitate coarse-to-fine solutions to the correspondence problem. Geiger and Yuille (1987) employed small vergence changes to disambiguate stereo correspondences. Coombs and Brown (1990) explored roles that vergence and binocular cues play in pre-attentive gaze-stabilization systems. Olson and Coombs (1991) developed a cepstral disparity filter that *estimates* vergence error, and use it to demonstrate real-time vergence control on the Rochester Robot. None of these efforts comprehensively addresses the role of accommodation.

Computer vision researchers have devoted relatively little effort to understanding cooperation of cues. Two voices in the wilderness belong to Abbott and Ahuja (1988), who described an approach to active surface reconstruction that integrates the use of stereo with the control of camera focusing and vergence, thus coupling image data acquisition with surface estimation. Inspired by Sperling's energy-based model (1970), they take an optimization approach to integrating the sensing operations. They seek to minimize an objective function that sums individual criteria to normalize image contrast, minimize image blur, minimize disparity at image centers, maximize surface smoothness, and minimize differences in depth estimates among individual depth cues. They demonstrate the approach by using 36 fixations to build a 3-D model of part of a chair. One limitation of the approach is the assumption that the Scene contains a single continuous surface with no depth discontinuities. Das and Ahuja (1990) addressed this limitation by developing an algorithm for selecting new fixation points during surface reconstruction. Another limitation is the requirement to choose weights for each of the individual criteria.

Table 1. Uncertainty of stereo ranging verified by focusing in 14 trials.

N	T (mm)	Stereo (%/m)	Focus (%/m)	MLE (%/m)
7	2692.4	1.30	0.45	0.53
9	2413.0	1.53	0.45	0.51
7	2159.0	1.47	0.60	0.37
9	1955.8	0.72	0.38	0.34
7	1524.0	0.71	*0.92*	0.84
8	1905.0	0.83	0.69	0.57
12	1397.0	0.31	1.25	1.10
3	1828.0	2.03	0.85	1.00
9	2104.4	1.40	1.37	1.35
14	2540.0	2.26	0.59	0.65
1	1524.0	3.33	2.47	2.58
5	1447.8	1.03	1.04	0.80
2	2286.0	3.58	1.33	1.60
7	1981.2	0.44	0.53	0.47
100	—	1.24	0.79	0.75

Table 2. Uncertainty of focus ranging verified by stereo 10 trials.

N	T (mm)	Stereo (%/m)	Focus (%/m)	MLE (%/m)
10	2908.3	0.83	0.81	0.68
15	2270.0	0.62	0.68	0.58
16	2032.0	0.75	0.72	0.51
14	1778.0	1.06	0.79	0.88
17	1524.0	1.13	0.77	0.81
9	2540.0	1.07	1.41	1.16
17	2159.0	1.07	0.96	*0.99*
17	1905.0	1.58	1.51	1.54
16	1651.0	2.28	2.01	2.14
13	2667.0	0.67	0.46	*0.46*
144	—	1.14	1.02	1.00

maximum-likelihood estimator. The results of the experiments show that (i) the Cooperative ranging procedure is robust, and (ii) that the computed range values are highly reliable, since mistaken combined range measurements are extremely rare, and (iii) they *are* more accurate than either of the computed *ranges* alone, as shown by the smaller rms percent error of the maximum-likelihood estimates. These three points deserve further discussion.

The procedure is not just an idea on paper or a program that ran successfully once, but a process that has been extensively tested in a complex environment and on a wide variety *of* scenes including curved surfaces, occluded objects, and specular reflectors. The implementation autonomously performs a sequence of dynamic, adaptive sensing operations, and exhibits robust behavior in the presence of signal noise and interference, measurement errors, measurement mistakes, and even moderate hardware failures.

While the sturdiness *of* the implementation is noteworthy, the reliability of the combined range measurements is especially significant. In 75 experiments considering 3000 object points, not one of the approximately 100 mistaken range measurements survived crosschecking and statistical consistency testing. Although this does not imply that mistakes cannot occur, it is convincing evidence that they are highly unlikely, and that cooperative range measurements *can* be used with a high degree of confidence.

The accuracy results are less than satisfying, because the relative error in the combined measurements is not significantly lower than one of the measurements (from focusing) alone. This is not surprising given that the focusing measurements are weighted more heavily than the stereo measurements, and leads to the general conclusion that as the differences in sensor accuracy *grow,* the benefits (from the point of view of accuracy alone) of combining their measurements diminishes.

Although the implementation adequately demonstrates the principle of cooperative ranging and practically illustrates the benefit of increased reliability, it is by no means a finished product. The remainder of this section discusses some improvements and extensions that might make it a more powerful system for applications.

The output range maps *are* sparse. In a number of applications, having a few range points with high confidence is of great help. For other applications, it is possible to increase both the quantity and the density of the range points, in at least three ways.

First, focusing and stereo currently operate under different image magnifications (six and one, respectively). In one sense, this is a strength of the implementation, because it shows that sensors with very different operating requirements and characterstics *can* indeed cooperate. In another sense, this is a weakness of the implementation, because it significantly decreases the common field of view, and consequently limits the possible quantity of computed range points. Using the same magnification would simplify the implementation, and could increase the *size* of the range maps by a factor of as much as 36 (the maximum increase in the area *of* the field of view). Alternatively, the cameras could be reoriented and/or repositioned several times for focus ranging and cross-checking.

Second, for Simplicity, the focusing and stereo processes currently consider only the midpoints of the

by the absolute error $|T - Z|$ nor by the relative error $(T - Z)/T$. However, the *distance-dependent relative error* $A = (T - Z)/T^2$ does capture this dependency. For a number of measurements of different quantities (i.e., the distances to different object points), we define the uncertainty U as the root-mean-square percent error over N measurements:

$$U = 100 \sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \Delta_i^2} \quad (1)$$

This figure of merit, whose units are percent per meter, reflects the distancedependent uncertainty for the set of measurements, and can be used to describe the range uncertainty of the measurement process as a whole. One interprets an uncertainty of 1 percent/m as follows: for an object point 1 m away the uncertainty on its range is 1 percent, or 1m x 0.01 = 1 cm; for an object at 2 m distance the relative error is 2 percent, resulting in 2m x 0.02 = 4 cm uncertainty; and so forth.

3.1 Focus *Ranging*

The focus-ranging procedure described in (Krotkov 1987) involves four steps:

1. Set the focal length to its maximum value (105 mm, or a magnification of 6×), to decrease the depth of field of the lens.
2. Select a small image path (typically 20 x 20 pixels) to serve as an evaluation window W.
3. Automatically focus the lens on W. A criterion function approximately measures the "sharpness" of focus by the magnitude of the gradient of the intensity function in W. A search procedure locates the focus motor position M of the lens eliciting the maximal response from the criterion function.
4. Solve an adapted version of the Gaussian lens law for the distance along the z-axis from the lens center to the point(s) projecting to W, using

$$Z_F = \frac{(\gamma M + f)f}{\gamma M} + \delta \quad (2)$$

where γ and δ are calibrated constants, M is the focus motor position determined in step 3, and f is the focal length of the lens.

