Multi-Sensor Fusion: A Perspective

Jay K. Hackett *and Mubarak Shah
Computer Science Department
University of Central Florida
Orlando, FL 32816
*Now with Martin Marietta Electronic Systems, Orlando, FL

Abstract

Multi-sensor fusion deals with the combination of complementary and sometimes competing sensor data into a reliable estimate of the environment to achieve a sum which is better than the parts. Multi-sensor systems have applications in automatic target recognition, autonomous robot navigation, and automatic manufacturing. This paper presents a current survey of the state of the art in multi-sensor fusion. We have surveyed papers related to fusion and classified them into six categories: scene segmentation, representation, 3-D shape, sensor modeling, autonomous robots, and object recognition. A number of fusion strategies have been employed to combine sensor outputs. These strategies range from simple set intersection, logical *and* operations, and heuristic production rules to more complex methods involving non-linear least square fit and maximum likelihood estimates. Sensor uncertainty has been modeled using Bayesian probabilities, and support and plausibility involving the Dempster-Shafer formalism.

1 Introduction

Sensor fusion combines the output from two or more devices that retrieve a particular property of the environment. Commonly used sensors are a video camera, range finder, sonar, and tactile sensor.

The advantages of using multiple sensors are many. Since each sensor output contains noise and measurement errors, multiple sensors can be used to determine the same property, but with the consensus of all sensors. In this way, sensor uncertainty can be reduced. The output of a single sensor in some cases may be ambiguous and misleading, in which case another sensor can be used to resolve the ambiguity. For instance, vision does a poor job in scenes with shadows, so a range sensor can be used because it does not have such problems. In some cases, multiple sensor data can be integrated in a way that can provide information otherwise unavailable, or difficult to obtain from any single type of sensor. Every sensor is sensitive to a different property of the environment; in order to sense multiple properties, it is necessary to use multiple sensors. The system can be made fault tolerant by designing redundancy into the system. This means that a system using multiple sensors that sense a single property can be used. In case of failure of any single sensor, the system will still be able to function.

When raw sensor measurements are imprecise and noisy, they need to be modeled and characterized. Methods need to be developed for determining consistency of data, and fusion of consistent data. The precise operations involved in fusion depend on the level of data used. The most simple case of fusion for a multi-sensor configuration that records the same property of the environment is to combine the data using averaging. Here it is assumed that nothing is known a priori about the sensors. Thus each sensor's measurement is equally likely. In the cases when it is known that a particular sensor reading is more reliable than the others, a weighted average can be be used instead. Appropriate weights can be assigned proportional to the reliability of the sensor. This simple heuristic method will have problems when a large number of sensors are used and sensor interaction is more complex. There are formal approaches involving probability distributions to model these situations. Sensor fusion deals with the selection of a proper model for each sensor, and identification of an appropriate fusion method.

This paper deals with a current survey of research in the multi-sensor fusion area. We have confined ourselves to the papers related to vision, AI, and robotics. We are not aware of any previous survey done on the multi-sensor fusion topic, except a summary paper by Mitiche and Aggarwal [34], which summarizes their own group's research in sensor integration. A workshop on Multisensor Integration in Manufacturing was held in 1987 and a technical report by Henderson *et al.* [20] summarizes the conclusions of this workshop. Another workshop on Spatial Reasoning and Multisensor Fusion in the same year was organized by AAAI [24]. A special issue of International Journal of Robotics and Automation, edited by Brady in December 1988 [3] contained many papers on multi-sensor fusion.

A detailed version of this paper is given in [15].

2 Multisensor Systems

There are several methods for combining multiple data sources. Some of them are *deciding, guiding, averaging, Bayesian statistics*, and *integration*. Deciding is the use of one of the data sources during a certain time of the fusion process. Usually the decision as to which source to use is based upon some confidence measures or the use of the most dominant or more certain data. Averaging is the combination of several data sources, possibly in a weighted manner. The weights can be assigned based upon confidence values. This type of fusion ensures that all sensors play a role in the fusion process, but not all to the same degree. Guiding is the use of one or more sensors to focus the attention of another sensor on some part of the scene. An example of guiding is the use of intensity data to locate objects in a scene, and then the use of a tactile sensor to explore some of the objects in more detail. Integration is the delegation of various sensors to particular tasks. For instance, the intensity image may be used to find objects, the range image can then be used to find close objects, and then a tactile sensor can be used to help locate and pick up the close objects for further inspection. In this case, the data is not fused but is used in succession to complete a task. Therefore, there is no redundancy in sensor measurements.

Approaches to sensor fusion can be put into one general framework as shown in Figure 1. In this figure, sensors are shown by circles, and their outputs are denoted by $X_1, X_2, \ldots, X_n$. Corresponding to each sensor i, there is an input transformation denoted by f_i, which is shown by the oval shape. The input transformation could be the identity transformation, which does nothing to the input; that is the input and output are the same. On the other hand, it could be a simple operation like edge detection which will output an ede image, or a more complex task like object recognition, which will output a list of possible interpretations of objects present in the scene. The fusion is performed in the large rectangular block. We have listed a number of possible fusion strategies which can be used. The most simple fusion strategy will be the one in which raw sensor measurements of the same property obtained by multiple sensors are combined. For instance, focus and stereo range data can be combined using Bayes' rule. In another case, the sonar and infra-red depth measurements can be combined using simple *if-then rules*, or the range and intensity edge maps can be fused by using the logical *and* operation. Alternatively, a more complex fusion strategy might use weighted least-squares fit to determine an object's location and orientation using multiple sensor measurements.

