Adaptive Control of Cooperating Sensors: Focus and Stereo Ranging with an Agile Camera System

Eric Krotkov †
GRASP Laboratory
Computer and Information Science Dept.
University of Pennsylvania
Philadelphia, PA 19104-6389

Ralf Kories
Fraunhofer-Institut fur Informations- und Datenverarbeitung
Sebastian-Kneipp-Str. 12-14
D-7500 Karlsruhe 1
Federal Republic of Germany

ABSTRACT

This paper presents a cooperative computer vision procedure in which focus ranging and stereo ranging operate together, verifying the results of each other in computing the position (but not the shape) of arbitrary objects in a stationary, unknown environment. The procedure increases the reliability of the position measurements by enforcing measurement consistency via mutual constraint, and increases their accuracy by combining them with a maximum likelihood estimator into an estimate of lower variance than any of the measurements alone. The final outcome of the procedure is a set of estimated three-dimensional points together with their estimated uncertainties built from a sequence of dynamic, adaptive sensing operations. The results of 75 experiments processing close to 3000 different object points lying between 1 and 3 meters distant from the cameras show that the integrated range values are (*i*) highly reliable, since no mistaken combined range measurements are observed, and (*ii*) more precise than either of the computed ranges alone. The cooperative methodology extends to more and different sensors, and the results lend practical credence to the view that multiple sensors (*i*) allow the limitations of a single sensor to be circumvented, (*ii*) provide larger, statistically more effective data sets, (*iii*) reduce mistakes generated by inaccurate interpretation models, using sensors outside of their known operating regions, and sensor failure.

1. Introduction

Three-dimensional computer vision measurements are in principle imprecise, and frequently are unreliable. One reason for this is that any single sensor is necessarily limited: since it is a real device, it is limited in bandwidth and frequency response; since its readings are interpreted by models which are approximations, it is not infinitely accurate; since it may at times produce spurious readings, it is not infinitely reliable; and since it is subject to the real world, it may from time to time fail altogether. This paper proposes to circumvent the limitations of a single sensor by using multiple visual sensors to make three-dimensional position measurements.

Multiple sensors can provide larger, statistically more effective data sets, and can allow data consistency to be enforced via mutual constraint. Further, multiple sensors can reduce mistakes generated by improper interpretation models, using sensors outside of their known operating regions, and sensor failure. But using multiple sensors opens strategic and tactical problems of control, and raises the question of how the sensors can operate together, i.e., cooperate.

A number of researchers have used different sensors together, including touch and stereovision [2], centroid and tactile [13], acoustic rangefinders and vision [5], range and reflectance [1], thermal and visual [11], ultrasound and infrared [4], laser rangefinders and odometers [3], color vision and laser range-finders [12], and stereovision, odometers and contact [6]. It is noteworthy that work in sensor cooperation has been limited to different sensor modalities, frequently according to the formula "Vision and something else." In particular, we are aware of no work in which multiple passive visual sensors operate simultaneously and conjunctively, a deficiency which we hope the research reported in this paper will begin to remedy.

While other work concentrates on multi-modal sensing, this paper explores *intra*-modal sensing by examining how two particular visual ranging processes can cooperate to improve measurement reliability and accuracy. It starts by reviewing two procedures — focus ranging and stereo ranging — for computing the three-dimensional position (but not the shape) of objects in a completely unknown environment. In section 3 it presents a cooperative ranging procedure in which focusing and stereo operate together, cross-checking each others' measurements. Next it describes a statistical consistency test, and develops a method for combining the measurements and estimating the most likely object position. In section 5 it presents the results of 75 experiments processing close to 3000 object points, analyzing the reliability and accuracy of the cooperative procedure. It concludes by critically discussing the approach and the results in detail.

2. Assumptions and Background

This section states some basic assumptions and sketches the research domain for studying cooperative sensing. To provide background information, it then reviews the principles of focus and stereo ranging, and summarizes the experimentally observed performance of their implementations. It concludes by comparing focusing and stereo as techniques for measuring range.

