Selection of fusion operations using rank-score diversity for robot mapping and localization

Damian M. Lyons  
*Fordham University*

D. Frank Hsu  
*Fordham University*

Qiang Ma  
*Fordham University*

Liang Wang  
*Fordham University*

Follow this and additional works at: [http://fordham.bepress.com/frcv_facultypubs](http://fordham.bepress.com/frcv_facultypubs)

Part of the [Robotics Commons](http://fordham.bepress.com/frcv_facultypubs)

**Recommended Citation**

Lyons, Damian M.; Hsu, D. Frank; Ma, Qiang; and Wang, Liang, "Selection of fusion operations using rank-score diversity for robot mapping and localization" (2007). *Faculty Publications*. 24.  

This Article is brought to you for free and open access by the Robotics and Computer Vision Laboratory at DigitalResearch@Fordham. It has been accepted for inclusion in Faculty Publications by an authorized administrator of DigitalResearch@Fordham. For more information, please contact jwatson9@fordham.edu.
Selection of fusion operations using rank-score diversity for robot mapping and localization

Damian M. Lyons, D. Frank Hsu, Qiang Ma and Liang Wang

Computer Vision & Robotics Laboratory
Department of Computer & Information Science
Fordham University, Bronx, NY 10458, USA

{lyons, hsu, ma, wang}@cis.fordham.edu

Abstract
In this paper, we evaluate the use of a rank-score diversity measure for selecting sensory fusion operations for a robot localization and mapping application. Our current application involves robot mapping and navigation in an outdoor urban search and rescue situation in which we have many similar and mutually occluding landmarks. The robot is a 4-wheel direct drive platform equipped with visual, stereo depth and ultrasound sensors.

In such an application it’s difficult to make useful and realistic assumptions about the sensor or environment statistics. Combinatorial Fusion Analysis (CFA) is used to develop an approach to fusion with unknown sensor and environment statistics. A metric is proposed that will indicate when fusion from a set of fusion alternatives will produce a more accurate estimation of depth than either sonar or stereo alone and when not. Experimental results are reported to illustrate that two CFA criteria are viable predictors to distinguish between positive fusion cases (the combined system performs better than or equal to the individual systems) and negative cases.

1.0 Introduction
Two key processes in the operation of a mobile robot platform are mapping and localization. Mapping provides a description of the environment - a map - that can be used for planning motion, and localization determines where the robot is with respect to the map. Simultaneous Localization and Mapping (SLAM) couples these two processes so that map and location are estimated recursively from sensory information [27]. By equipping the robot platform with multiple, diverse sensor types, the information available to mapping and localization, and hence the quality of the map and robot location in the map, can be improved, since one kind of sensor may provide information not available from other kinds of sensor.

For example, a stereo camera based depth sensor may work well in regions of high visual texture. However, if that visual texture arises from multiple overlapping surface edges, the angle of those edges may impair depth estimation by a sonar sensor.

To leverage this advantage, a sensor fusion algorithm needs to be developed that takes the information from each sensor and fuses it in such a fashion that the result is at least as good, in terms of accurately measuring the environment, as each sensor. Sensor fusion for robot mapping is a topic that has received attention in the research literature [3][5][7][22][25][27]. The problem addressed in this paper is a robust approach to low-level fusion [14], the collection of depth information – distances from the robot to points in the environment - from multiple sensors and their fusion to produce a more accurate depth description for use in mapping and localization.

If the statistics of the sensor and environment are known, then they can be used to construct a Kalman Filter or Extended Kalman Filter (EKF) formulation for this problem. Neira et al. [22] describe an EKF based approach for fusing range and intensity information from a laser ranging device for robot localization. Arras and Tomatis [1] use an EKF for fusing edge information from laser ranging and monocular vision. However, for sensors that have significantly different principles of operation and for cluttered, complicated environments (e.g., outdoors, debris pile [22]), it is difficult to model the statistics of sensor and environment in a useful fashion [13]. One approach is to use empirically determined rules to fuse sensors. Duffy et al. [6] use sonar to detect features and then use monocular vision to extract more information about the features. Another approach is to explore the limits of fusion with unknown sensor and environment statistics. Rao [25] addresses the problem of how well sensor probability distributions can be characterized if a finite number of calibration samples, pairs of sensed
versus actual features, can be taken before fusion begins.