A general block diagram for multi-sensor integration is shown in Figure 2. Multi-sensor integration is the *use* of several sensors in a sequential manner to achieve a particular task. With integration, a particular sensor performs a subtask or provides a particular piece of information. The next sensor may proceed by using any previous information to perform its own tasks. In this way, each sensor is used as an expert in its modality, but an overall consensus of all of the sensors is not achieved. Therefore, there is no redundancy in sensor measurements. For integration to work properly, there must be some type of control which organizes the flow of data from one sensor to the other. Integration is a much simpler process than fusion since the controller uses the data from only one sensor at a time to guide the actions of the other sensors. With fusion, however, the data from different sensors must usually be put into equivalent forms to allow the fusion to occur. The benefit of fusion occurs because the output is achieved by the consent of all of the sensors.

In order for data from multiple sensors to be fused, there must be some method to relate data points from one sensor with corresponding data points from the other sensors. The *registered* data points will easily allow for gathering of sensor information about one particular point in the scene. The registration can be done rather easily between some sensors, for instance intensity and range data can be registered by using known fields of view, tilt, and pan angles of the sensors. In case of different fields of view, image data for one sensor may not have corresponding data for the other sensor. The main problem we face in multi-sensor systems, and the one we want to solve through registration, is that sensors might provide data from different physical parts of an object.

3 Fusion Strategies

Each fusion approach is unique to some extent, however, certain key fusion methods and their variations have been employed by many authors. In this section, we will summarize some commonly used fusion strategies. In section five, we discuss current research that uses these strategies. Fusion methods can be classified broadly into two categories: direct and indirect fusion. In methods related to direct fusion, the raw sensor measurements are combined, while in indirect methods a transformation of the sensor measurements is fused. Before the sensor measurements can be fused, whether directly or indirectly, their consistency has to be checked (discussed in section 4). Bayesian theory has traditionally been used to model uncertainty in many disciplines for some time, thus there exists a well developed body of literature in this area. Therefore, a great number of approaches surveyed use Bayesian statistics as a fusion strategy. This will be discussed in subsection 3.1. Shafer-Dempster theory is another formalism that is used to model uncertainty. Since, it has certain advantages over Bayesian approaches a few authors have also used the Shafer-Dempster approach for fusion. We will summarize this method in subsection 3.2.

3.1 Bayesian Approaches

Bayesian statistics is very useful in combining multiple sensor values since sensor uncertainty can easily be incorporated. The state of the environment is decided based upon sensor measurements, knowledge about the types of states expected, as well as sensor uncertainty. New measurements can change the probability of a state occurring. A number of approaches surveyed in this paper make use of maximum likelihood, a well known Bayesian approach, as a fusion strategy. In this section we will review some of the basic concepts related to Bayesian approaches.

3.1.1 Direct Methods

The simpler forms of fusion employ raw sensor measurements directly. In this section, we will describe the maximum likelihood and Bayes' law for direct fusion. Assume that the sensor output is denoted by the vector $X = (x_1, x_2, \ldots x_n)$, and the object property (e.g., position, orientation, etc.) being estimated is denoted by Θ. We will be using two conditional probabilities: $p(X|\Theta)$ and $p(\Theta|X)$. $p(X|\Theta)$ is the probability of sensor output being X given that the object property is Θ, and $p(\Theta|X)$ is the probability of object property being Θ given that the sensor output is X. In our case, $p(X|\Theta)$ can be computed from the sensor model, while $p(\Theta|X)$ is the *a posteriori* probability which we want to determine. These two probabilities are related by *Bayes' Law*, which states:

$$p(\Theta|X) = \frac{p(X|\Theta)p(\Theta)}{p(X)} \tag{1}$$

where $p(X)$ and $p(\Theta)$ are the unconditional probabilities of the sensor output and object property being X and Θ respectively.

Assume that in our system there are k sensors, which give the following readings: $\mathcal{X} = (X^1, X^2, \ldots, X^k)$. We would like to develop the best estimate of the object property Θ using these k sensors readings. This can achieved by using the *likelihood estimate*. In the likelihood estimate we compute Θ such that the following is maximized:

$$p(\mathcal{X}|\Theta) = \prod_{l=1}^{k} p(X^l|\Theta) \tag{2}$$

It is usually easier to deal with the logarithm of the likelihood than the likelihood itself, because the product can be changed to the sum, and the terms involving exponents can be simplified. Let $L(\Theta)$ be the log-likelihood function:

$$L(\Theta) \;=\; \log p(\mathcal{X}|\Theta) = \sum_{l=1}^{k} \log p(X^l|\Theta) \tag{3}$$

Assume that the readings from the sensors follow Gaussian density functions. Then $p(X^l/\Theta)$ is given by:

$$p(X^l|\Theta) = \frac{1}{(2\pi)^{n/2}|C_l|^{1/2}} \exp(-\frac{1}{2}(X^l - \Theta)^t C_l^{-1}(X_l - \Theta)) \tag{4}$$

where C_l is the variance-covariance matrix, t denotes the transpose, and $|\;|$ denotes the determinant. Now, the expression for likelihood (equation 3) becomes:

$$\begin{aligned} L(\Theta) &= \sum_{l=1}^{k} \log p(X^l|\Theta) \\ &= \sum_{l=1}^{k} \left(-\frac{1}{2}\log[(2\Pi)^n|C_l|] - \frac{1}{2}(X^l - \Theta)^t C_l^{-1}(X^l - \Theta)\right) \end{aligned}$$

The best estimate $\tilde{\Theta}$ of Θ can be found by differentiating L with respect to Θ, equating the result to zero, and computing the value of Θ, as follows:

$$\tilde{\Theta} = \frac{\sum_{l=1}^{k} C_l^{-1} X^l}{\sum_{l=1}^{k} C_l^{-1}} \tag{5}$$