The scene is assumed to be stationary and populated by multiple structured (not featureless) objects lying at unknown locations indoors. The sensing apparatus is an agile stereo camera system [10]. Four servomotors position and orient the stereo cameras inside a supporting gantry, allowing horizontal and vertical translations, plus pan and tilt rotations. The cameras lie on a platform which, actuated by a fifth servomotor, allows the cameras to converge and diverge. Six other servomotors adjust the focal length of the two lenses (zooming in and out), open and close the aperture, and change the focusing distance. In addition, the intensities of up to ten incandescent lamps providing scene lighting are independently controlled by a host processor, which drives all of the devices.

The lenses are modeled as pin-holes which obey the Gaussian thick lens law. A left-handed coordinate frame is assigned to each camera, whose origin is at the optical center of the lens, with the x-axis parallel to the photoreceptor scan-lines, and the z-axis directed toward the scene, coincident with the optic axis.

The basic problem considered here is to measure the three-dimensional position of objects in the scene, which can be reduced to the problem of computing their z-components (their *ranges*) using the pin-hole projection equations (see section 4.3, *ff*). Two techniques, focus ranging and stereo ranging, will be used to make

† Currently with the Robotics and Artificial Intelligence Group, LAAS du CNRS, 7 avenue du Colonel Roche, 31077 Toulouse Cedex, France.

these measurements. Before discussing the techniques, it is useful to first define a measurement error model.

2.1. Error Model

In general, it is more difficult with these techniques to accurately measure the distance to objects which lie farther away. Let Z^* represent the true distance along the z-axis from the lens center to an object point, and Z a distance computed by focus or stereo ranging. Earlier studies of these techniques suggest that the difference $|Z^* - Z|$ grows quadratically with Z^*. The quadratic dependence of accuracy on distance is reflected neither by the absolute error $|Z^* - Z|$ nor by the relative error $|Z^* - Z| / Z^*$. However, the *distance-dependent relative error* $\Delta = Z^* - Z / Z^{*2}$ does capture this dependency. Assuming that the measurements are not biased (i.e., the sensors are properly calibrated), the first moment of Δ is $\mu = E[\Delta] = 0$, and the variance is

$$E[(\Delta - \mu)^2] = \frac{1}{N^2} \sum_{i=1}^{N} \Delta_i^2 ,$$

where N is the number of measurements of different object distances. Define the uncertainty U based on the variance of Δ by the root-mean-square (rms) percent error:

$$U = \frac{100}{N} \sqrt{\sum_{i=1}^{N} \Delta_i^2} , \qquad (1)$$

whose units are percent per meter. An uncertainty of 1 percent/m is to be interpreted as follows. For an object point 1 m away the uncertainty on its range is 1 percent, or 1 cm. For an object at 2 m distance the relative error is 2 percent, resulting in 4 cm uncertainty. This figure of merit U reflects the distance-dependent uncertainty for N measurements of different quantities (the distances to different object points), and can be used to describe the accuracy of the measurement process as a whole.

2.2. Focus Ranging

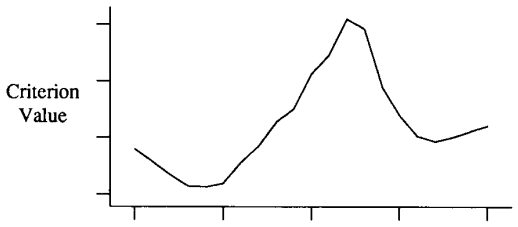
The focus ranging procedure [8,9] involves four steps: (1) Set the focal length f to its maximum value (105 mm, or a magnification of 6x), to decrease the depth of field of the lens. (2) Select a small image patch (typically 20x20 pixels) to serve as an evaluation window W. (3) Automatically focus the 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

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

where γ and δ are calibrated constants, and M is the focus motor position determined in step (3).