Experimentally, under typical operating conditions and object distances between 1 and 3 m, the uncertainty of the range computation is approximately $\sigma_f = 1$ percent/m, commensurate with the depth of field of the lens.

Focus ranging encounters problems when W contains projections of objects lying at different distances. Figure 2 illustrates three images, digitized at different focus Settings, of a scene containing three objects lying at different distances. It also plots the criterion function computed over all focus settings while treating the entire field of view as W. The criterion function has three local maxima, one for each object. Using the focus *setting* corresponding to the mode of the criterion function to compute a range produces a meaningless, mistaken result. That the criterion function is not unimodal is considered a *mistake*, as distinct from an error or an inaccurate measurement, because it violates the assumptions that W contains projections of objects lying at roughly the Same distance.

3.2 *Stereo Ranging*

The stereo-ranging procedure (Krotkov et al. 1990) performs five steps:

1. Set the lens focal lengths to their minimum values (17.5 mm, or a magnification of Jx) to maximize the field of view.
2. Acquire a stereo pair of images.
3. Extract line segments from each image.
4. Identify corresponding line segments using a recursive hypothesize-and-verify algorithm. Compute a disparity vector $\vec{d} = (d_x, d_y)$ as the distance (mm) between the midpoints of corresponding line segments.
5. For each correspondence, triangulate approximately on the object, taking the vergence angle into account:

$$Z_s = \frac{b(f\cos\theta - x_L \sin\theta)(f\cos\theta + x_R \sin\theta)}{f(x_L - x_R)\cos 2\theta + (f^2 + x_L x_R)\sin 2\theta} + Z_0 \quad (3)$$

where $x_L - x_R = d_x$, b is the stereo baseline, f is the focal length of the lens, the right camera rotates about the right lens center by θ, the left camera rotates about the left lens center by $-\theta$, and Z_0 is an offset from the baseline to a plane, defined for measurement convenience, to be attached to the camera platform.

For correct solutions to the correspondence problem, the uncertainty of (3) has been experimentally determined to be approximately 2.5 percent/m, for object distances between 1 and 3 m. However, one problem with stereo (not unique to this matching

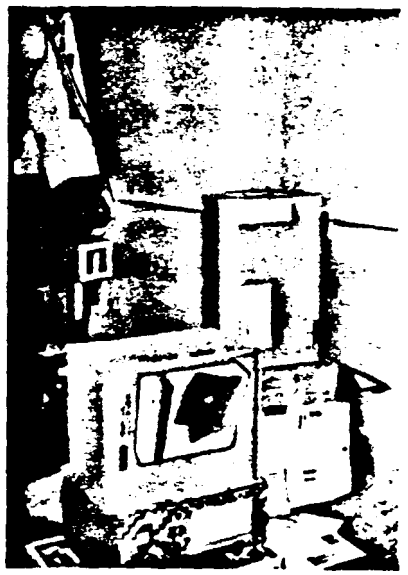

Fig. 6. Typical scene containing a robot arm, gripper, boxes, and envelopes.

We conducted 75 experiments, processing close to 3000 different object points. As a result of the limited field of view of the lens at the maximum magnification, only a small number of line-segment features are extracted (several tens), and as a result of the conservative matching policy, there are even fewer correspondences computed (tens). In general, the range map computed in each experiment is fairly sparse, typically consisting of ten to fifteen points. This sparsity is not necessary, and section 7 discusses strategies for increasing the size and density of the range maps.

6.1 *Reliability*

During the experiments, many measurements were not verified by the cooperative ranging process: some because they were mistaken; others because they were not in the common field of view, occluded, or too close to other points; and still others because of hardware failures. The stereo-ranging procedure computed mistaken ranges for at least 67 points, while the focus-ranging procedure identified mistaken ranges for at least 26 points.

In none of the experiments was a mistaken match ever verified by focusing, nor was a mistaken focusing range ever confirmed by stereo. Indeed, crosschecking was so effective that no more than four points survived to fail the statistical consistency test. We conclude that the range measurements are highly reliable; if a range measurement is confirmed by both stereo and focusing, the hypothesis that it is mistaken *can* be prima facie rejected.

6.2 *Accuracy*

The distancedependent quantity U in (1) defines the uncertainty of the cooperative range measurements. In this case, Z represents the MLE range computed by (10), T represents the manually measured range, and the units of U are percent per meter, as in section 3.

We tabulate accuracy results for a subset of the experimental data in which, to facilitate the manual distance measurements, the studied objects are planar and lie reasonably close to perpendicular to the optic axes of the unconverged lenses. Results for all the data are not available, since careful manual measurement of the object distances is quite time-consuming.

The summary of the data at the bottom of table 1 reveals that, considering an average over 100 points, focus ranging is somewhat more accurate than stereo ranging, consistent with the previous studies of their dative accuracies, and that the MLE is marginally superior to focus ranging alone. The summary of table 2 shows that, considering an average over 144 points, the MLE is slightly more accurate than either of the focus-range measurements alone. Examination of both tables reveals that in a number of experiments, the MLE is actually less accurate than one of the measurements alone. This could be accounted for by the fact that the focus and stereo measurements do not have exactly the same mean values, because they are not perfectly calibrated. Even if they were perfectly calibrated, occasional departures from the expected values would not be surprising, since the predicted variance of the MLE is an expected quantity bounded in the long run, but not guaranteed to fall inside the expected bounds for each data set.

7 Discussion

Here, we have presented a procedure for autonomous cooperative ranging, using focusing and stereo behaviors, which consists of ensuring measurement consistency—by cross-checking and by statistical testing—and combining Consistent measurements by a