In the case when there are only two sensors in the system, and each sensor measurement, X_i, is a scalar, then the best estimate from equation 5 will be:

$$\tilde{\Theta} = \frac{({\sigma_2}^2)x^1 + ({\sigma_1}^2)x^2}{({\sigma_1}^2) + ({\sigma_2}^2)}$$

This equation shows the weighted average of two sensor readings; the weight is inversely proportional to the standard deviation of each sensor. In some cases, Bayes' law can be used directly to fuse the data coming from one sensor at multiple time instances, or the data from multiple sensors at one time instance. Matthies and Elfes, [32], use *Occupancy grids* to represent the space around a robot that is occupied by objects so that obstacle avoidance may occur. The occupancy grid is a 2D array of cells that contains probability values which denote the chance of the cell containing an object or part of an object. The probability of a cell being occupied $p(OCC|R)$, given sensor reading R, using Bayes law is given by:

$$p(OCC|R) = \frac{p(R|OCC)\,p(OCC)}{p(R|OCC)\,p(OCC) + p(R|EMP)\,p(EMP)} \quad (6)$$

where $p(OCC)$ and $p(EMP)$ are *a priori* probabilities of a cell being occupied, and empty, respectively. This equation is easily modified for sequential updating based on multiple readings.

3.1.2 Indirect Methods

The previous section dealt with the direct fusion of raw sensor measurements for a multiple sensor, single property configuration. In some cases, one sensor measurement can be related to other sensor measurements by some known transformation. In these cases sensor measurements can be fused indirectly. The work of Heeger and Hager [17] is an example of indirect fusion. They fuse optical flow data and camera motion parameters in order to obtain consistent object motion and depth information, and use it for segmenting the scene into moving and stationary objects. They develop a linear equation that relates image velocity (optical flow) to camera motion.

In a model based system where the object model (in terms of 3-D vectors) and the sensor locations are known, unknown translation and rotation parameters of the object with respect to the sensors can be computed by fusing the sensor readings. Shekhar *et al.*, [44], use a weighted least squares fit to fuse various sensor readings. Assume that the kth object point is denoted by vector p_k in the object coordinate system. If the object can rotate and translate, then its coordinates P_k in the sensor coordinate system are given by:

$$P_k = R\,p_k + h$$

where R is the rotation matrix, and h is the translation vector. Our aim is to find R and h which are consistent with several sensor readings. Shekhar *et al.* treat translation and rotation separately. For instance, for computing h they consider the following. For any distance measurement D_i along a direction n_i, there is a distance d_i in the model computed along n_i. These two measurements are related as follows:

$$D_i = d_i + n_i^t h \quad (7)$$

$$n_i^t h = D_i - d_i \quad (8)$$

For multiple measurements the above equation becomes:

$$\boldsymbol{C}h = \boldsymbol{d}$$

where $\boldsymbol{C} = [n_1, n_2, \ldots, n_n]^t$, $\boldsymbol{d} = [d_1, d_2, \ldots, d_n]^t$. Now, introducing the weights we get:

$$w_p \boldsymbol{C} h = w_p \boldsymbol{d}$$

where the weights w_p are the δd_i^2 (expected errors in distance). This equation is solved using the standard least squares fit with pseudo inverse method.

3.2 Dempster-Shafer Approaches

In Dempster theory the probability is assigned to propositions, (i.e., to subsets of a frame of discernment Θ). This is a major departure from the Bayesian formalism in which probability masses can be assigned to only singleton subsets. When a source of evidence assigns probability masses to the propositions represented by subsets of Θ, the resulting function is called a basic probability assignment (bpa). Formally, a bpa is a function $m : 2^\Theta \Longrightarrow [0,1]$ where $0.0 \leq m \leq 1.0$, $m(\Phi) = 0$, $\sum_{x \subseteq \Theta} m(X) = 1$. Dempster's rule of combining states that two bpa's, m_1 and m_2, corresponding to two independent sources of evidence, may be combined to yield a new bpa m:

$$m(X) = K \sum_{X_1 \cap X_2 = X} m_1(X_1) m_2(X_2) \quad (9)$$

where

$$K^{-1} = 1 - \sum_{X_1 \cap X_2 = \Phi} m_1(X_1) m_2(X_2). \quad (10)$$

This combination is termed the *orthogonal sum*. For each sensor, a mass distribution is formed which divides the input data into portions that provide belief for different propositions. Thus the sum of all masses which contribute to the belief for a proposition X will be denoted by m(X). Thus we can compute the orthogonal sum by an intersection of the components of the mass functions to find the total belief attributed to a proposition by both sensors.

The Shafer theory is based upon an interval of uncertainty, [s(A),p(A)]. Here s(A) denotes the support for a proposition A being true and p(A) is the plausibility of proposition A. The interval between p(A) and s(A) denotes the uncertainty about proposition A. If the uncertainty is zero then we simply have a Bayesian approach since the support for the proposition A is equal to the maximum likelihood. Support may be interpreted as the total positive effect a body of evidence has on a proposition, while plausibility represents the total extent to which a body of evidence fails to refute the proposition. Support for a proposition A is the total mass ascribed to A and to its subsets, the plausibility of A is one minus the sum of the mass assigned to $\neg A$, the uncertainty of A is equal to the mass remaining. More formally, $S(A) = \sum_{A_i \subset A} m(A_i)$, $P(A) = 1 - S(\neg A)$.

The Dempster-Shafer method is different from Bayesian approaches since an interval of uncertainty is used, while a Bayesian approach uses only one value which represents the probability of a proposition being true. The Bayesian approach also has difficulty in maintaining consistency when propositions are related. This is the case since many Bayesian approaches require independent measurements about the environment. Bayesian methods also require more complete information since a single point value must be computed. While the Dempster-Shafer method allows only an interval based upon the uncertainty to be computed.