In this paper we report results for an alternate approach to this problem: characterizing fusion as the problem of combining multiple lists of depth estimates, where a score value is associated with each depth estimate in each list. This approach, based on Hsu, Shapiro and Taksa [10], has the advantage of not depending on a model of the sensors or environment. It can therefore be used as a uniform multisensor fusion layer for many different sensor types used in a wide range of environments. Here we show that a direct application of this approach produces reasonable, but not strongly motivating results for a mobile robot generating depth information about an environment constructed to pose difficulties for sonar and stereo depth sensing. However, a combination of the approach with the probabilistic methods typically used in SLAM produces strong results.

2. The Approach

Hsu and Taksa [10] introduce an approach using the scoring behavior, the relationship between the scores assigned by an expert (e.g., a classifier, a filter, etc.) to a list of candidates and the ranks of the candidates, to determine how to combine multiple lists to produce a result that satisfies a performance metric at least as well as any single list. This approach, called Combinatorial Fusion Analysis or CFA ([8][10][19][30]) uses scoring behavior diversity and performance measures to select the best performing fusion operation between lists. In previous work on target tracking of people in surveillance video ([9][11][12][20][21]), we showed that it was possible to develop a metric that predicted, in the absence of any statistics on the sensors and environments, which fusion operation would perform most accurately.

In this paper we determine whether the same criteria can be applied to the problem of fusing stereo information with ultrasound ranging to generate depth information necessary for applications such as mapping and localization. Figure 1 shows our system architecture (similar to [3]). We collect depth information from ultrasound sensors (US) and a movable stereo camera (ST) as a mobile robot traverses a path in front of complicated environment. By complicated, we mean the environment consists of a cluttered scene with surfaces difficult for stereo or sonar or both. Each local map (unlike [3]) consists of candidate depth lists generated by a single sensor type. Calibration and registration information is used to translate the maps to a common frame of reference, at which point fusion can take place.

CFA selects a fusion operation from a set of available fusion operations based on a scoring behavior diversity metric. That metric is identified experimentally. The principle thrust of this paper is a proposal for, and an evaluation of a metric to select from a set of fusion operations for Stereo and Ultrasound sensors. We evaluate a set of fusion operations of the stereo and sonar data with respect to ground truth information on depth. We also evaluate for each fusion operation the CFA criteria developed in [8][12], namely, a feature performance ratio metric \( PR(A,B) \) for features \( A \) and \( B \), and a feature rank-score diversity metric \( d(f_A,f_B) \). We show that, in the absence of any assumptions about the statistics of sensors or environments, or any calibration sampling, these two features can be used to predict when fusion will produce a more accurate depth measurement and when not.

3. Combinatorial Fusion Analysis

We consider each feature measured by a sensor (which may measure multiple features as in the case of video) as a scoring system for the depth of a surface in the environment from the robot. Let \( D = \{d_1, d_2, ..., d_n\} \) be the set of depth estimates that a sensor can produce. For sensor \( A \), let \( s_A(d) \) be the scoring function which
assigns a real number to each $d_i$ in $D$ capturing the confidence of A that $d$ is the correct depth.

If the function $s_d(x)$ is treated as an array of real numbers, then after sorting the $s_d(x)$ array into descending order and assigning a rank (a natural number) to each of the $d_i$ in $D$, it leads to a rank function $r_d(x)$. The resulting rank function $r_d(x)$ is a function from $D$ to $\mathbb{N} = \{1, 2, \ldots, n\}$ (and note that $|D| = n$). We will assume that scoring functions for all sensors are range normalized to $s_d(x) : D \rightarrow [0, 1]$.

