

# Combinatorial Fusion Criteria for Robot Mapping

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

We address the problem of sensor fusion for stereo and ultrasound depth measurements for map building for a robot operating in a cluttered environment. In such a situation it's difficult to make useful and realistic assumptions about the sensor or environment statistics. Combinatorial Fusion Analysis is used to develop an approach to fusion with unknown sensor and environment statistics. A metric is proposed that shows 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. The metric consists of two criteria: (a) the performance ratio  $PR(A,B)$  between sensors A and B, and (b) the diversity  $d(A,B)$  between A and B as captured by the rank-score function  $f_A$  and  $f_B$ . Experimental results are reported to illustrate that these two CFA criteria are viable predictors to distinguish between positive cases (the combined system performs better than or equal to the individual systems) and negative cases.

## 1. Introduction

An important first step in the process of mapping and navigation for a mobile robot is extracting information from sensors about the physical environment surrounding the robot. A key advantage to equipping a mobile robot with a diverse set of sensors is that 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][24][26][28]. Our work here falls into what is typically regarded as low-level fusion [15]. 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. [24] 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.

For sensors that have significantly different principles of operation and for cluttered, complicated environments, it is difficult to model the statistics of sensor and environment in a useful fashion [14]. One approach is to use empirically determined rules to fuse sensors. For example, 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. For example, Rao [26] 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 previous work on target tracking of people in surveillance video, we have proposed an approach to fusion with unknown statistics using Combinatorial Fusion Analysis (CFA) [10],[12],[9],[21]. That approach, based on the work of Hsu, Shapiro and Taksa [11], characterizes the *scoring behaviour*, 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. In [9], we investigated the problem of tracking on a single camera, using multiple feature cue information, in situations where targets engage in multiple mutual occlusions. Investigating a set of fusion operations between ranked and scored lists generated by measuring color, shape and position target information in the video image, 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. We collect depth information from two ultrasound sensors and a movable stereo camera 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. A scored and ranked list of depth estimates is collected from each sensor for each measurement. In this paper, we adopt the frequency of measurement of a depth value over a spatial range, or over a time interval, as the score for that value. (If only the top ranked value is used, then this corresponds to a median filter.) We evaluate a set of fusion operations of the stereo and sonar data with respect to ground truth. We also evaluate for each fusion operation the CFA criteria developed in [9][13], 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.

## 2. Combinatorial Fusion Analysis

Our principal tool in identifying which features or pieces of evidence are most useful is the emerging field of CFA ([8]-[11], [19], [23], [24], [30]). The use of CFA has at least three distinct characteristics which are clearly advantages over the existing data and information fusion approaches (see e.g., [3], [27][29]). CFA considers: (A) both score and rank function for each feature/evidence and explores the interaction between the two functions using the rank-score function, (B) for a set of multiple scoring systems, both combinatorial possibility and computational efficiency of combining multiple scoring systems, and (C) fusion at both the data and decision levels, where (1) at the data level the multiple scoring systems are determined by sensors or characterized by features, and (2) at the decision level, a variety of scoring systems are obtained by different methods such as probability, statistics, analysis, combinatorics and computation. In our project, we (a) explore the scoring behavior of each of the features/evidence, (b) adopt CFA to inspect and analyse the *space of possible combinations* of the features or evidence, (c) use the difference between the two rank/score functions  $d(f_A, f_B)$  to characterize the diversity between A and B, and (d) use the rank/score function  $f_A$  to represent the scoring behavior of a feature or piece of evidence A.

We consider **each feature measured by a sensor** (which may measure multiple features) **or each piece of the evidence reported by a multiple sensor system** as a scoring system for the depth of a

surface in the environment. Let  $\mathbf{D} = \{d_1, d_2, \dots, d_n\}$  be the set of depth estimates. Let  $s_A(x)$  be the scoring function which assigns a real number to each  $d_i$  in  $\mathbf{D}$ . We view the function  $s_A(x)$  as the score function with respect to the scoring system (feature/evidence) A from  $\mathbf{D}$  to  $\mathbf{R}$  (the set of real numbers). When treating  $s_A(x)$  as an array of real numbers, it would lead to a rank function  $r_A(x)$  after sorting the  $s_A(x)$  array into descending order and assigning a rank (a positive natural number) to each of the  $d_i$  in  $\mathbf{D}$ . The resulting rank function  $r_A(x)$  is a function from  $\mathbf{D}$  to  $\mathbf{N} = \{1, 2, \dots, n\}$  (we note that  $|\mathbf{D}| = n$ ).