4 Consistency Check

Before sensor measurements can be combined, we have to make sure that the measurements represent the same physical entity. Therefore, we need to check consistency of sensor measurements. The *Mahalanobis* distance, T, is very useful for determining which data should be fused. It is defined as:

$$T = \frac{1}{2}(X_1 - X_2)^t C^{-1} (X_1 - X_2) \quad (11)$$

where X_1 and X_2 are two sensor measurements, and C is the sum of variance-covariance matrices related to the two sensors. The minimal 'distance' will indicate a consistency between the two measurements. The Mahalanobis distance will be larger when two measurements are very inconsistent and it will decrease when uncertainty becomes less.

Krotkov and Kories, [26], use Mahalanobis distance as a consistency test for a system with two sensors; the reading from each sensor is a scalar denoted by x^1, x^2. In this case, the equation 11 simplifies to:

$$T = \frac{(x^1 - x^2)^2}{\sqrt{\sigma_1{}^2 + \sigma_2{}^2}} \quad (12)$$

where σ_1 and σ_2 are the standard deviations of sensor measurements x^1 and x^2. If $T \leq T_\alpha$, where T_α is some threshold, then the sensor measurements are consistent.

Luo *et al.* [28] use probability distances d_{ij}, and d_{ji} as the consistency check between sensors i and j.

$$d_{ij} = |\int_{x_i}^{x_j} P_i(x|x_i)P_i(x_i)dx| \quad (13)$$

$$d_{ji} = |\int_{x_j}^{x_i} P_j(x|x_j)P_j(x_j)dx| \quad (14)$$

where P_i, and P_j are *a priori* probabilities related to sensors i and j, and $P_i(x|x_i)$, and $P_j(x|x_j)$, are the conditional probabilities.

5 Survey of Existing Methods

This section surveys papers dealing with multi-sensor fusion and other related topics. We have attempted to classify the papers into six broad categories: segmentation, representation, 3-D shape, sensor modeling, autonomous robots, and recognition. This classification, however, is not strict; there might be some papers which belong to more than one category. For each group of papers, we have included a summary table listing the authors, sensors used, and fusion type employed. The largest group of papers deals with segmentation, and the smallest group of papers discusses object recognition. Due to the fact that segmentation is the earliest perception task, and involves lower level processing, fusion at that level is simpler, in general. Object recognition is the most sophisticated task, and involves fusion at the feature level.

5.1 Segmentation

Segmentation is one of the most basic low-level processes in computer vision. There are two types of segmentation: region-based and edge-based. In the region based segmentation, an attempt is made to group pixels in an image based on their similarity to one another. In edge-based segmentation, the boundaries of objects are identified by locating pixels where the change in pixel values is high. If segmentation is done accurately then each region should correspond to one object or one area of interest in the image. The majority of the papers deal with segmentation using range and intensity images. This is due to the fact that range and intensity images are readily available in a registered form from a laser range finder or a structured light sensor. Fusion is performed mostly at a lower level using pixel values and their distributions.

Duda *et al* [6] fuse intensity and range information by locating all horizontal and vertical surfaces by using a Hough Transform of the range information, and by finding all remaining surfaces employing a histogram of the intensity data. Hackett and Shah [14] fuse intensity and range data by choosing one of data sources at a time. In their method they use histograms of each type of data. A region merging algorithm is then used to refine the segmentation by using boundary strengths in both the range and intensity data. Wong and Hayrapetian [48] use range data to select thresholds based on distance. The intensity data is then partitioned by using these thresholds. Zuk *et al* [49] reject range data where the corresponding registered data point in the intensity (reflectance) information is very low. This is not real fusion, but they are using two sources of data. Gil *et al* [12] fuse range and intensity data by computing an edge map for each source of data. The edge maps are then combined by using a type of logical AND operator. Hu and Stockman [22] fuse intensity and sparse light-striping data. They use the intensity data to fill in where 3-D data is not available. This is done by triggering rules based upon the type of edge found in the intensity data and the type of stripe pattern found.

The use of other sensors like thermal with vision has been limited, since a thermal sensor would mostly be useful for scenes containing objects with large temperature variations. Also, a thermal sensor does not provide direct 3-D information like a range sensor does. Nandhakumar and Aggarwal [36] use thermal and intensity data to determine the surface heat flux of objects in outdoor scenes. Different heat fluxes can be used to discriminate between objects, thus segmentation is achieved.

Mitiche [33] uses the rigid body theorem, which directly relates optical flow and depth information as a test for rigidity. When optical flow and range information are supplied, the theorem can be tested. Wherever in the image the theorem fails, a different motion is present. Thus segmentation of moving objects is achieved. Moerdler and Kender [35] use a Hough-like transformation to consolidate various orientation constraints. They do not fuse data from different sensors but they fuse the outputs of several shape from texture algorithms. Duncan *et al* [7] use a *learning automata* that allows for rewarding and penalizing of the use of certain sensors. As segmentation progresses, the automata determines whether a particular sensor should be used. If segmentation is proceeding badly then most likely it will be penalized and control will be relinquished to another sensor. See Figure 3 for a table that summarizes the reviewed research and the method of fusion employed for segmentation approaches.

5.2 Representation

A number of papers deal with the building of representations of objects and space by using multiple sensors, as shown in Figure 4. Representation techniques include octree, occupancy grid, spherical octree, 3-D position and orientation vectors, tactile and other visual features. An octree is a hierarchical representation of space, in which a cube of space is decomposed into eight equal volumes. Each volume (octant) may be split if it is not homogeneous, giving rise to a tree representing the workspace. A spherical octree structure is a generalization of the rectangular octree structure. The occupancy grid is a 2-D array of cells that contains probability values which denote the chance of the cell containing an object or part of an object. If a value in the grid is high then an object is probably occupying that space.