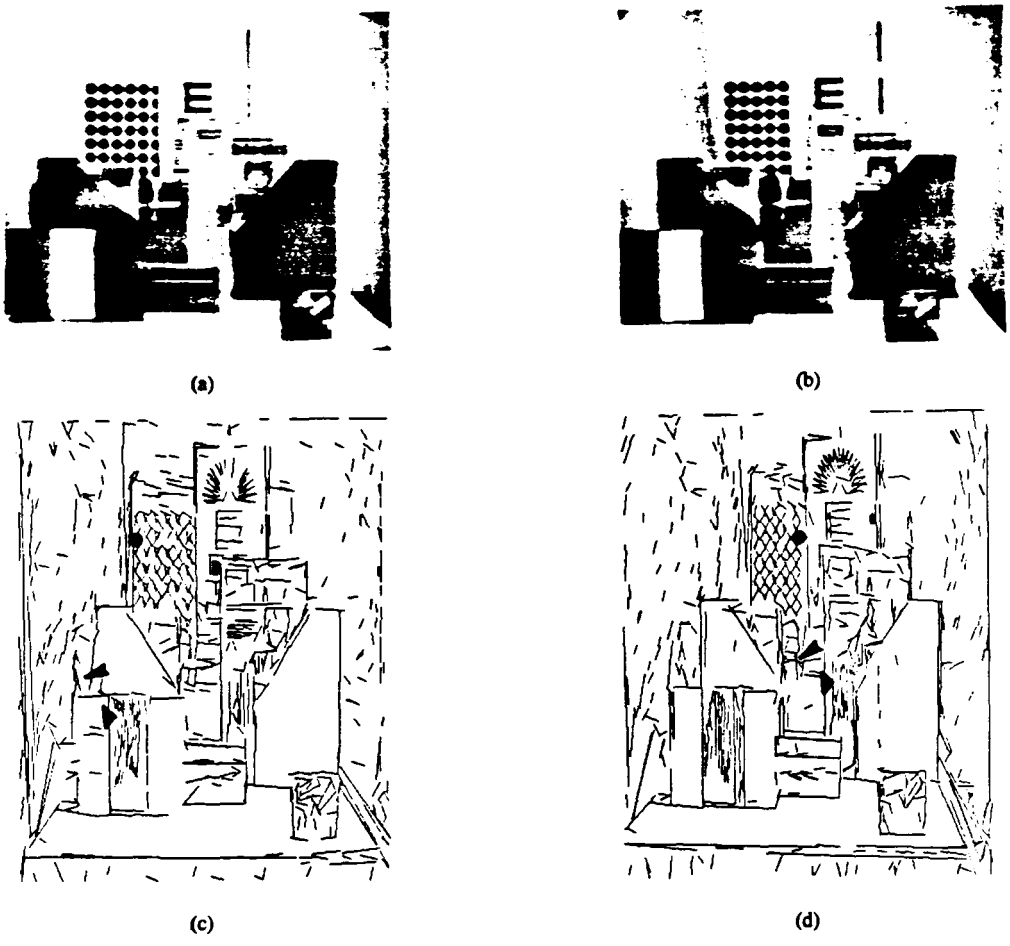

Fig. 3. Stereo mistakes. This figure illustrates the line segments extracted from a stereo pair of images, and identifies three incorrect solutions to the correspondence problem. Using incorrectly matched features to compute range results in a measurement *mistake*.

performed on-line, since everchanging images have to be processed; stereo *can* be performed off-line, since only one pair of images is required. Focusing works best with a *small* depth of field to increase the resolution of the range computation; stereo profits from a large depth *of* field *to* keep as much as possible *of* the scene in *sharp* focus, thus decreasing feature-localization errors. Focusing requires a longer focal length, *so* that the criterion-functionmode has a *sharp* peak; stem *can* be performed at any focal length, but it covers a larger field of view with a smaller focal length. Focusing produces range information *to* patches; stereo produces range information *to* points and lines. Focusing is prone to making mistakes when not operating on meaningful, structured image areas; stereo is prone *to* mistakes in solving the correspondence problem.

Although both make visual-range measurements, focusing and stereo *are* based on different principles, exhibit different accuracies, and involve very different processing steps. The challenge faced now *is* for them *to* cooperate.

4 Cooperative Ranging

This section describes several cooperative behaviors, whose final outcome is two sets of three-dimensional points (figure 4). The sequence of operations at the top *of* the figure (boxes 1–3) autonomously position and orient the cameras *so* that they *can* capture a stereo pair of images and extract line segments from them. The point of the operations in the two branches is for focus and stereo ranging to verify the results of each other

causing minor image instability. The adaptive window refinement compensates for the instability fairly effectively by adjusting the window location. But if this fails and a mistaken range is computed, the failure is likely to cause crosschecking to fail or to be incomplete. The overall performance is poorer since it provides one less data point, but this does not halt the processing of others.

Second, significant changes in Scene illumination occur during the course of a trial, due for instance to shadows cast by passing people, and room lights being turned on or off. Frequently, such distmbrances are accommodated by the adaptive aperture control. If not, a ranging or verification procedure will probably fail, but the system does not grind to a halt. Instead, its performance degrades gracefully.

5 Combination Policy

The outcome of the cooperative ranging process so far (through boxes 10 and 17) is two sets of pairs of range measurements (Z_s, Z_{fm}) and (Z_{fm}, Z_{fs}), whose union can be viewed as one set of pairs of measurements (Z_1, Z_2). It is natural to seek now a single, better estimate that is more accurate than Z_1 or Z_2. For the purposes of this article, better means most likely, and more accurate means lower expected variance. Of course, many other interpretations of these terms can be found in the substantial literature on statistical filtering and estimation techniques. This section describes a policy for combining each pair of measurements into a single one, yielding a set of most likely range values Z, as well as an upper bound on their uncertainty.

5.1 Measurement Consistency and Combination

Let Z_1 and Z_2 represent independent measurements of the range T of an object point. Here they represent the ranges computed independently by focusing and stereo, but the following analysis can be extended to apply to any number of measurements from any kind of measuring device or procedure. For simplicity, the measurements are treated as scalars, although they could equally well be treated as vectors representing, for example, computed three-dimensional positions.

Suppose that the measurements Z_i are normally distributed $N(\mu_i, \sigma_i^2)$, $\sigma_i \neq 0$. Further suppose that the sensors are not biased, so that the μ_i are identical, and in particular, that $\mu_1 = \mu_2$. This amounts to the hypothesis that the measurements are of the same physical quantity, which is justified by the fact that the measurements have so far survived crosschecking. To futher ensure (with a given probability) that the measurements are consistent, a statistical test attempts to reject this hypothesis.

Since Z_1 and Z_2 are independent, a zero-mean, unit-variance random variable χ can be defined by

$$\chi = \frac{Z_1 - Z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{7}$$

The absolute value of χ grows with the difference between the Z_i, and so can be used as a measure of their consistency, testing the hypothesis that the Z_i represent the same value. Define a threshold function by

$$\text{consistent}(Z_1, Z_2) = \begin{cases} 1 & \text{if } |\chi| \leq \chi_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

If $|\chi|$ exceeds the threshold value χ_α then we reject the hypothesis that the given measurements are consistent, that is, represent the same physical quantity. An appropriate χ_α can be chosen basad on an acceptable (see Section 5.2) error probability α considering that χ obeys a standard normal distribution.

Let Z_1 and Z_2 be two independent normally distributed measurements, as above, and in addition require that they be consistent. The maximum likelihood estimate (MLE) Z of T can be shown (Krotkov 1989) to be

$$Z = \frac{\sigma_2^2 Z_1 + \sigma_1^2 Z_2}{\sigma_1^2 + \sigma_2^2} \tag{9}$$

This expression is essentially a weighted average, where the weights are the variances of the measurement processes. The MLE is a lower-variance estimator of T than any of the Z_i (Krotkov 1989).

5.2 Implementation

Figure 5 illustrates the implemented combination policy. The most important implementation issue for both the statistical consistency test and the maximum likelihood estimator is to determine the variances of the measurements Z_i. As discussed in section 3, they are simply the squared uncertainties σ_f^2 and σ_s^2. The uncertainty σ_f on the focusing-range estimate is given by the depth of field, which under typical operating

first be oriented so that objects of interest will be visible from both cameras when the images are later magnified *six* times for focus ranging. To simplify the problem of determining a desirable camera orientation, only vergence movement is considered, excluding other camera translations and rotations. To determine a useful vergence angle, an initial, rough estimate of the range of objects in the scene must first be computed. Here this is *accomplished* by *gross* (low-resolution) focus ranging with one of the cameras, which is arbitrarily assigned to the *muster,* and the other to be the *slave.* This visual behavior is functionally identical to the accommodative convergence behavior of the human visual system, in which accommodation drives convergence.