In order to properly compare and correctly combine score functions from multiple scoring systems (multiple features for a single sensor, or multiple items of evidence from multiple sensors) normalization is needed. We simply adopt the following transformation from  $s_A(x): \mathbf{D} \rightarrow \mathbf{R}$  to

$$s^*_A(x): \mathbf{D} \rightarrow [0, 1] \text{ where } s^*_A(x) = \frac{s_A(x) - s_{\min}}{s_{\max} - s_{\min}}, x \in \mathbf{D}$$

and  $s_{\max} = \max\{s_A(x) | x \in \mathbf{D}\}$  and

$$s_{\min} = \min\{s_A(x) | x \in \mathbf{D}\}.$$

Given  $m$  scoring systems  $A_i, i=1, 2, \dots, m$ , with score functions  $s_{A_i}(x)$  and rank function  $r_{A_i}(x)$ , there exist several different ways of combining the output of the scoring systems, including score combination, rank combination, voting, average combination and weighted combination. Initially we will use the average rank (or score) combination as follows. For the  $m$  scoring systems  $A_i$  with  $s_{A_i}(x)$  and  $r_{A_i}(x)$ , we define the score functions  $s_R$  and  $s_S$  of the rank combination (RC) and score combination (SC) respectively as:

$$\begin{aligned} s_R(x) &= \sum_{i=1}^m \left[ \frac{r_{A_i}(x)}{m} \right], \text{ and} \\ s_S(x) &= \sum_{i=1}^m \left[ \frac{s_{A_i}(x)}{m} \right]. \end{aligned} \quad (1)$$

As we did before,  $s_R(x)$  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.

When  $m$  scoring systems (features or evidence)  $A_i, i=1, 2, \dots, m$ , together with the score function  $s_{A_i}(x)$  and rank function  $r_{A_i}(x)$  are used,

combinatorially there are  $2^m - 1$  (  $= \sum_{k=1}^m \binom{m}{k}$  )

possible combinations for these  $m$  scoring systems using either rank or score functions. The order of complexity is exponential and becomes prohibitive when  $m$  is large. The study of multiple scoring systems on large data sets  $\mathbf{D}$  involves sophisticated

mathematical, statistical, and computational approaches and techniques (see e.g., [9] and refs). For example, each of the rank functions of the scoring system  $A_i, i=1,2,\dots,m$ , on  $\mathcal{D}, |\mathcal{D}|=n$ , can be mapped to a point in the  $n$ -dimensional polyhedra called the **rank space**. The  $n$ -dimensional polyhedron  $\mathcal{Q}_n$  is also a Cayley graph with the symmetric group  $S_n$  as the vertex set and the adjacency between vertices is defined by a set of generators (a subset of permutations) acting on its vertices.

**Remark 1:** Previous work using CFA ([8][9], [11], [13], [19],[24],[30]) in various application areas have demonstrated that: (1) the combination of multiple scoring systems (features or evidence) 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), and (2) rank combinations perform better than score combinations under conditions (a) and (b) and other restrictions.

**Remark 2:** The diversity  $d(A,B)$  (dissimilarity or difference) between A and B has been studied using the score functions  $d(s_A, s_B)$  and rank functions  $d(r_A, r_B)$  as correlation and rank correlation respectively. The approach of the current proposal, following the practice of [9], [11], [13], [30], [30], is to also use the concept of the rank/score function to measure the diversity between A and B. That is, we include  $d(f_A, f_B)$  as defined in formula (2) below *in addition to*  $d(s_A, s_B)$  and  $d(r_A, r_B)$ , where  $f_A, f_B$  are the rank/score functions of A and B respectively. The inclusion of  $d(f_A, f_B)$  in the measurement of the diversity between scoring systems A and B is one of the novelties of our approach.