Theoretically, the resolution of the range computation is limited by the depth of field D of the lens. Experimentally, under typical operating conditions and object distances between 1 and 3 m, the uncertainty of the range computation has been determined to be approximately σ_f=1 percent/m, commensurate with the depth of field of the lens. But focus ranging encounters problems when W contains projections of objects lying at different distances. As an example, figure 1 plots the criterion function computed over all focus settings for a scene containing three objects lying at different distances while treating the entire field of view as W. The criterion function has three modes, 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.

Figure 1. Example of Focus Ranging Mistake.



This figure plots the focus criterion function computed over all focus settings evaluated on a scene containing three objects lying at different distances from the camera while treating the entire field of view as the evaluation window W.

2.3. Stereo Ranging

The stereo ranging procedure [7] performs five steps: (1) Set the lens focal lengths to their minimum values (17.5 mm, or a magnification of 1x) 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 $\underline{d} = (d_x, d_y)$ as the distance (in millimeters) between the midpoints of corresponding line segments. (5) For each correspondence, triangulate approximately on the object, taking the convergence angle into account, by:

$$Z_S = \frac{b f}{d_x + k_1 V^2 + k_2 V + k_3} + k_5 , \qquad (3)$$

where d_x is the x-component of the disparity vector (in millimeters), b is the stereo baseline, the k_i are calibrated constants, and V is the vergence motor position.

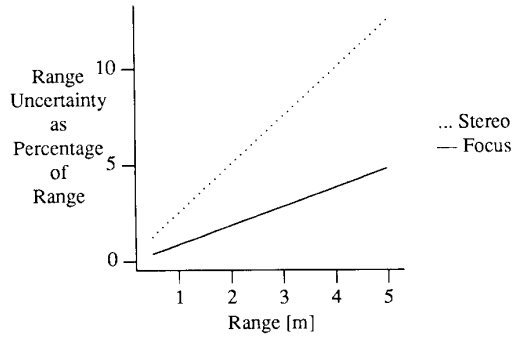
Equation (3) is an approximation, and for correct solutions to the correspondence problem, its uncertainty has been experimentally determined to be approximately σ_s=2.5 percent/m, for object distances between 1 and 3 m. However, one of the problems with stereo (not unique to this matching algorithm) is that the solutions to the correspondence problem are occasionally mistaken. It is possible for two line segments which are not projections of the same object feature to satisfy all the hypothesis generation and verification constraints and be incorrectly matched. This can be the case when a scene structure for some reason does not appear as a feature in both of the images, or when similar scene structures (e.g. a periodic pattern) produce similar image features which are difficult to disambiguate. Using incorrectly matched features results in a measurement *mistake*, as distinct from a measurement error, i.e., an inaccurate measurement.

2.4. Comparison of Focus and Stereo Ranging

Figure 2 illustrates the modeled uncertainty on the two range computations. Note that these estimates of uncertainty are valid only for this particular implementation of the focus and stereo ranging processes, which uses a focal length six times longer for focusing (105 mm) than for stereo (17.5 mm). In principle, one would expect the stereo uncertainty to decrease by a factor of six using the longer focal length, although this has not been experimentally verified.

Focusing is monocular; stereo is binocular. 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 function has a sharp peak; stereo

Figure 2. Uncertainty on the Range Measurements.



This figure plots the modeled uncertainties σ_f and σ_s.

can be performed at any focal length, but it covers a larger field of view with a smaller focal length. 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.

3. A Cooperative Ranging Procedure

This section describes a cooperative ranging procedure, whose final outcome is two sets of computed three-dimensional points computed by focus and stereo ranging. Figure 3 illustrates an overview of the procedure. The sequence of seven operations at the top of the figure (described in detail elsewhere [9]) 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 by cross-checking in order to increase the reliability of the range measurements each produces alone. Specifically, the purpose of the left branch is to perform stereo matching and to identify mistaken matches with focusing, and the purpose of the right branch is to perform focus ranging and to identify mistaken focus points using stereo. This section describes these two procedures. Section 4 discusses the operations at the bottom of the branches.

3.1. Stereo Ranging with Verification by Focusing

Descending the left branch in figure 3, the stereo ranging procedure matches the extracted line segments, calculates disparities, and uses equation (3) to compute a range Z_S to the midpoint of each matched line segment. As described in section 2.3, 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 6.