The gross focusing procedure starts by turning off all the lights, opens the aperture as wide as possible without saturating, adaptively illuminates the lamps until just before saturation, and searches for the focus motor position M bringing objects in the field of view into reasonably sharp focus. It then uses M to compute the range Z_{gross} using (2).

To orient the cameras so that objects at distance Z_{gross} will lie in the common field of view and have cloe to zero disparity, the required vergence angle is

$$\alpha = \tan^{-1} \frac{b}{2Z_{gross}} \tag{4}$$

The orienting procedure calculates the vergence motor position corresponding to a, and servos the vergence motor to this position. After verging, some of the objects *may* have drifted out of view. To reacquire these objects, a corrective pan by amount a is executed. For the purposes of this article, the camera poses are now fixed.

For line extraction and matching, it is desirable to keep as much as possible of the scene in sharp focus. In turn, this requires increasing the depth of field of the lenses, which can be accomplished by decreasing the aperture diameters. For this, the procedure starts by entirely closing each aperture, and turning up the lights as much as possible. Then, independently, each aperture adaptively opens until the image intensities saturate. This implements the behavior of aperture adjustment to vary depth of field.

4.3 Stereo Ranging with Verification by Focusing

The imaging procedure now acquires a pair of stereo images (box 2), and extracts line segments from them (box 3), considering only the portion of the images that would be visible when magnified for focusing. Only on this part of the image can the lenses be focused without reorienting or repositioning the cameras.

Descending the left branch in figure 4, in box 4 the stereo ranging procedure matches the extracted line segments, calculates disparities, and in box 6 uses (3) to compute the range to the midpoint of each matched line segment. As described in section 3, one of the problems with stereo is that the computed solutions to the correspondence problem are occasionally mistaken. It is therefore desirable to verify the computed matches to identify the mistaken ones, which *can* either be eliminated or recomputed. In this work, only the former alternative has been explored, although the latter holds great promise and is discussed further in section 7.

The verfication procedure attempts to confirm a match by focusing on the matched feature, that is, by crosschecking stereo with focusing. This operation is analogous to convergence accommodation in the human visual system, where disparity estimation serves as virtual convergence. It begins in box 7 with the master image coordinates (u_m, v_m) of the midpoint of a matched line segment, and predicts its location in the magnified image using knowledge of the image position of the lens center and the magnification, which are known from calibration procedures. Next, the procedure defines a window P around this predicted location, and zooms in the master lens. Since the predicted location is inexact, the procedure adaptively refmes the predicted locations, using the edge content of the predicted window. Specifically, it transforms P into a refined window Wof the same size, whose center lies at the centroid $\bar{C}$ of the Sobel gradient magnitude distribution S in P, where

$$S(i, j) = \|\nabla I(i, j)\| \quad , \tag{5}$$

$$C_x = \frac{\Sigma_P iS(i, j)}{\Sigma_P S(i, j)},$$

$$C_y = \frac{\Sigma_P jS(i, j)}{\Sigma_P S(i, j)}, \ (i, j) \in P \tag{6}$$

This effectively pulls the predicted window toward the midpoint of the line segment, on which W is now centered.

In box 8, the verfication p d u r e predicts the focus-motor position M corresponding to a computed range Z_s by solving (2) for M. Next, the procedure establishes an interval $[M_1, M_2]$ of focus-motor positions symmetric about M, whose size is chosen in

first be oriented so that objects of interest will be visible from both cameras when the images are later magnified six times for focus ranging. To simplify the problem of determining a desirable camera orientation, only vergence movement is considered, excluding other camera translations and rotations. To determine a useful vergence angle, an initial, rough estimate of the range of objects in the scene must first be computed. Here this is accomplished by gross (low-resolution) focus ranging with one of the cameras, which is arbitrarily assigned to the *master*, and the other to be the *slave*. This visual behavior is functionally identical to the accommodative convergence behavior of the human visual system, in which accommodation drives convergence.

The gross focusing procedure starts by turning off all the lights, opens the aperture as wide as possible without saturating, adaptively illuminates the lamps until just before saturation, and searches for the focus motor position M bringing objects in the field of view into reasonably sharp focus. It then uses M to compute the range Z_{gross} using (2).

To orient the cameras so that objects at distance Z_{gross} will lie in the common field of view and have cloe to zero disparity, the required vergence angle is

$$\alpha = \tan^{-1} \frac{b}{2Z_{gross}} \tag{4}$$

The orienting procedure calculates the vergence motor position corresponding to α, and servos the vergence motor to this position. After verging, some of the objects *may* have drifted out of view. To reacquire these objects, a corrective pan by amount $\alpha/2$ is executed. For the purposes of this article, the camera poses are now fixed.

For line extraction and matching, it is desirable to keep as much as possible of the scene in sharp focus. In turn, this requires increasing the depth of field of the lenses, which can be accomplished by decreasing the aperture diameters. For this, the procedure starts by entirely closing each aperture, and turning up the lights as much as possible. Then, independently, each aperture adaptively opens until the image intensities saturate. This implements the behavior of aperture adjustment to vary depth of field.

4.3 Stereo Ranging with Verification by Focusing

The imaging procedure now acquires a pair of stereo images (box 2), and extracts line segments from them (box 3), considering only the portion of the images that would be visible when magnified for focusing. Only on this part of the image can the lenses be focused without reorienting or repositioning the cameras.

Descending the left branch in figure 4, in box 4 the stereo ranging procedure matches the extracted line segments, calculates disparities, and in box 6 uses (3) to compute the range to the midpoint of each matched line segment. As described in section 3, one of the problems with stereo is that the computed solutions to the correspondence problem are occassionally mistaken. It is therefore desirable to verify the computed matches to identify the mistaken ones, which can either be eliminated or recomputed. In this work, only the former alternative has been explored, although the latter holds great promise and is discussed further in section 7.

The verfication procedure attempts to confirm a match by focusing on the matched feature, that is, by crosschecking stereo with focusing. This operation is analogous to convergence accommodation in the human visual system, where disparity estimation serves as virtual convergence. It begins in box 7 with the master image coordinates (u_m, v_m) of the midpoint of a matched line segment, and predicts its location in the magnified image using knowledge of the image position of the lens center and the magnification, which are known from calibration procedures. Next, the procedure defines a window P around this predicted location, and zooms in the master lens. Since the predicted location is inexact, the procedure adaptively refmes the predicted locations, using the edge content of the predicted window. Specifically, it transforms P into a refined window W of the same size, whose center lies at the centroid $\vec{C}$ of the Sobel gradient magnitude distribution S in P, where

$$S(i, j) = \|\nabla I(i, j)\| \tag{5}$$

$$C_x = \frac{\Sigma_P iS(i, j)}{\Sigma_P S(i, j)}, \quad C_y = \frac{\Sigma_P jS(i, j)}{\Sigma_P S(i, j)}, \ (i, j) \in P \tag{6}$$

This effectively pulls the predicted window toward the midpoint of the line segment, on which W is now centered.

In box 8, the verfication procedure predicts the focus-motor position M corresponding to a computed range Z_s by solving (2) for M. Next, the procedure establishes an interval $[M_1, M_2]$ of focus-motor positions symmetric about M, whose size is chosen in

causing minor image instability. The adaptive window refinement compensates for the instability fairly effectively by adjusting the window location. But if this fails and a mistaken range is computed, the failure is likely to cause cross-checking to fail or to be incomplete. The overall performance is poorer since it provides one less data point, but this does not halt the processing of others.