When plotting the graph of the rank/score function (hence it is called the rank/score graph) of scoring systems A and B on the same coordinate plane, the diversity measure can be easily visualized. Different diversity measurements have been considered in other application domains ([1], [5]-[9], [12], [13], [19], [24], [30]).

Let  $s_A(x)$  and  $r_A(x)$  be the score function and the rank function of the scoring system A. The rank/score function  $f_A(x) : N \rightarrow [0, 1]$  is defined as:

$$f_A(i) = (s_A^* \circ r_A^{-1})(i) = s_A^*(r_A^{-1}(i)) \quad (2)$$

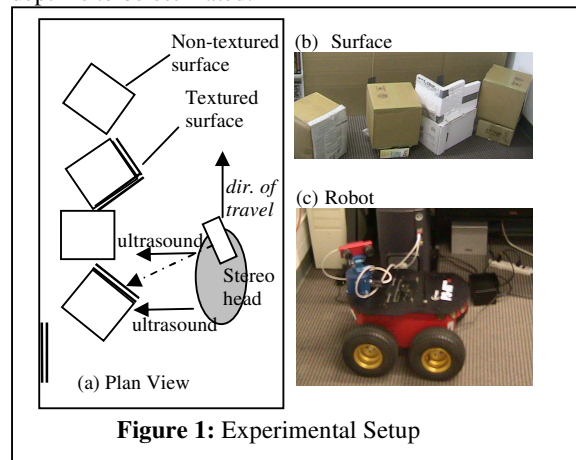
We note that the set  $N$  is different from the set  $\mathcal{D}$  which is the set of  $n$  depth estimates. The set  $N$  is used as the index set for the rank function value. The rank/score function so defined signifies the scoring (or ranking) behavior of the scoring system and is independent of depth estimates under consideration. Again, the diversity measure  $d(A,B)=d(f_A, f_B)$  can be defined in several different fashions. Here we use the following:

$$d(f_A, f_B) = \sum_{i=1}^n (f_A(i) - f_B(i))^2. \quad (3)$$

### 3. Experimental Investigation

#### 3.1 Design of Experiment

The objective in this experiment is to determine whether the diversity criterion for selecting fusion operations previously studied in video target tracking [13] can be of value in fusion of depth information from stereo vision and ultrasound sensors. The experimental setup is shown in Figure 1 (a-c). Figure 1(a) shows a sketch of the plan view of the experiment. Figure 1(b) shows a photograph of the robot and 1(c) a photograph of the surface whose depth is to be estimated.



**Figure 1:** Experimental Setup

The robot is driven along a straight line roughly parallel to the surface. The surface was chosen so that it offers multiple overlapping objects, whose position or appearance provide challenges to sonar (angled surfaces) and to stereo (non-textured surfaces) or to both. Sonar and stereo depth measurements are made at 24 locations along a 1.3m long path. Ground truth is measured by hand from the sonar sensors to the surface at each location.

A ranked list of depth measurements is obtained from sonar and from stereo camera sensors (implementation details in next section). The performance  $P$  of a sensor measurement or fused sensor measurements is calculated as the sum of the squared error of the measurement with respect to ground truth for the first  $q$  measurements in the list.

Two fusion alternatives were evaluated, an average score fusion and an average rank fusion, as described in formula (1). The fusion results are divided into *positive* and *negative* cases. A combined system C that uses sensors A and B is positive if the performance of C is better than the performance of A and the performance of B, i.e.:

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

For each combination, two performance metrics are evaluated. The *rank-score diversity*, calculated for a combination of features A and B as

$$d(f_A, f_B) = \sum_{i=1}^n (f_A(i) - f_B(i))^2$$

and the *performance ratio* metric,  $PR(A, B)$ , calculated as:

$$PR(A, B) = \frac{\min(P(A), P(B))}{\max(P(A), P(B))}$$

On each step, for each combination, the value of  $d(f_A, f_B)$ ,  $PR(A, B)$ , and whether the combination was positive or negative was recorded to a log file.