The verification procedure attempts to confirm a match by focusing on the matched feature, i.e., by cross-checking stereo with focusing. It begins 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. Next, the procedure defines a window W_{pred} around this predicted location, and zooms in the master lens. Since the predicted location is inexact, the procedure adaptively refines the predicted locations, using the edge content of the predicted window. Specifically, it transforms W_{pred} into a refined window W_{ref} of the same size, whose center lies at the centroid of the Sobel gradient magnitude distribution in W_{pred}. This effectively

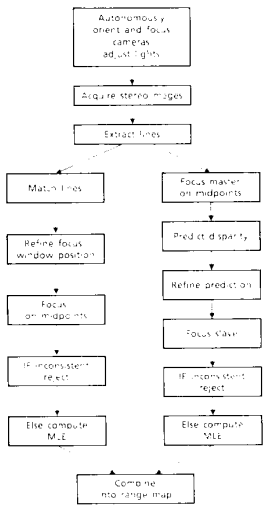


Figure 3. Overview of Sensing Operations

"pulls" the predicted window toward "interesting areas," on which W_{ref} is now centered.

The verification procedure predicts the focus motor position M corresponding to a computed range Z_s by solving equation (2) for M. Next, the procedure establishes an interval $[M_1, M_2]$ of focus motor positions symmetric about M. It then finely quantizes and searches this interval, servoing to each motor position and evaluating the focus criterion function, identifying the motor position M^* eliciting the maximal response of the criterion function.

If the criterion function is not unimodal, implying that W_{ref} contains projections of points lying at significantly different distances (cf. figure 1), the verification procedure reports that cross-checking could not be completed. If the criterion function is monotonic over the search interval (i.e., the "mode" lies at M_1 or M_2), then the procedure reports that cross-checking fails and that the match is mistaken. Otherwise the criterion function values are unimodal, suggesting that the match is not unreasonable, and the procedure uses M^* in equation (2) to compute the range from focusing Z_{fm}. Finally, if cross-checking succeeds, the verification procedure stores the master image point location (u_m, v_m), and the ranges Z_s and Z_{fm} for further processing. The verification procedure is repeated for each of the stereo matches.

3.2. Focus Ranging with Verification by Stereo

Proceeding down the right branch in figure 3, now the goal is to perform focus ranging and to verify its results using stereo information. To accomplish this, first the master camera executes focus ranging. Then, using stereo geometry to predict corresponding feature locations, the slave camera verifies the results, cross-checking focusing by stereo.

The focus ranging procedure considers a line segment L extracted from the master image to be an "interesting area," even if no range value from stereo is available there (i.e., the segment was not matched). All points along L are legitimate candidates for focus ranging, but for simplicity, the procedure calls for focusing on only the midpoint L_m. The procedure starts with L_m, predicts its location in the magnified image, builds a window W_{pred} around it, and adaptively refines the W_{pred} (as in the previous section) to compute an evaluation window W_{ref}. This process of prediction and refinement is repeated for each line segment.

In contrast to the verification by focusing described in section 3.3, which is guided by the range computed by stereo, now there is no *a priori* knowledge about the distances of imaged structures in

the W_{ref} to guide the focus ranging procedure, which must search for the best motor position M^* for each W_{ref}. If the criterion function is monotonic or not unimodal, then the master camera process reports that it can not complete the range measurement, and goes on to consider the next window. Otherwise, the procedure computes the range Z_{fm} (in the master camera coordinate frame) to the point(s) projecting to the window, using equation (2). Once finished making all the range measurements, the verification procedure begins.

The stereo process described in section 2.3 computes range from disparity. Conversely, in order to verify the master focusing measurements with the slave camera, now the range is used instead to predict a disparity. A given range Z_{fm} measured at a given vergence angle implies a stereo disparity $\underline{d}$, which can be identified by solving equation (3) for the x-component of $\underline{d}$. The predicted location of the center of an evaluation window W_S in the slave image is the sum (difference) of its location in the master image and the stereo disparity (in image coordinates): $\underline{C}_{slave} = \underline{C}_{master} \pm \underline{d}$, with the sign depending on which camera is the master.