Second, significant changes in Scene illumination occur during the course of a trial, due for instance to shadows cast by passing people, and room lights being turned on or off. Frequently, such distrubrances are accommodated by the adaptive aperture control. If not, a ranging or verification procedure will probably fail, but the system does not grind to a halt. Instead, its performance degrades gracefully.

5 Combination Policy

The outcome of the cooperative ranging process so far (through boxes 10 and 17) is two sets of pairs of range measurements (Z_s, Z_{fm}) and (Z_{fm}, Z_{fs}), whose union can be viewed as one set of pairs of measurements (Z_1, Z_2). It is natural to seek now a single, better estimate that is more accurate than Z_1 or Z_2. For the purposes of this article, better means most likely, and more accurate means lower expected variance. Of course, many other interpretations of these terms can be found in the substantial literature on statistical filtering and estimation techniques. This section describes a policy for combining each pair of measurements into a single one, yielding a set of most likely range values Z, as well as an upper bound on their uncertainty.

5.1 Measurement Consistency and Combination

Let Z_1 and Z_2 represent independent measurements of the range Tof an object point. Here they represent the ranges computed independently by focusing and stereo, but the following analysis can be extended to apply to any number of measurements from any kind of measuring device or procedure. For simplicity, the measurements are treated as scalars, although they could equally well be treated as vectors representing, for example, computed three-dimensional positions.

Suppose that the measurements Z_i are normally distributed $N(\mu_i, \sigma_i^2)$, $\sigma_i \neq 0$. Further suppose that the sensors are not biased, so that the μ_i are identical, and in particular, that $\mu_1 = \mu_2$. This amounts to the hypothesis that the measurements are of the same physical quantity, which is justified by the fact that the measurements have so far survived crosschecking. To futher ensure (with a given probability) that the measurements are consistent, a statistical test attempts to reject this hypothesis.

Since Z_1 and Z_2 are independent, a zero-mean, unit-variance random variable χ can be defined by

$$\chi = \frac{Z_1 - Z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{7}$$

The absolute value of χ grows with the difference between the Z_i, and so can be used as a measure of their consistency, testing the hypothesis that the Z_i represent the same value. Define a threshold function by

$$\text{consistent}(Z_1, Z_2) = \begin{cases} 1 & \text{if } |\chi| \leq \chi_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

If $|\chi|$ exceeds the threshold value χ_α then we reject the hypothesis that the given measurements are consistent, that is, represent the same physical quantity. An appropriate χ_α can be chosen basad on an acceptable (see Section 5.2) error probability a considering that χ obeys a standard normal distribution.

Let Z_1 and Z_2 be two independent normally distributed measurements, as above, and in addition require that they be consistent. The maximum likelihood estimate (MLE) Z of T can be shown (Krotkov 1989) to be

$$Z = \frac{\sigma_2^2 Z_1 + \sigma_1^2 Z_2}{\sigma_1^2 + \sigma_2^2} \tag{9}$$

This expression is essentially a weighted average, where the weights are the variances of the measurement processes. The MLE is a lower-variance estimator of Tthan any of the Z_i (Krotkov 1989).

5.2 Implementation

Figure 5 illustrates the implemented combination policy. The most important implementation issue for both the statistical consistency test and the maximum likelihood estimator is to determine the variances of the measurements Z_i. As discussed in section 3, they are simply the squared uncertainties σ_f^2 and σ_i^2. The uncertainty σ_f on the focusing-range estimate is given by the depth of field, which under typical operating

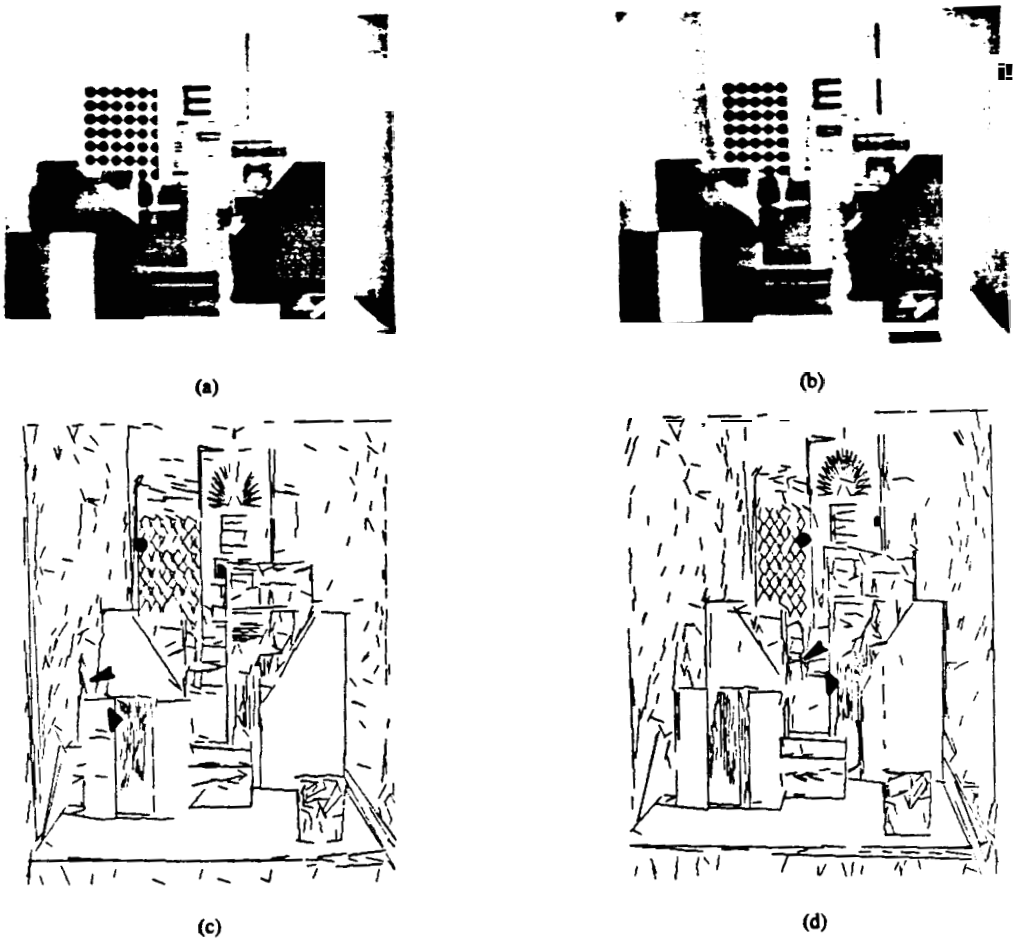

Fig. 3. Stereo mistakes. This figure illustrates the line segments extracted from a stereo pair of images, and identifies three incorrect solu

performed on-line, since everchanging images have to be processed; stereo *can* be performed off-line, since only one pair of images is required. Focusing works best with a small depth of field to increase the resolution of the range computation; stem profits from a large depth *of* field to keep as much as possible *of* the scene in *sharp* focus, thus decreasing feature-localization errors. Focusing requires a longer focal length, *so that* the criterion-function mode has a *sharp* peak; stereo *can* be performed at any focal length, but it covers a larger field of view with a smaller focal length. Focusing produces range information to patches; stem produces range information to points and lines. Focusing is prone to making mistakes when not operating on meaningful, structured image areas; stereo is prone to mistakes in solving the correspondence problem.