Given $m$ scoring systems $A_i$, $i = 1, 2, \ldots, m$, with score functions $s_A(x)$ and rank functions $r_A(x)$, there exist many different ways of combining the output of the scoring systems, including score combination, rank combination, voting, average combination and weighted combination [10][29]. We recognize that each of these has advantages and disadvantages. For the purpose of this paper, we will use CFA to evaluate the use of two combination operations: the average score combination, and the average rank combination. We restrict to just two for simplicity for this experiment and we select these two as examples of metric score combination (e.g., weighted combination, Mahalanobis combination, and so forth), and of order combination (e.g., min, max, voting, etc.).

For the $m$ scoring systems $A_i$ with $s_A(x)$ and $r_A(x)$, we define the score functions $s_R$ and $s_S$ of the rank combination (RC) and score combination (SC) respectively as:

$$s_R(x) = \frac{\sum_{i=1}^{m} r_A(x)}{m}, \quad \text{and}$$

$$s_S(x) = \frac{\sum_{i=1}^{m} s_A(x)}{m}. \quad \text{(1)}$$

$s_R(x)$ and $s_S(x)$ are then sorted into ascending and descending order to obtain the rank function of the rank combination $r_R(x)$ and the score combination $r_S(x)$, respectively.

The study of multiple scoring systems on large data sets $D$ involves sophisticated mathematical, statistical, and computational approaches and techniques (see e.g., [8] and refs) and is outside the scope of this paper, which is an application of this approach to a robotics problem.

Hsu, Chung and Kristal [8] have demonstrated that the combination of multiple scoring systems would improve the prediction or classification accuracy rate only if (a) each of the scoring systems has a relatively good performance, and (b) the individual scoring systems are distinctive (or diversified). Each of these needs to be quantified to be of use in selecting a fusion combination.

### 3.1 Diversity

The diversity $d(A,B)$ (dissimilarity or difference) between A and B has been studied previously using the score function diversity $d(s_A,s_B)$, correlation, and rank function diversity $d(r_A,r_B)$, rank correlation. Different diversity measurements have been considered in other application domains ([1], [5]-[8], [11], [12], [18], [22], [30]). A unique aspect of CFA is its definition of the rank-score function and associated diversity.

Let $s_d(x)$ and $r_d(x)$ be the score function and the rank function of the scoring system A. The rank-score function $f_d(x) : N \rightarrow [0,1]$ is defined as:

$$f_d(i) = (s_A^*(r_A^{-1})(i)) = s_A^*(r_A^{-1}(i)). \quad \text{(2)}$$

We note that the set $N$ is different from the set $D$ of depth hypotheses. The set $N$ is used as the index set for the rank function value. The rank-score function so defined characterizes the scoring behavior of the scoring system and is independent of any sensor or environment models. The rank-score diversity measure $d(A,B)=d(f_A,f_B)$ can be defined in several different fashions. Here we use absolute sum of differences between the two rank-score functions, as follows:

$$d(f_A,f_B) = \left| \sum_{i=1}^{\infty} (f_A(i) - f_B(i)) \right|. \quad \text{(3)}$$

### 3.2 Relative Performance

To evaluate whether each scoring system has good performance, we need to compare the scoring with ground truth.

Let $gt(d)$ be a ground truth scoring error of a depth estimate based on the absolute error, the absolute value of the difference between the depth estimate and the actual depth. We define the performance of a scoring system to be the inverse of the average sum of errors for the top $q$ ranked estimates, as follows:

$$P(A) = \frac{1}{q} \sum_{x:x \in \{r_A(x)\}^{q}} gt(x). \quad \text{(4)}$$

And the relative performance of two scoring systems can be quantified as:
A combined scoring system $C$ that uses systems $A$ and $B$ is positive if the performance of $C$ is better than or equal to the performance of $A$ and the performance of $B$:

$$P(C) \geq \max(P(A), P(B))$$  \hfill (6)

If (6) holds, we refer to this as a positive case of combination (fusion), otherwise we refer to it as a negative case.