The verification procedure evaluates the focus criterion function on W_S using the slave camera, searching a small, finely quantized interval around the best master focus position M^*. If the criterion function is not unimodal, then the slave camera reports that cross-checking could not be completed. If the criterion function is monotonic, then the slave camera reports that it can not verify the point, and that cross-checking fails. If cross-checking succeeds, the verification procedure computes the range Z_{fs} from focusing the slave camera, and records the master image point location (u_m, v_m), and the ranges Z_{fm} and Z_{fs} for further processing. The verification procedure is repeated for each of the master's range measurements.

4. Combination Policy

The outcome of the cooperative ranging process so far (up to the middle of the branches in figure 3) 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 $\{(Z_1, Z_2)\}$ of the underlying state of nature $\{Z^*\}$. It is natural to now seek a single, "better" estimate of Z^* which is "more accurate" than Z_1 or Z_2. For the purposes of this paper, "better" means "most likely," and "more accurate" means "lower expected variance." 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.

4.1. Assumptions

Let Z_1 and Z_2 represent independent measurements of the range Z^* of an object point. Here they represent the scalar ranges computed independently by focusing and stereo, but the following analysis can be extended to apply to any number of vector measurements from any kinds of measuring devices.

Suppose that the measurements Z_i are normally distributed $N(\mu_i, \sigma_i^2)$, $\sigma_i \neq 0$. As discussed in section 2, the variances are simply the squared uncertainties σ_f^2 and σ_s^2. 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 cross-checking, which can be viewed as a first consistency test. To further ensure (with a given probability) that the measurements are consistent, a statistical test attempts to reject this hypothesis.

4.2. Statistical Test of Measurement Consistency

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

$$T = \frac{Z_1 - Z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad . \tag{4}$$

The absolute value of T 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

$$consistent(Z_1, Z_2) = \begin{cases} 1 & |T| \leq T_\alpha \\ 0 & |T| > T_\alpha \end{cases} \quad . \tag{5}$$

If $|T|$ exceeds the threshold value T_α then we reject the hypothesis that the given measurements are consistent, i.e., represent the same physical quantity. An appropriate T_α can be chosen based on an acceptable error probability α considering that T obeys a standard normal distribution. Here they are implemented with an acceptable error probability of $\alpha = 0.05$, making $T_\alpha = 1.96$.

4.3. Maximum Likelihood Estimation

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 Z^* can be shown [9] to be:

$$Z = \frac{\sigma_2^2 Z_1 + \sigma_1^2 Z_2}{\sigma_1^2 + \sigma_2^2} \quad . \tag{6}$$

This expression is essentially a weighted average, where the weights are the variances of the measurement processes. It is demonstrated in [9] that the MLE is a more accurate (in the sense of lower variance) estimator of Z^* than any of the Z_i, since $\sigma_Z^2 < \min_i \{\sigma_i^2\}$. The right-hand side of this expression is an upper bound on the accuracy of the MLE.

For consistent measurements, the combination procedure uses equations of the form (6) to compute MLEs for stereo verified by focusing and for focusing verified by stereo. Finally, the combination procedure computes the three-dimensional point position by projecting the image points (u_m, v_m) in master image coordinates (cf. sections 3.1 and 3.2) through the pin-hole lens by

$$X = \frac{u_m Z}{f} \quad , \quad Y = \frac{v_m Z}{f} \quad . \tag{7}$$

The final outcome of the combination procedure is a set of computed points $\{(X, Y, Z)^T\}$ referred to the master camera coordinate frame.

5. Experimental Results

This section presents some experimentally obtained results of the cooperative ranging processes described in sections 3 and 4. It first describes the experimental procedure and test scenes, and then analyzes the reliability and accuracy of the implementation of cooperative ranging.