### 3.2 Implementation

The robot used for these experiments was a Pioneer AT3 robot, equipped with 16 ultrasound sensors and a *Videre Design* firewire stereo camera mounted on a *Biclops* pan-tilt base.

The SRI Small Vision [17] 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 [25] 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  $C_n$  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.

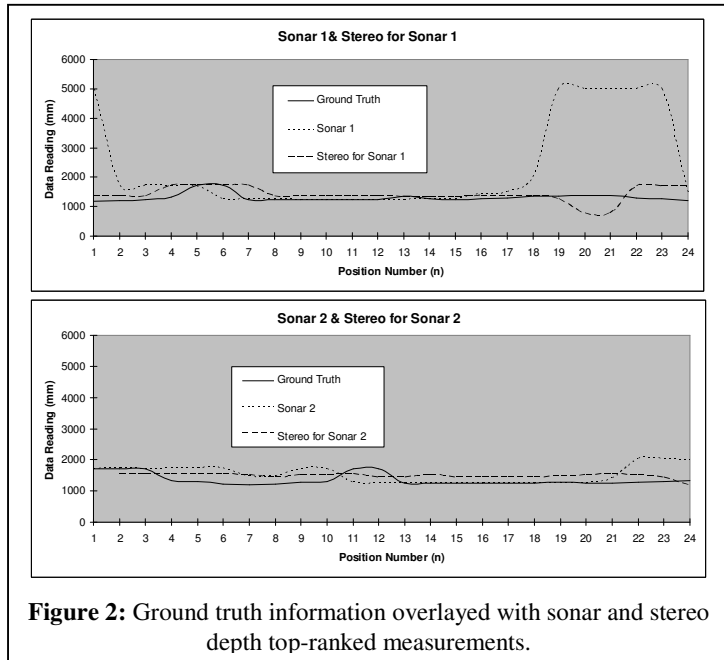
The following procedure was used to generate a ranked list of depth estimates from sonar and stereo:

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

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



**Figure 2:** Ground truth information overlaid with sonar and stereo depth top-ranked measurements.

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. In fact all the stereo information is measured from a single stereo head positioned differently for each sonar so as to observe the portion of the environment sensed by the sonar.

### 3.3 Results

Figure 2 shows raw data from the sensors overlaid 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.

The results of the combinatorial fusion analysis are shown in the scatter graph shown in Figure 3. Looking at the graph, it can be seen that the negative combinations, the combinations for which the performance of the combination, its closeness 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 is very close to what we observed in a more comprehensive experiment for video target tracking [13], and indicates that this approach has value also for selecting feature fusion alternatives in robot mapping applications.