4. Experimental Setup

Figure 2 shows the robot platform, a Pioneer P3 equipped with a Pan-Tilt stereovision head which could be moved to allow the stereo to view the same range as a given sonar sensor. Figure 3 (A) shows an image of the experimental surface – a depth surface constructed so as to offer challenges for stereo and sonar in a cluttered fashion. Textured regions will produce better stereo estimates than non-textured regions. Orthogonal (with respect to the sonar normal) regions produce better sonar estimates than angled regions. Figure 3(B) is a ground truth depth profile (measured by hand from the sonar sensors to the surface at each location) of the surface consisting of 24 measurements spaced along a 2.4m path parallel to the experimental surface. Figure 4 illustrates this 2.4m path taken by the robot during the experiment. At each of the 24 positions along this path, the front and back left-side sonar reading were taken, the stereo camera move to view each sonar’s field, and stereo readings taken. The SRI Small Vision [16] system was used to generate stereo depth maps.

These points $p_s$ were translated to a robot-centered coordinate system by:

$$p = p_s T_c R_b$$

where $T_c$ is the coordinate transformation matrix between the camera and robot systems with the pan-tilt in home position, and $R_b$ is the pan-tilt rotation matrix. Sonar range data is read using the Aria software [24] and also translated to robot-centered coordinates.

$$p = p_{u,n} T_{u,n}$$
where $T_{u,n}$ is the transformation for ultrasound sensor $n$. A cylinder $C_n$ is identified for each sonar in the robot-centered frame, a fixed radius $r_n$ around a line that is the central axis of the sonar. Whenever a sonar measurement is made with sonar $n$ then the cylinder $C_n$ is used to determine which points from the stereo depth map correspond with the sonar reading. $C_n$ was calculated by hand for each sonar and refined using a sequence of calibration experiments.

The following procedure was used to generate a ranked list of depth estimates from sonar and stereo:

(a) **Sonar**: A sequence of 100 sonar measurements are made for each of two sonar sensors facing the experimental surface. A (temporal) histogram is made from these values and used to produce a ranked list, where the score of each value is its frequency.

(b) **Stereo**: The set of depth values associated with each sonar sensor is collected into a (spatial) histogram, and these values used to produce a ranked list, where the score of each value is again its frequency.

5. **Experimental results**

The 24 measurements were made for each of the two ultrasound sensors, and associated stereo depth measurements were collected, resulting in 48 ranked lists. Average score and average rank fusions for each associated stereo and sonar list pair were calculated. In the case where a depth measurement value occurred in both lists, the fusion was straightforward. In the case where a value occurred in one list but not the other (as happens in many cases), a common ranked list was made by normalizing the scores in each list, merging the lists and re-ranking the merged list. The fusion score was calculated using the rank of the depth value in the original list and the merged list.

There are four scoring systems in this implementation: two sonar scoring systems and two stereo scoring systems. Each sonar system is paired with a stereo system. The results are shown in Figure 6 with the raw data from the sensors overlayed on ground truth. The highest ranked depth measurement for sonar and for stereo is shown for each of the 24 measurements and for each of the two sonars. The horizontal measurements correspond to the measurement number (from 1 to 24) which corresponds closely to the distance travelled by the robot parallel to the experimental surface. Sonar 1 is closer to the front of the vehicle than sonar 2. Thus the surface dip shown at positions 5 and 6 for Sonar 1, appear in positions 11 and 12 for Sonar 2. Notice that for Sonar 1, the measurements from position 18 onwards display large error with respect to ground truth. For the stereo head turned to sonar 1’s field of view, the stereo information from position 19 onwards also shows error.