The experimental procedure begins by selecting scenes with structured objects appearing simultaneously and possibly occluding each other. The scenes contain such objects as a robot arm, a tool box, hand tools, a telephone (these first four objects have specularly reflecting surfaces), an eye chart, postal parcels, a scotch tape dispenser, a stapler, a pitcher, blocks, a stripe pattern, a circle pattern, a sweater, labeled envelopes, a striped tube, and a bottle of spray cleaner. Due to constrained laboratory space, they lie at distances between 1 and 3 m from the cameras. Next, we manually measure a number of the object distances. Then the sensing and cross-checking processes described in section 3 execute autonomously and completely unsupervised, with a focus evaluation window size of 20x20 pixels and fixed stereo matching thresholds, and no interactive "parameter tuning." After this, statistically consistent measurements (equation 5) are combined by a maximum likelihood estimator (equation 6) into a single range map. Finally,

the computed range map is compared to the known object distances.

We conducted 75 such 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 relatively small number of line segment features are extracted (on the order of twenty), and as a result of the conservative matching policy, there are even fewer correspondences computed (on the order of ten). 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 6 discusses strategies for increasing the size and density of the range maps.

5.1. Reliability

In general, the reliability of a measuring process decreases with its probability of failure. Here the measurements are MLEs of the range of an object point, and failures are mistaken measurements. As described in section 3, such failures occur when the stereo and focusing ranges are inconsistent, e.g., when distances to different object points are computed. So the reliability of the cooperative ranging process can be judged by how well it rejects inconsistent measurements.

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 (strictly speaking) 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.

5.2. Accuracy

In section 2.1 we observed that the relative error of focusing and stereo range measurements increases quadratically with object distance, and expressed their accuracies by the distance-dependent root-mean-square (rms) percent error U. The same figure here defines the accuracy of the cooperative range measurements. Accuracy results are tabulated 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 parallel (unconverged) lenses. Table 1 illustrates the accuracies for stereo ranging verified by focusing on 100 points computed over 14 experiments, and table 2 depicts the accuracies for focus ranging verified by stereo on 144 points computed over 10 experiments.

The summary of the data at the bottom of table 1 reveals that, considering 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 also shows that, considering 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, due to the fact that 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.

Table 1. Stereo Ranging Verified by Focusing.

N	Z^* [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. Focus Ranging Verified by Stereo.

N	Z^* [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

In both tables, N represents the number of measurements per experiment, equation (1) computes the observed range errors, and equation (6) computes the MLE.

6. Discussion

This paper has presented a procedure for autonomous cooperative ranging using focusing and stereo which consists of ensuring measurement consistency — by cross-checking and by statistical testing — and combining consistent measurements by a maximum likelihood estimator. This appears to be the first description and implementation of dynamic, adaptive, conjunctive operation of two visual sensors. The results of the experiments show that (*i*) the cooperative ranging procedure is robust, and that the computed range values are (*ii*) highly reliable, since mistaken combined range measurements are extremely rare, and (*iii*) more accurate than either of the computed ranges alone, as shown by the smaller rms percent error of the maximum likelihood estimates. These three central 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. These capabilities attest to the effectiveness of the procedure, and qualify it for many real-world applications.

While the sturdiness of the implementation is noteworthy, the reliability of the combined range measurements is especially significant. In 75 experiments considering approximately 3000 object points, not one mistaken range measurement survived cross-checking and statistical consistency testing. Although this does not imply that mistakes can not 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 reported in section 5.2 are less 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. Further work might improve the accuracy results by using combination rules other than maximum likelihood estimation.

Although the implementation adequately demonstrates the principle of cooperative ranging and practically illustrates the benefits of increased reliability and accuracy, 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 quite 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), decreasing the common field of view, and consequently limiting the possible quantity of computed range points. Using the same magnification 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 extracted lines. More dense three-dimensional information can be computed by using all or many points along these lines.

Third, rather than ignoring unverified measurements, they could be marked "unverified" and included in the output range map. It is also possible to feed back the information that cross-checking failed or could not be completed, and to reinitiate verification sensing with different parameters. For example, this could be accomplished for stereo matching by changing the line extraction parameters or the matching constraints or the vergence angle, and for focus ranging by changing the focus motor search interval or adjusting the location of the evaluation window.