Matthies and Elfes [32] use Bayes Law directly by forming a separate occupancy grid for the sonar and range data. As new data is obtained it is added to the octree. Then a combined world octree is formed by using Bayes Law again. Kent *et al* [25] use the least squares method in order to match an octree created from a model of an object to the objects found in an image. Shekhar *et al* [44] use the weighted least square technique for fusion. Transformations are defined by relating the sensor outputs with object models, and the least squares fit is used to estimate the unknown parameters. Stansfield [45] uses stereo vision and a tactile sensor to develop 3-D edge and tactile features such as roughness and corners. A hierarchical shape representation is developed from both sources of data. The vision is used first to gather rough position parameters of objects and then the tactile sensor is guided to locations where more information can be gathered. Crowley [5] uses several range sensors in order to develop a complete 3-D surface model of objects encountered. Generalized surface patches are created by considering conflicting and complementary data from all of the sensors. Chen [4] updates a spherical octree over time by using multiple vision sensors so that obstacle avoidance may occur. As new information becomes available it is simply added to the octree.

5.3 3-D Shape

The papers in this category deal with methods for computing the depth information by using multiple sensors, see Figure 5. Krotkov and Kories [26] fuse focus and stereo ranging at the lowest level using the maximum likelihood method. However, before values are fused they are checked for consistency.

In the remaining papers the data is fused at the intermediate level. For instance, Heeger and Hager [17] develop a linear equation that relates optical flow to camera motion to provide depth information. The Mahalanobis distance is used to provide consistency and the maximum likelihood is used to fuse both sensor measurements. Henderson *et al* [19] determine 3-D structure in the scene by using several visual views. They assume that angular invariance holds so that by looking at the relationships between lines in the images they can directly determine 3-D structure by using different views. Shaw and deFigueiredo [42] fuse microwave radar data and surface orientation obtained from a visual image in a consistent manner. Wang and Aggarwal [47] combine surface information obtained from occluding contour and light striping data by using rules that determine which sensor's data will be used to provide 3-D structure for a particular portion of the scene.

5.4 Sensor Modeling

Modeling of sensor characteristics is very important for a multi-sensor fusion system. Sensor measurements in general are imprecise and contain errors and uncertainties. The measurement error can be approximated by a probability distribution. The Gaussian distribution is commonly used. The estimate of various distribution parameters like mean and variance-covariance matrices is needed in a sensor fusion system. Durrant-Whyte [8] employs the summation of two Gaussians to model uncertainty in the sensor measurements and the maximum likelihood is used to fuse only the consistent measurements. Porrill [37] discusses a method for iteratively updating the variance-covariance matrix. The variance-covariance matrix includes errors involved with sensor calibration, actual feature location, and sensor measurements of that feature. If an error is present then the other values in the variance-covariance matrix must be adjusted to provide a better estimate of the parameters obtained from the scene.

A simulation module which allows a multi-sensor system to be modeled and tested before construction is very useful. Henderson *et al.* [18,20], in a series of reports, have advocated the use of a logical sensor system to simplify simulation. Another important step in a multiple sensor, multiple data configuration is the problem of correspondence or registration; Fernandez [9] discusses several algorithms such as the maximum likelihood and Mahalanobis distance to measure the similarity among data vectors from different sensors. Luo *al* [28] compute the distance between two probability distribution functions obtained from sensor measurements. This is used to obtain consistent information so that the maximum likelihood can be used to combine sensor measurements. Harmon *et al* [16] propose the use of averaging, guiding, and deciding for fusing data from different sensors that provide data about the same observations in the scene. Huntsburger and Jayaramamurthy [23] employ the Dempster-Shafer formalism to fuse information obtained from a sequence of visual images. Figure 6 outlines all reviewed sensor modeling approaches.

5.5 Autonomous Robots and Navigation

As can be seen in Figure 7, almost all papers reviewed in this section use guiding as a fusion strategy. This means that one sensor's data is used to instruct the other sensor what to do or what to look at next. These methods can be termed integration methods.

The paper by Ruokangas *et al.* [40] is a good example of sensor integration for an autonomous robot. They use vision, acoustic ranging, and force torque sensors in a controlled robot workcell. They consider a task of acquiring bolts from known positions. While the vision system provides scene gauging and object location, the acoustic ranger provides the data to determine the camera's correct focal distance, and hence, the vision system's gauge scale. The papers by Turk *at al* [46], Shafer *et al* [41], and Giralt *et al* [13] are similar in that the result is to navigate either along a road or inside a laboratory. They all search for the path or road regions using one sensor and then use other sensors to look for obstacles. Flynn [10] uses sonar and an infra-red distance sensor to build a map of the environment so that path planning can occur. Since each sensor is more accurate for different ranges of distances, rules are used to determine which sensor's data should be used depending upon the value returned by each sensor.

5.6 Recognition

One of the important goals of a multi-sensor system is to be able to recognize objects from its sensory inputs. Object recognition is a well developed area of research in computer vision, and there are a number of approaches for recognizing objects using vision [27], range [38], and tactile [21] sensors. The aim of using multiple sensors for object recognition is to decrease feature ambiguity and to reduce the search space during matching. We have been able to find only four papers dealing with object recognition using multiple sensors, and one paper dealing with emitter detection, see Figure 8. Three of these papers demonstrate methods by using real scenes. Rodger and Browse [39] use visual and tactile features to recognize objects in the synthetic scenes, however, they do not distinguish among features from different sensors. They use positional and placement constraints to reduce the number of possible interpretations. Luo *et al.* [30] use a straightforward two stage method for recognizing objects. First visual features are used to discriminate objects, then tactile features are employed, if necessary. Allen *et al.* [1] use stereo vision to guide a tactile sensor, and then the tactile sensor is used to fill in 3-D data where the stereo has obtained information. Magee *et al.* [31] use range and intensity data. All of the above approaches consider sensor output which is a two dimensional array. The paper by Garvey *et al* [11] uses Dempster-Shafer theory to combine the evidences for the pulse width and RF frequency of a one-dimensional signal. This evidence will be used to determine the emitter type which produced the signal.