Although both make visual-range measurements, focusing and stereo are based on different principles, exhibit different accuracies, and involve very different processing steps. The challenge faced now is for them to cooperate.

4 Cooperative *Ranging*

This section describes several cooperative behaviors, whose final outcome is two sets of three-dimensional points (figure 4). The sequence *of* operations at the top *of* the figure (boxes 1–3) autonomously position and orient the cameras *so* that they *can* capture a stereo pair of images and extract line segments from them. The point of the operations in the two branches is for focus and stem ranging to verify the results of each other

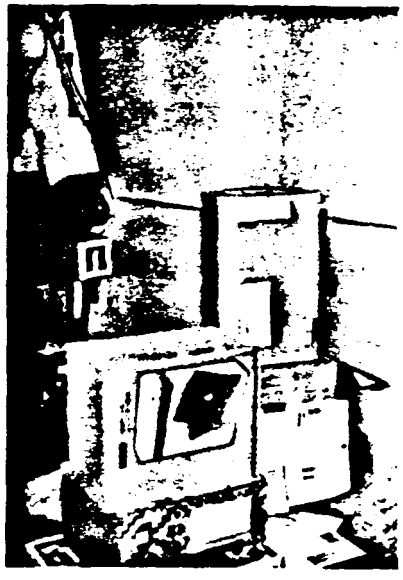

Fig. 6. Typical scene containing a robot arm, gripper, boxes, and envelopes.

We conducted 75 experiments, processing close to 3000 different object points. As a result of the limited field of view of the lens at the maximum magnification, only a small number of line-segment features are extracted (several tens), and as a result of the conservative matching policy, there are even fewer correspondences computed (tens). In general, the range map computed in each experiment is fairly sparse, typically consisting of ten to fifteen points. This sparsity is not necessary, and section 7 discusses strategies for increasing the size and density of the range maps.

6.1 *Reliability*

During the experiments, many measurements were not verified by the cooperative ranging process: some because they were mistaken; others because they were not in the common field of view, occluded, or too close to other points; and still others because of hardware failures. The stereo-ranging procedure computed mistaken ranges for at least 67 points, while the focus-ranging procedure identified mistaken ranges for at least 26 points.

In none of the experiments was a mistaken match ever verified by focusing, nor was a mistaken focusing range ever confirmed by stereo. Indeed, cross-checking was so effective that no more than four points survived to fail the statistical consistency test. We conclude that the range measurements are highly reliable; if a range measurement is confirmed by both stereo and focusing, the hypothesis that it is mistaken can be prima facie rejected.

6.2 *Accuracy*

The distance-dependent quantity U in (1) defines the uncertainty of the cooperative range measurements. In this case, Z represents the MLE range computed by (10), T represents the manually measured range, and the units of U are percent per meter, as in section 3.

We tabulate accuracy results for a subset of the experimental data in which, to facilitate the manual distance measurements, the studied objects are planar and lie reasonably close to perpendicular to the optic axes of the unconverged lenses. Results for all the data are not available, since careful manual measurement of the object distances is quite time-consuming.

The summary of the data at the bottom of table 1 reveals that, considering an average over 100 points, focus ranging is somewhat more accurate than stereo ranging, consistent with the previous studies of their relative accuracies, and that the MLE is marginally superior to focus ranging alone. The summary of table 2 shows that, considering an average over 144 points, the MLE is slightly more accurate than either of the focus-range measurements alone. Examination of both tables reveals that in a number of experiments, the MLE is actually less accurate than one of the measurements alone. This could be accounted for by the fact that the focus and stereo measurements do not have exactly the same mean values, because they are not perfectly calibrated. Even if they were perfectly calibrated, occasional departures from the expected values would not be surprising, since the predicted variance of the MLE is an expected quantity bounded in the long run, but not guaranteed to fall inside the expected bounds for each data set.

7 Discussion

Here, we have presented a procedure for autonomous cooperative ranging, using focusing and stereo behaviors, which consists of ensuring measurement consistency—by crosschecking and by statistical testing—and combining consistent measurements by a

by the absolute error $|T - Z|$ nor by the relative error $(T - Z)/T$. However, the *distance-dependent relative error* $A = (T - Z)/T^2$ does capture this dependency. For a number of measurements of different quantities (i.e., the distances to different object points), we define the uncertainty U as the root-mean-square percent error over N measurements:

$$U = 100 \sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \Delta_i^2} \quad (1)$$

This figure of merit, whose units are percent per meter, reflects the distancedependent uncertainty for the set of measurements, and can be used to describe the range uncertainty of the measurement process as a whole. One interprets an uncertainty of 1 percent/m as follows: for an object point 1 m away the uncertainty on its range is 1 percent, or $1\text{m} \times 0.01 = 1\ \text{cm}$; for an object at 2 m distance the relative error is 2 percent, resulting in $2\text{m} \times 0.02 = 4$ cm uncertainty; and so forth.

3.1 Focus Ranging

The focus-ranging procedure described in (*Krotkov* 1987) involves four steps:

1. Set the focal length to its maximum value (105 mm, or a magnification of 6×), to decrease the depth of field of the lens.
2. Select a small image path (typically 20×20 pixels) to serve as an evaluation window W.
3. Automatically focus the lens on W. A criterion function approximately measures the "sharpness" of focus by the magnitude of the gradient of the intensity function in W. A search procedure locates the focus motor position M of the lens eliciting the maximal response from the criterion function.
4. Solve an adapted version of the Gaussian lens law for the distance along the z-axis from the lens center to the point(s) projecting to W, using

$$Z_F = \frac{(\gamma M + f)f}{\gamma M} + \delta \quad (2)$$

where γ and δ are calibrated constants, M is the focus motor position determined in step 3, and f is the focal length of the lens.

Experimentally, under typical operating conditions and object distances between 1 and 3 m, the uncertainty of the range computation is approximately $\sigma_f = 1$ percent/m, commensurate with the depth of field of the lens.

Focus ranging encounters problems when W contains projections of objects lying at different distances. Figure 2 illustrates three images, digitized at different focus settings, of a Scene containing three objects lying at different distances. It also plots the criterion function computed over all focus settings while treating the entire field of view as W. The criterion function has three local maxima, one for each object. Using the focus setting corresponding to the mode of the criterion function to compute a range produces a meaningless, mistaken result. That the criterion function is not unimodal is considered a *mistake,* as distinct from an error or an inaccurate measurement, because it violates the assumptions that Wcontains projections of objects lying at roughly the same distance.