![Figure 6: Ground truth information overlayed with sonar and stereo depth top-ranked measurements.](image-url)
The results of the combinatorial fusion analysis are shown in the scatter graph shown in Figure 7. Looking at the graph, it can be seen that the negative combinations, the combinations for which the performance of the combination, its closeness to ground truth depth, is worse than the performance of at least one of the combined features, cluster in the lower left of the graph. That is, in the area of low relative performance and low diversity. The positive combinations are more evenly scattered through the space, and cluster at a higher relative performance and diversity than the negative combinations. This result indicates that the diversity and performance metrics are separating positive and negative combination cases, but not in a very motivating fashion.

The principle problem is that the stereo and sonar lists have very few depth estimates in common, and a common list is crucial for CFA score and rank combinations. The use of the merged list as a way to handle this non-overlap is not very effective. We can devise an improvement by considering the usual probabilistic framework for mapping and localization.

5.1 Augmented Sensor Lists. A typical approach to SLAM uses a probabilistic model to interpret sensor results [26]. When a sonar is read, the measurement is convoluted with a probability density function (pdf), typically Gaussian. And this function is used to update the map. This pdf models sensor noise and uncertainty due to the beam spread of the sonar. We modify our approach to include this as follows.

Each of the estimates in the list of depth estimates from a sensor is used as the mean for a Gaussian kernel (see Fig. 8 (a) and then (b)). The kernel is then discretely sampled to produce extra depth estimates (Fig. 8(c)). In this manner, the number of common depth estimates between the sonar and stereo lists can be increased.

Figure 7: Scatter Graphs using sensor lists

Figure 8: Augmenting Ranked Sensor List
be improved. If the score of an estimate $d$ is $s(d)$, then the score of an added estimate $x$ is given by:

$$s(d) \ast p(x)$$

where $x \sim N(d, \sigma)$ and $\sigma$ is selected to produce the desired overlap of sensor lists.

The sonar lists from the experiment were augmented as indicated and the experiment repeated: the two fusions regenerated for all 24 measurements and the diversity and relative performance measures evaluated. The scatter graphs for the augmented lists are shown in Figure 9. The distinction in this case is now markedly better, with the positive cases clustering mostly in the region of high diversity and relative performance, and the negative cases clustering in low diversity and relative performance.

4. Conclusion

We have proposed an approach to the problem of multisensor fusion for robot mapping and localization that is widely applicable to many different kinds of sensors and operating environments because it does not rely upon modelling the sensor or environment. The approach draws upon the work of Hsu and Taks [10] in characterizing fusion as the general problem of combining multiple scoring systems. We introduced an architecture and framework for combining stereo and sonar depth measurements using this approach. We reported the results of an experiment to determine whether a diversity and performance measure could be used to identify fusion combinations of stereo and sonar that perform better (with respect to ground truth) than both the sonar or stereo measurement – that is – selecting the cases where combination makes effective use of the redundancy in sensors.

Our first result (Fig. 7) was obtained by generating an ordered list of depth estimates from a temporal histogram of sonar measurements and a spatial histogram of stereo depth measurements. The proposed metric showed some ability to separate positive and negative fusion cases. However, this was hampered by the fact that the sensor lists have very few depth estimates in common, and a common set of depth estimates is crucial for implementing the CFA approach. The problem was addressed by using the probabilistic sensor model typical in SLAM. The sensor lists were augmented in this manner so as to produce more estimates in common, and the resulting scatter graph (Fig. 9) shows a much improved separate of positive and negative cases clustering as expected – positive cases being fewer and clustered in the high diversity high relative performance regions and negative examples being more plentiful and clustered in the low diversity, low relative performance region.

There are still some outliers in Fig 9. and this may be due to several factors, including the registration between sonar and stereo which is modelled as a cylinder around the sonar axis. In fact, it is a cone.

The principal outcome of this paper is the metric developed. Note however, that the relative performance component of the metric cannot be used in practice to select fusion operations, since performance, and hence ground truth, will not be known. Rank-score diversity can be calculated in practice since it does not use ground truth. In [20] a similar metric to (3) is used to select fusion operations for video target tracking. The next step in this work is to evaluate the use of the diversity measure in practice in selecting between fusion operations and to compare its performance with other approaches to multisensory fusion.

References

|---------------|