6 Summary

We have examined papers which describe various approaches to multi-sensor fusion. The sensors which have been employed in a multi-sensor environment include a video camera, tactile sensor, range finder, sonar, infra-red sensor, and torque sensor. The researchers have investigated the use of multiple sensors for scene segmentation, object recognition, autonomous robot navigation, building 3-D representation, and simulation and modeling of multi-sensor systems. Strategies for combining sensor measurements include Bayes law, maximum likelihood, Dempster-Shafer, logical *and*, set intersection, weighted least squares, Hough transform, guiding, integration, deciding, verification, and global consistency.

7 Future Work

It has become obvious from this survey that the current state of the art in multi-sensor fusion is in its infancy. There are, therefore, promising areas of future work in almost all categories discussed in this paper, and other related topics to multi-sensor fusion. One of the most important areas which will have a significant impact on the research in multi-sensor fusion is in sensor design. The majority of currently available sensors are slow, less robust, and expensive. Due to high cost of sensors (e.g. range finder), very few laboratories are equipped with more than two sensors. This scenario is reminiscent of vision research ten years ago, when the cameras and digitizing equipment were beyond the reach of every institution. Now, inexpensive cameras and digitizer boards for PC's with high resolution monitors are available at an affordable cost, which has made vision research a wide spread activity. Therefore, it is expected that the situation related to the availability of other sensors (range, infra-red, etc.) will improve in the future, and more research groups will be involved in multi-sensor fusion research. Another related issue is the availability of registered sensor data, for example registered intensity, thermal and range data, which will be useful for fusion at the lower level to achieve scene segmentation. In the future, with the availability of registered data we will experience an increase in the research activity in the segmentation area.

Uncertainty management has been active in the past, and will remain popular due to its mathematical elegance. The validity of sensor uncertainty models need to be justified, and robust methods for approximating the model parameters (e.g. variance-covariance matrix) need to be explored. In the past the effort in uncertainty management has mostly been limited to fusion at the lower level (depth values, or position and orientation vectors). In the future we will see more work being done on fusion at the feature level. The features in one sensor's data need to be interrelated to the features in another sensor's data. This will also necessitate a design of generalized representation techniques which can be employed for multiple sensors.

Another important area where multi-sensor systems can really make a difference is in object recognition. Surprisingly, very little work has been done on this topic. Since each sensor is sensitive to different modality, multiple sensors not only can provide multiple views of objects, but they can also impose more constraints to reduce the search space during matching. For certain features there will be a direct correlation between sensors. A line segment in the visual image, for instance, should match to an edge in the tactile image, assuming both sensors have similar view and range. A feature supported by two sensors should have precedence over a feature supported by only a single sensor.

Finally, the implementation of multi-sensor fusion systems in real time needs special architectures employing parallel processing. There are several promising areas for the future work including the work on: mapping the current sensor fusion algorithms to available parallel architectures, and implementation of the fusion methods in specialized hardware so that chips can be designed, and manufactured.

Acknowledgement: The authors are thankful to Donna Williams, and Kevin Bowyer for their thoughtful comments on the earlier draft of this paper. The research reported in this paper was supported by the Center for Research in Electro Optics and Lasers (CREOL) University of Central Florida under grant number 20-52-043 and by the National Science Foundation under grant number IRI 87-13120.

References

[1] Bajcsy, R. and Allen, P. *Robotics Research: The 2nd International Symposium*, chapter Converging Disparate Sensory Data, pages 81–86. The MIT Press, Cambridge, MA, 1985.

[2] Barnes, D.P. and Lee, M.H. and Hardy, N.W. A control and monitoring system for multiple-sensor industrial robots. In *Proc. 3rd Int. Conf. on Robot Vision and Sensory Controls*, pages 457–465, November 1983.

[3] M. Brady, editor. *Int. Journal of Robotics and Automation.* MIT Press, Dec., 1988.

[4] Chen, S. A geometric approach to multisensor fusion and spatial reasoning. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 201–210, October 1987.

[5] Crowley, J.L. Representation and maintenance of a composite surface model. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 1455–1462, April 1986.

[6] Duda, R.O. and Nitzan, D. and Barrett, P. Use of range and reflectance data to find planar surface regions. *IEEE PAMI*, 1:259–271, 1979.

[7] Duncan, J.S. and Gindi, G.R. and Narenda, K.S. Low level information fusion: Multisensor scene segmentation using learning automata. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 323–333, October 1987.

[8] Durrant-Whyte, H.F. *Integration, Coordination and Control of Multi-Sensor Robot Systems.* Kluwer Academic Publishers, Norwell, Mass., 1988.

[9] Fernandez, M.F. Data association techniques for multisensor fusion. In *IEEE*, pages 277–281, 1985.

[10] Flynn, A.M. Combining sonar and infrared sensors for mobile robot navigation. *The International Journal of Robotics Research*, pages 5–14, 1988.

[11] Garvey, T.D. and Lowrance, J.D. and Fischler, M.A. An inference technique for integrating knowledge from disparate sources. *IJCAI*, pages 319–325, 1983.

[12] Gil, B. and Mitiche, A. and Aggarwal, J.K. Experiments in combining intensity and range edge maps. *CGIP*, 21:395–411, 1983.