This work considers two particular sensors — focusing and stereo — but there is certainly room for more and for others; motion stereo appears to be a plausible candidate. The addition of more sensors is straightforward for the combination policy (consistency test and maximum likelihood estimator) provided that their variances are known and that they measure like quantities.

While other research has attempted to increase measurement reliability by multi-modal sensing, confirming vision results with other sensors, this work shows how using passive visual sensors together can provide reliable three-dimensional measurements. In summary, we have found the principal virtue of cooperative ranging to be increasing the reliability of the measurements, while for the particular experimental conditions, the observed increase in measurement accuracy is marginal. The overall results lend credence to the view that the limitations of a single sensor can be circumvented by employing multiple sensors, and suggest that the cooperative methodology presented will allow teams of visual sensors to be used effectively in applications of computer vision to a wide variety of problems.

Acknowledgements. This work was performed at the University of Pennsylvania and was supported in part by NSF/DCR, the Air Force, DARPA/ONR, the ARMY, NSF-CER, DEC Corp., IBM Corp., LORD Corp., and the German Federal Ministry of Defense.

References

1. Aggarwal, J. K. and M. J. Magee, "Determining Motion Parameters Using Intensity Guided Range Sensing," *Pattern Recognition*, vol. 19, no. 2, pp. 169-180, 1986.
2. Allen, P. and R. Bajcsy, "Two Sensors Are Better Than One: Example of Integration of Vision and Touch," *Technical Report MS-CIS-85-29*, Computer Science Department, University of Pennsylvania, Philadelphia, 1985.
3. Chatila, R. and J.-P. Laumond, "Position Referencing and Consistent World Modeling for Mobile Robots," *Proc. IEEE Conference on Robotics and Automation*, pp. 138-145, St. Louis, March, 1985.
4. Flynn, A., "Redundant Sensors for Mobile Robot Navigation," *Technical Report 859*, MIT Artificial Intelligence Laboratory, October, 1985.
5. Harmon, S. Y., G. L. Bianchini, and B. E. Pinz, "Sensor Data Fusion Through a Distributed Blackboard," *Proc. IEEE Conference on Robotics and Automation*, pp. 1449-1454, San Francisco, April, 1986.
6. Kriegman, D. J. and E. Triendl, "A Mobile Robot: Sensing, Planning and Locomotion," *Proc. IEEE Conference on Robotics and Automation*, pp. 402-408, Raleigh, March, 1987.
7. Krotkov, E., R. Kories, and K. Henriksen, "Stereo Ranging with Verging Cameras: A Practical Calibration Procedure and Error Analysis," *Technical Report MS-CIS-86-86*, Computer Science Department, University of Pennsylvania, Philadelphia, December, 1986.
8. Krotkov, E., "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, In press, 1987.
9. Krotkov, E., "Exploratory Visual Sensing for Determining Spatial Layout with an Agile Stereo Camera System," *Ph.D. Dissertation MS-CIS-87-29*, Computer Science Department, University of Pennsylvania, Philadelphia, May, 1987.
10. Krotkov, E., J. F. Summers, and F. Fuma, "An Agile Stereo Camera System for Flexible Image Acquisition," *IEEE Journal of Robotics and Automation*, In press, 1988.
11. Nandhakumar, N. and J. K. Aggarwal, "Multisensor Integration - Experiments in Integrating Thermal and Visual Sensors," *Proc. International Conference on Computer Vision*, pp. 83-92, IEEE Computer Society, London, June, 1987.
12. Shafer, S., A. Stentz, and C. E. Thorpe, "An Architecture for Sensor Fusion in a Mobile Robot," *Proc. IEEE Conference on Robotics and Automation*, pp. 2002-2011, San Francisco, April, 1986.
13. Shekhar, S., O. Khatib, and M. Shimojo, "Sensor Fusion and Object Localization," *Proc. IEEE Conference on Robotics and Automation*, pp. 1623-1628, San Francisco, April, 1986.