3.2 Stereo Ranging

The stereo-ranging procedure (Krotkov et al. 1990) performs five steps:

1. Set the lens focal lengths to their minimum values (17.5 mm, or a magnification of 1×) to maximize the field of view.
2. Acquire a stereo pair of images.
3. Extract line segments from each image.
4. Identify corresponding line segments using a recursive hypothesize-and-verify algorithm. Compute a disparity vector $\vec{d} = (d_x, d_y)$ as the distance (mm) between the midpoints of corresponding line segments.
5. For each correspondence, triangulate approximately on the object, taking the vergence angle into account:

$$Z_s = \frac{b(f\cos\theta - x_L \sin\theta)(f\cos\theta + x_R \sin\theta)}{f(x_L - x_R)\cos 2\theta + (f^2 + x_L x_R)\sin 2\theta)} + Z_0 \quad (3)$$

where $x_L - x_R = d_x$, b is the stereo baseline, f is the focal length of the lens, the right camera rotates about the right lens center by θ, the left camera rotates about the left lens center by $-\theta$, and Z_0 is an offset from the baseline to a plane, defined for measurement convenience, to be attached to the camera platform.

For correct solutions to the correspondence problem, the uncertainty of (3) has been-experimentally determined to be approximately 2.5 percent/m, for object distances between 1 and 3 m. However, one problem with stereo (not unique to this matching

Table 1. Uncertainty of stereo ranging verified by focusing in 14 trials.

N	*T* (mm)	Stereo (%/m)	Focus (%/m)	MLE (%/m)
7	2692.4	1.30	0.45	0.53
9	2413.0	1.53	0.45	0.51
7	2159.0	1.47	0.60	0.37
9	1955.8	0.72	0.38	0.34
7	1524.0	0.71	0.92	0.84
8	1905.0	0.83	0.69	0.57
12	1397.0	0.31	1.25	1.10
3	1828.0	2.03	0.85	1.00
9	2184.4	1.40	1.37	1.35
14	2540.0	2.26	0.59	0.65
1	1524.0	3.33	2.47	2.58
5	1447.8	1.03	1.04	0.80
2	2286.0	3.58	1.33	1.60
7	1981.2	0.44	0.53	0.47
100	—	1.24	0.79	0.75

Table 2. Uncertainty of focus ranging verified by stereo 10 trials.

N	*T* (mm)	Stereo (%/m)	Focus (%/m)	MLE (%/m)
10	2908.3	0.83	0.81	0.68
15	2270.0	0.62	0.68	0.58
16	2032.0	0.75	0.72	0.51
14	1778.0	1.06	0.79	0.88
17	1524.0	1.13	0.77	0.81
9	2540.0	1.07	1.41	1.16
17	2159.0	1.07	0.96	0.99
17	1905.0	1.58	1.51	1.54
16	1651.0	2.28	2.01	2.14
13	2667.0	0.67	0.46	0.46
144	—	1.14	1.02	1.00

maximum-likelihood estimator. The results of the experiments show that (i) the cooperative ranging procedure is robust, and (ii) that the computed range values are highly reliable, since mistaken combined range measurements are extremely rare, and (iii) they are more accurate than either of the computed *ranges* alone, as shown by the smaller rms percent error of the maximum-likelihood estimates. These three points deserve further discussion.

The procedure is not just an idea on paper or a program that ran successfully once, but a process that has been extensively tested in a complex environment and on a wide variety of scenes including curved surfaces, occluded objects, and *specular* reflectors. The implementation autonomously performs a sequence of dynamic, adaptive sensing operations, and exhibits robust behavior in the presence of signal noise and interference, measurement errors, measurement mistakes, and even moderate hardware failures.

While the sturdiness of the implementation is noteworthy, the reliability of the combined range measurements is especially significant. In 75 experiments considering 3000 object points, not one of the approximately 100 mistaken range measurements survived cross-checking and statistical consistency testing. Although this does not imply that mistakes cannot occur, it is convincing evidence that they are highly unlikely, and that cooperative range measurements *can* be used with a high degree of confidence.

The accuracy results are less than satisfying, because the relative error in the combined measurements is not significantly lower than one of the measurements (from focusing) alone. This is not surprising given that the focusing measurements are weighted more heavily than the stereo measurements, and leads to the general conclusion that as the differences in sensor accuracy *grow,* the benefits (from the point of view of accuracy alone) of combining their measurements diminishes.

Although the implementation adequately demonstrates the principle of cooperative ranging and practically illustrates the benefit of increased reliability, it is by no means a finished product. The remainder of this section discusses some improvements and extensions that might make it a more powerful system for applications.

The output range maps are *sparse.* In a number of applications, having a few range points with high confidence is of great help. For other applications, it is possible to increase both the quantity and the density of the range points, in at least three ways.

First, focusing and stereo currently operate under different image magnifications (six and one, respectively). In one sense, this is a strength of the implementation, because it shows that sensors with very different operating requirements and characteristics *can* indeed cooperate. In another sense, this is a weakness of the implementation, because it significantly decreases the common field of view, and consequently limits the possible quantity of computed range points. Using the same magnification would simplify the implementation, and could increase the size of the range maps by a factor of as much as 36 (the maximum increase in the area of the field of view). Alternatively, the cameras could be reoriented and/or repositioned several times for focus ranging and cross-checking.

Second, for simplicity, the focusing and stereo processes currently consider only the midpoints of the

The human visual system couples accommodation and convergence. One aspect of this coupling is *convergence accommodation*: as the eyes converge, they accommodate as if to focus objects nearer and nearer. Studies of this phenomenon show that convergence alone, in the absence of blur, *can* drive accommodation (Fincham & Walton 1957; Kersten & Stark 1977). The converse aspect of this coupling is *accommodative convergence:* when one eye accommodates to a target, the visual axes converge to fixate that target. Studies of this phenomenon reveal that a subject, when accommodating to a monocularly presented near target, will exhibit convergence (as well as accommodation in the other eye) even though the monocular presentation eliminates the need for convergence (Westheimer 1976). In short, Convergence in the absence of disparity *can* drive accommodation, and accommodation in the absence of disparity can drive convergence and also accommodation of the other eye.

A third aspect of this coupling is variability: the control parameters vary with the optical stimulus. Miles et al. (1987) studied human subjects before and after wearing various optical devices (periscopes and prisms), and confirmed the existence of adaptive elements that regulate the bias in the vergence and/or accommodation systems.

The general form of cooperation exhibited by the human visual system-one cue triggering the other—inspires our approach. The four implemented visual behaviors have biological analogues:

1. The aperture adjustment behavior is analogous to pupil constriction.
2. The focus-fixate behavior is analogous to accommodative convergence.
3. The stereo-focus behavior is analogous to convergence accommodation, where disparity estimation serves as virtual convergence.
4. The focus-predict-focus behavior is analogous to accommodative convergence, where disparity prediction acts as virtual convergence.

We take the biological examples of cooperation as inspiration, but not more; we do not attempt to implement proposed models—e.g., (Krishnan & Stark 1977; Schor 1979; Sperling 1970)—of the human visual system, nor do we seek to synthesize its mechanisms. We note that our approach exhibits a looser and more sequential coupling between convergence and accommodation than the approach taken by nature for the human visual system.