[13] Giralt, G. and Chatila, R. and Vaisset, M. *Robotics Research: The First International Symposium*, chapter An Integrated Navigation and Motion Control System for Autonomous Multisensory Mobile Robots, pages 191–214. The MIT Press, Cambridge MA, 1985.

[14] Hackett, J.K. and Shah, M. Segmentation using intensity and range data. *Optical Engineering*, pages 667–674, June 1989.

[15] Hackett, J.K. and Shah, M. Multi-sensor fusion. Technical Report CS-TR-16-89, University of Central Florida, Computer Science Department, October, 1989.

[16] Harmon, S.Y., and Bianchini, G.L., and Pinz, B.E. Sensor data fusion through a distributed blackboard. In *Proc. IEEE 1986 Int. Conf. on Robotics and Automation*, pages 1449–1454, April, 1986.

[17] Heeger, D.J., and Hager, G. Egomotion and the stabilized world. In *IEEE ICCV*, pages 435–440, 1988.

[18] Henderson, T. and Shilcrat, E. Logical sensor systems. Technical report, The University of Utah, Technical Report UUCS-84-002, March, 1984.

[19] Henderson, T. and Weitz, E. and Hansen, C. and Mitiche, A. Multisensor knowledge systems: Interpreting 3d structure. *The Int. Journal of Robotics Research*, 7:114–137, Dec. 1988.

[20] Henderson, T., et al. Workshop on multisensor integration in manufacturing automation. Technical report, The University of Utah, Technical Report UUCS-84-002, February, 1987.

[21] Hillis, W. A high resolution imaging touch sensor. *The International Journal of Robotics Research*, pages 33–44, 1982.

[22] Hu, G. and Stockman, G. 3-d scene analysis via fusion of light striped image and intensity image. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 138–147, October 1987.

[23] Huntsburger, T.L. and Jayaramamurthy, S.N. A framework for multisensor fusion in the presence of uncertainty. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 345–350, October 1987.

[24] A. Kak and S. Chen, editors. *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion.* AAAI, Oct., 1987.

[25] Kent, E.W. and Sheiner, M.O. and Hong, T. Building representations from fusions of multiple views. In *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pages 1634–1639, 1986.

[26] Krotkov, E. and Kories, R. Adaptive control of cooperating sensors: Focus and stereo ranging with an agile camera system. In *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pages 548–553, April 1986.

[27] Lowe, D.G. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, Boston, 1985.

[28] Luo, R.C. and Lin, M. and Scherp, R.S. Dynamic multi-sensor data fusion system for intelligent robots. In *IEEE Journal of Robotics and Automation*, volume 4, pages 386–396, August 1988.

[29] Luo, R.C. and Mullen Jr., R.E. and Wessell, D.E. An adaptive robotic tracking system using optical flow. In *Proc. 1988 IEEE Int. Conf. on Robotics and Automation*, pages 568–573, April 1988.

[30] Luo, R.C. and Tsai, W. Object recognition using tactile image array sensors. In *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pages 1248–1253, April 1986.

[31] Magee, M. and Boyter, B. and Chieu, C. and Aggarwal, J. Experiments in intensity guided range sensing recognition of three-dimensional objects. *IEEE PAMI*, 7:629–637, November 1985.

[32] Matthies, L. and Elfes, A. Integration of sonar and stereo range data using a grid-based representation. In *Proc. 1988 IEEE Int. Conf. on Robotics and Automation*, pages 727–733, April 1988.

[33] Mitiche, A. On combining stereopsis and kineopsis for space perception. In *Proc. IEEE First Conf. on Artificial Intelligence Applications*, pages 156–160, Dec. 1984.

[34] Mitiche, A. and Aggarwal, J.K. Multiple sensor integration/fusion through image processing: a review. *Optical Engineering*, 25:380–386, March 1986.

[35] Moerdler, M. and Kender, J. An approach to the fusion of multiple shape from texture algorithms. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 272–281, October 1987.

[36] Nandhakumar, N. and Aggarwal, J.K. Multisensor integration - experiments in integrating thermal and visual sensors. In *IEEE ICCV*, pages 83–92, 1987.

[37] Porrill, J. Optimal combination and constraints for geometrical sensor data. *The International Journal of Robotics Research*, pages 66–77, 1988.

[38] Reeves, A. and Taylor, R. Identification of three-dimensional objects using range information. *IEEE PAMI*, 11:403–410, April, 1989.

[39] Rodger, J.C. and Browse, R.A. An object-based representation for multisensory robotic perception. In *Proc. of the 1987 Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 13–20, October 1987.

[40] Ruokangas, C. and Black, M. and Martin, J. and Schoenwald, J. Integration of multiple sensors to provide flexible control strategies. In *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pages 1947–1953, 1986.

[41] Shafer, S.A. and Stentz, A. and Thorpe, C. An architecture for sensor fusion in a mobile robot. Technical report, Carnegie Mellon University, Technical Report CMU-RI-TR-86-9, April, 1986.

[42] Shaw, S.W. and deFigueiredo, R.J.P. and Krishen, K. Fusion of radar and optical sensors for space robotic vision. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 1842–1846, April 1988.

[43] Shekhar, S. and Khatib, O. and Shimojo, M. Sensor fusion and object localization. In *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pages 1623–1628, April 1986.

[44] Shekhar, S., and Khatib, O. and Shimojo, M. Object localization with multiple sensors. *The Int. Journal of Robotics Research*, 7:34–44, Dec. 1988.

[45] Stansfield, S.A. A robotic perceptual system utilizing passive vision and active touch. *The International Journal of Robotics Research*, pages 138–161, 1988.