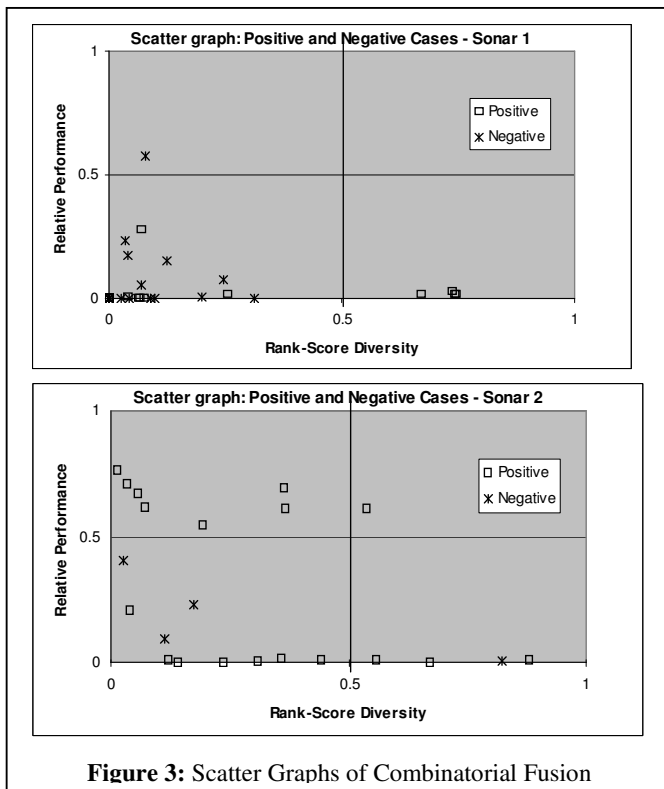


Figure 3: Scatter Graphs of Combinatorial Fusion

#### 4. Conclusion

In this paper, we have applied Combinatorial Fusion Analysis to the problem of fusing depth information from a stereo camera and an ultrasound sensor that are operating in a complex environment. We make no assumption about the statistics associated with the sensors or with the environment, and we do not take any sample measurements to attempt to estimate these statistics. Instead, we look at the scoring behaviour of the sensors, and show that a metric composed of a performance and a diversity component can be used to predict the performance of fusion operations.

The negative examples in Figure 3 cluster well in the area of low diversity and performance, however, the positive values are widely spread. This may be due to several issues:

- (1) The registration between sonar and stereo is modelled as a cylinder around the sonar axis. In fact, this is a cone.
- (2) The sonar range information produced lists of small length, due most likely to pre-filtering and smoothing within the Aria software [25].
- (3) Despite the technique for addressing the lack of common depth estimates in sonar and stereo lists still resulted in few common values.

Future work includes addressing these three potential causes of error.

#### References

- [1] Arras, K., Tomatis, N., Improving Robustness and Precision in Mobile Robot Localization by Using Laser Range Finding and Monocular Vision. *1999 3<sup>rd</sup> European Workshop on Advanced Mobile Robots*, Zurich, Switzerland, Sept 1999, pp177-185.
- [2] Brown, G., Wyatt, J., Harris, R., and Yao, X.; Diversity Creation Method: A survey and categorization. *Inf. Fusion* 6 (2005), pp5-20.
- [3] Castellanos, J., Neira, J., and Tardos, J., Multisensor Fusion for Simultaneous Localization and Map Building. *IEEE Trans Robt & Aut.* V17 N6 2001. pp. 908-914.
- [4] Dasarathy, B.V. (Editor); *Elucidative Fusion Systems – An Exposition*. Information Fusion 1 (200) pp.5-15.
- [5] DeSouza, G., and Kak, A., Vision for Mobile Robot Navigation: A Survey. *IEEE PAMI* V24, N2, Feb 2002. pp.237-267.
- [6] Duffy, B., Garcia, C., Rooney, C., and O'Hare G., *Sensor Fusion for Social*