2.2 Machine Accommodation and Convergence

Computer vision researchers have devoted significant effort to understanding the individual depth cues. For accommodation, pioneering efforts include papers by Horn (1968), Tenenbaum (1970), and Jarvis (1976). Pentland (1987) initiated an effort to recover *range* from blur that precedes contributions from Grossman (1987), Subbarao (1988), and Ens (1990). Other contributions to the literature come from Krotkov (1987). Nayar and Nakagawa (1990), and Cardillo and Sid-Ahmed (1991). None of these efforts seriously addresses the role of convergence.

For convergence, Grimson (1981) investigated vergence movements as an adjunct stereo process responsible for aligning the images on the retina so as to facilitate coarse-to-fine solutions to the correspondence problem. Geiger and Yuille (1987) employed small vergence changes to disambiguate stereo correspondences. Coombs and Brown (1990) explored roles that vergence and binocular cues play in pre-attentive gaze-stabilization systems. Olson and Coombs (1991) developed a cepstral disparity filter that estimates vergence error, and use it to demonstrate real-time vergence control on the Rochester Robot. None of these efforts comprehensively addresses the role of accommodation.

Computer vision researchers have devoted relatively little effort to understanding cooperation of cues. Two voices in the wilderness belong to Abbott and Ahuja (1988), who described an approach to active surface reconstruction that integrates the use of stereo with the control of camera focusing and vergence, thus coupling image data acquisition with surface estimation. Inspired by Sperling's energy-based model (1970), they take an optimization approach to integrating the sensing operations. They seek to minimize an objective function that sums individual criteria to normalize image contrast, minimize image blur, minimize disparity at image centers, maximize *surface* smoothness, and minimize differences in depth estimates among individual depth *cues.* They demonstrate the approach by using 36 fixations to build a 3-D model of part of a chair. One limitation of the approach is the assumption that the Scene contains a single continuous surface with no depth discontinuities. Dias and Ahuja (1990) addressed this limitation by developing an algorithm for selecting new fixation points during surface reconstruction. Another limitation is the requirement to choose weights for each of the individual criteria.

Krotkov, E., Summers, J.F., and Fuma, E 1988. An agile stereo camera *system* for flexible image acquisition. *IEEE J. Robot. Autom.* 4(1):108-113.

Lesser, V. and Corkill. D. 1981. Functionally-accurate, cooperative distributed systems, *IEEE Trans.* Sysr. *Man, Cybern.* 11(1):81-96.

Lesser, V. and Corkill. D. 1983. The distributed vehicle monitoring testbed: A tool for investigating distributed problem solving networks, *AI Magazine* 4(3):15-33.

Miles, F.. Judge. S., and Optican, L. 1987. Optically induced changes in the couplings between vergence and accommodation, *J. Neuroscience* 7&8):2576-2589.

Nayar, S. and Nakagawa, Y. 1990. Shape from focus: An effective approach for rough surfaces, *Proc. IEEE Intern. Conf. Robot. Autom.*, pp. 218-225, Cincinnati, *May.*

Olson, T. and Coombs. D. 1991. *Real-time* vergence control for binocular robots, *Intern. J. Comput. Vu.* 7(1):67-89.

Pentland, A. 1987. A new sense for depth of field, *IEEE Trans. Patt. Anal. Mach. Intell.* 9(4):523-531.

Schor, C. 1979. The relationship between fusional vergence eye movements and fixation disparity, V i *Research* 19(12):1359-1367.

Shmuel, A. and Werman, M. 1990. Active vision: 3D from an *image* sequence. *Proc. 10th Intern. Conf. Ran. Recog.*, pp. 48-54, Atlantic City, June.

Smith. R. and Davis R. 1981. Frameworks for cooperation in distributed problem solving, *IEEE Trans. Syst. Man. Cybern.*, 11:61-70.

Sperling, G. 1970. Binocular vision: A physical and a *neural* theory, *Amer. J. Psych.* 83:461-534.

Subbarao, M. 1988. Parallel depth recovery by changing camera parameters, *Proc. IEEE Intern. Conf. Comput. Vis.*. pp. 149-155, Tarpon Springs, FL.

Swain, J. and Stricker, M. eds. 1991. *Promising Directions* in *Active Vision*, available as University of Chicago Tech. Rep. CS 91-27, November.

Tenenbaum, J. 1970. *Accommodation in Computer Vision*. Ph.D. thesis. Stanford University. November.

Westheimer, G. 1976. Oculomotor control: The vergence system. In R. Monty and J. Senders, eds., *Eye Movements and psychological Processes*, pp. 55-64, Erlbaum: Hillsdale, NJ.

Whaite, P. and Ferrie, F. 1991. From uncertainty to visual exploration, *IEEE Trans. Patt. Anal. Mach. Intell.* 13(10):1038-1049.

Zhang, C. 1992. Cooperation under uncertainty in distributed expert systems, *Artificial Intelligence* 56:21-69.

International Journal of Computer Vision, 11:2, 187–203 (1993)

Active Vision for Reliable Ranging: Cooperating Focus, Stereo, and Vergence

ERIC KROTKOV
The Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891

RUZENA BAJCSY
Computer and Information Science, University of Pennsylvania, GRASP Laboratory, 3401 Walnut Street, Philadelphia, PA 19104-6228

Received September 30, 1992; Revised May 8, 1993.

Abstract

This article addresses the problem of measuring reliabily the absolute threedimensional position of objects in an unknown and cluttered scene. It circumvents the limitations of a single sensor or single algorithm by using several range recovery techniques together, so that they cooperate in visual behaviors similar to those exhibited by the human visual system. Implemented visual behaviors include (i) aperture adjustment to vary depth of field and contrast, (ii) focus ranging followed by fixation, (iii) stereo ranging followed by focus ranging, and (iv) focus ranging followed by disparity prediction followed by focus ranging. The main contribution is a demonstration that two particular visual ranging processes—focusing and stereo—can cooperate to improve measurement reliability. The results of 75 experiments processing close to 3000 different object points lying at distances between 1 and 3 meters demonstrate that the computed range values are highly reliable.

1 Introduction

This article addresses the problem of measuring reliably the three-dimensional position of objects is an unknown and cluttered scene. Although relative distances suffice for many tasks, we seek to estimate absolute position, say, for motion planning or map making.

Computer vision research has established various three-dimensional recovery techniques, but the fundamental problem remains open. One reason is that any single sensor or single algorithm is necessarily limited. We propose to circumvent these limitations by using several range recovery techniques in conjunction, so that they cooperate in visual *behaviors* (actions and reactions in specific circumstances) similar to those exhibited by the human visual system. The approach fits within the framework of sensor fusion, but differs from traditional methods by addressing directly the data-acquisition process.

This article presents and analyzes four implemented visual behaviors:

1. Aperture adjustment to vary depth of field and improve contrast (section 4.1, 4.3).
2. Focus ranging followed by fixation (section 4.2).
3. Stereo ranging followed by focus ranging (section 4.3).
4. Focus ranging of one camera followed by prediction of binocular disparity followed by focus ranging of the other cameras (section 4.4).

The main contribution of this work is that it demonstrates that two particular visual ranging processes—focusing and stereo—can cooperate to improve measurement reliability. The advance is not in developing the individual ranging processes, but in enabling their behavioral cooperation. Benefits of cooperation include (i) providing statistically more effective data sets, (ii) enforcing data consistency via mutual constraint, and (iii) reducing mistakes generated by improper