[46] Turk, M.A. and Morgenthaler, D.G. and Gremban, K.D. and Marra, M. Video road-following for the autonomous land vehicle. In *Proc. 1987 IEEE Int. Conf. on Robotics and Automation*, pages 273–280, March 1987.

[47] Wang, Y.F. and Aggarwal, J.K. On modeling 3d objects using multiple sensory data. In *Proc. 1987 IEEE Int. Conf. on Robotics and Automation*, pages 1098–1103, March 1987.

[48] Wong, R.Y. and Hayrapetian, K. Image processing with intensity and range data. In *Proc. 1982 IEEE Computer Society Conference on Pattern Recognition and Image Processing*, pages 518–520, June 1982.

[49] Zuk, D. and Pont, F. and Franklin, R. and Dell'Eva, M. A system for autonomous land navigation. Technical report, Environmental Research Institute of Michigan, Ann Arbor, Michigan, November 5, 1985.

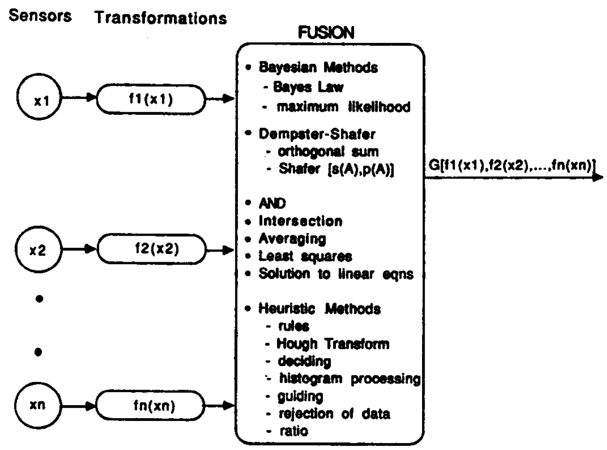

Figure 1: A Block diagram of Multi-sensor Fusion

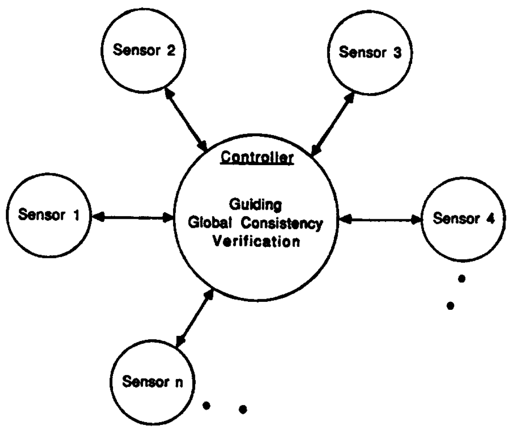

Figure 2: A Block diagram of Multi-sensor Integration

Authors	Sensor Data	Fusion Type
Henderson *et al.* [19]	abstract	logical sensors
Luo, Lin, Scherp [28]	abstract	maximum likelihood
Porrill [37]	stereo	update covariance matrix
Durrant-Whyte [8]	geometric sensors	maximum likelihood
Fernandez [9]	abstract	maximum likelihood
Harmon *et al.* [16]	abstract	guiding, deciding, averaging
Huntsburger *et al.* [23]	4 visual views	Shafer-Dempster

Figure 6: Approaches to sensor modeling

Authors	Sensor Data	Fusion Type
Duda *et al* [6]	range - vision	selection
Hackett, Shah [14]	range - vision	deciding
Wong *et al* [48]	range - vision	registration
Zuk, *et al.* [49]	range - vision	rejection of data
Gil *et al* [12]	range - vision	*and*
Nandhakumar *et al* [36]	thermal - vision	heat flux
Mitiche [33]	stereo range - motion	rigid body theorem
Hu, Stockman [22]	light stripes - vision	rules
Moerdler, Kender [35]	shape from texture	Hough transform
Duncan *et al.* [7]	-	deciding

Figure 3: Approaches to segmentation

Authors	Sensor Data	Fusion Type
Ruokangas *et al.* [40]	vision - range - force	guiding
Barnes *et al* [2]	vision - tactile - proximity	guiding
Flynn [10]	sonar - IR distance	rules
Luo *et al* [29]	sonar - vision	guiding
Shafer *et al* [41]	sonar - stereo - range	guiding
Giralt *et al* [13]	sonar - vision - range	guiding
Turk *et al.* [46]	color - range	guiding

Figure 7: Approaches to Autonomous robots and navigation

Authors	Sensor Data	Fusion Type
Kent *et al* [25]	multiple visual views	least squares
Shekhar *et al.* [43]	tactile - centroid	weighted least squares
Matthies, Elfes [32]	sonar - stereo range	Bayes law
Stansfield [45]	stereo - tactile	guiding
Crowley [5]	multiple range sensors	abstract
Chen [4]	multiple visual views	spherical octree

Figure 4: Representation methods

Authors	Sensor Data	Fusion Type
Rodger, Browse [39]	vision - tactile	consistency
Bajcsy, Allen [1]	vision - tactile	guiding/filling in
Luo, Tsai [30]	vision - tactile	hierarchical
Magee *et al.* [31]	vision - range	guiding
Garvey *et al.* [11]	frequency - pulse width	Shafer-Dempster

Figure 8: Methods for recognition

Authors	Sensor Data	Fusion Type
Heeger, Hager [17]	camera motion - optical flow	maximum likelihood
Henderson *et al.* [18]	multiple visual views	angular invariance
Shaw, deFigueiredo [42]	microwave radar - vision	guiding, minimization
Krotkov, Kories [26]	focus ranging - stereo ranging	verification
Wang, Aggarwal [47]	light striping - occluding contours	rules

Figure 5: Methods for determining 3D shape information