- Robotics. 31<sup>st</sup> Int. Symp. On Robotics, May 14-17, Montreal Canada.
- [7] Ge, W., and Cao, Z., Mobile Robot Navigation Based on Multisensory Fusion, *LCIS 3612/2005* Springer-Verlag 2005.
- [8] Ho, T.K., Hull, J.J., and Srihari, S.N.; Decision Combination in Multiple Classifier Systems, *IEEE Trans on Pattern Analysis and Machine Intelligence*, 16(1), (1994) pp.66-75.
- [9] Hsu, D.F., Chung, Y.S., and Kristel, B.S.; Combinatorial Fusion Analysis: Methods and Practice of Combining Multiple Scoring Systems. In: (H.H. Hsu, editor) *Advanced Data Mining Technologies in Bioinformatics*, Ideal Group Inc, (2005) in press.
- [10] Hsu, D.F., Lyons, D.M., Usandivaras, C., and Montero, F. RAF: A Dynamic and Efficient Approach to Fusion for Multi-target Tracking in CCTV Surveillance. *IEEE Int. Conf. on Multisensor Fusion and Integration*. Tokyo, Japan; (2003) pp.222-228.
- [11] Hsu, D.F. and Taksa, I., Comparing rank and score combination methods for data fusion in information retrieval, *Information Retrieval* 8 (2005). pp.449-480.
- [12] Hsu, D.F., and Lyons, D.M., A Dynamic Pruning and Feature Selection Strategy for Real-Time Tracking. *19th IEEE International Conference on Advanced Information Networking and Applications*, March 28-30 (2005) pp. 117-124.
- [13] Hsu, D.F., and Lyons, D.M., Combinatorial Fusion Criteria for Real-Time Tracking. *20th IEEE International Conference on Advanced Information Networking and Applications*, March 28-30 (2006).
- [14] Hu, H., and Gan, J., Sensors and Data Fusion Algorithms in Mobile Robotics. Technical Report CSM-422 Univ. Essex, Dept of Comp. Sc., Colchester, UK, January 2005.
- [15] Kam, M., Zhu, X., and Kalata, P., Sensor Fusion for Mobile Robot Navigation. *Proceedings of the IEEE* V85 N1 Jan. 1997, pp.108-119.
- [16] Kittler, J., and Alkoot, F., Sum versus Vote Fusion in Multiple Classifier Systems. *IEEE Trans. on PAMI* (2003) 25(1): pp. 110-115.
- [17] Konolige, K. and Beymer, D., SRI Small Vision System --- Software User Manual 3.2g Nov. 2004.
- [18] Koschan, A., Kang, S., Paik, J., Abidi, B., and Abidi, M., Color active shape models for tracking non-rigid objects. *Pattern Recognition Letters* 24: pp. 1751-1765, July 2003.
- [19] Kuncheva, L., Diversity in Multiple Classifier Systems. *Information Fusion* 6(1), March 2005.
- [20] Lin, C.Y., Lin, K.L., Huang, C.D., Chang, H.M., Yang, C.Y., Lin, C.T., Tang, C.Y., and Hsu, D.F.; Feature Selection and Combination Criteria for improving Predictive Accuracy in Protein Structure Classification. *IEEE Symp. On Bioinformatics & Bioengineering* (2005) in press.
- [21] Lyons, D., and Hsu, D.F., Combinatorial Fusion for Target Tracking Using Rank-Score Characteristics. *Sub: Information Fusion 2005*.
- [22] Lyons, D., and Hsu, D.F., Rank-based Multisensory Fusion in Multitarget Video Tracking. *IEEE Intr. Conf. on Advanced Video & Signal-Based Surveillance*. Como, Italy. (2005).
- [23] Melnik, O., Vardi, Y., and Zhang, C-H., Mixed Group Ranks: Preference and Confidence in Classifier Combination. *IEEE PAMI* V26, N8, August 2004, pp.973-981.
- [24] Neira, J., Tardos, J., Horn, J., and Schmidt, G., Fusing Range and Intensity Images for Mobile Robot Localization. *IEEE Trans. Rob. And Aut.* V15 N1 Feb 1999. pp.76-84.
- [25] Pioneer 3 Operations Manual. MobileRobots Inc. Jan. 2006.
- [26] Rao, N.S.V., Multisensor Fusion under Unknown Distributions. In: Multisensor Fusion, A.K. Hyder, E. Shahbasian and E. Waltz Eds. Kluwer Academic 2002.
- [27] Thrun, S., Burgard, W., and Fox, D., A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots. *Machine Learning and Aut. Robots* 31/5 1998, pp. 1-25.
- [28] Varshney, P.K., Special Issue on Data Fusion. *Proc. of the IEEE* 85 (1) 1997.
- [29] Xu, L., Krzyzak, A., and Suen, C.Y., Method of Combining Multiple Classifiers and their Application to Handwriting Recognition. *IEEE Trans. SMC*, 22 (3): (1992). pp. 418-435.
- [30] Yang, J.M., Chen, Y.F., Shen, T.W., Kristal, B.S., and Hsu, D.F.; Consensus scoring criteria for improving enrichment in virtual screening. *J. of Chemical Inf. & Mod.* 45 (2005), pp 1134-1146.