In *The Latin American Robotics Symposium* (LARS 09),
Valparaiso, Chile, October 29-30 2009.

Autonomous Information Fusion for Robust Obstacle Localization on a Humanoid Robot

Mohan Sridharan
Department of Computer Science
Texas Tech University, USA
mohan.sridharan@ttu.edu

Xiang Li
Department of Computer Science
Texas Tech University, USA
xiang.li@ttu.edu

***Abstract*—Recent developments in sensor technology [1], [2] have resulted in the deployment of mobile robots equipped with multiple sensors, in specific real-world applications [3]–[6]. A robot equipped with multiple sensors, however, obtains information about different regions of the scene, in different formats and with varying levels of uncertainty. In addition, the bits of information obtained from different sensors may contradict or complement each other. One open challenge to the widespread deployment of robots is the ability to fully utilize the information obtained from each sensor, in order to operate robustly in dynamic environments. This paper presents a probabilistic framework to address autonomous multisensor information fusion on a humanoid robot. The robot exploits the known structure of the environment to autonomously model the expected performance of the individual information processing schemes. The learned models are used to effectively merge the available information. As a result, the robot is able to robustly detect and localize mobile obstacles in its environment. The algorithm is fully implemented and tested on a humanoid robot platform (Aldebaran Naos [7]) in the robot soccer scenario.**
Keywords: Information fusion, Robot Vision, Obstacle localization, Humanoid robot soccer.

I. Introduction

Recent developments in sensor technology [1], [2] have resulted in the deployment of mobile robots equipped with multiple sensors, in specific real-world applications such as surveillance, navigation and disaster rescue [3]–[6], [8], [9]. The ability to accurately sense and interact with the environment is however still missing. One key challenge to the widespread deployment of mobile robots is the ability to operate autonomously in dynamic environments. The major aspects of this challenge include:

- *Autonomous learning*: How to enable a robot to learn environmental models based on sensory input, and adapt the learned models in response to changes?
- *Processing management*: Given multiple sources of information, which bits should be processed, and what processing should be performed to achieve a desired goal reliably and efficiently?
- *Multisensor information merging*: How to enable a robot equipped with multiple sensors to effectively merge the information obtained from the individual sensors?

There has been significant research on autonomous learning from sensory input [4], [6], [10], and processing management on mobile robots [3], [11]. However, the ability to autonomously exploit the complementary properties of the available sensors to effectively merge the available information, is still lacking. Each sensor mounted on a robot typically provides information on different regions of the scene, with varying levels of uncertainty. The bits of information obtained by processing the sensory inputs may contradict or complement each other. The visual input from a color camera, for instance, is a high-bandwidth source of information as compared to the range input from a laser range finder. Visual input is however more noisy and the visual information processing algorithms are typically computationally expensive. In order to operate robustly in dynamic environments, humanoid robots need to fully utilize the information obtained from the different sensors. However, the sophisticated algorithms used for motion control [12], [13] and the requirement of operating in real-time, make it a challenge to efficiently merge the information obtained from the different sources.

There has been extensive research on information fusion on a robot equipped with multiple sensors, for different applications [4], [14]–[16]. However, a major shortcoming of existing methods is that manually encoded heuristic constraints specify the conditions under which the information obtained from each specific sensor is given precedence. In the DARPA grand challenges, for instance, range and GPS information were used for most of the decision-making, while the visual input was predominantly used for only close-range obstacle avoidance [4], [6]. Such an approach that does not utilize all the available information, is likely to be at a disadvantage in a dynamically changing environment.

This paper advocates a probabilistic approach to effectively merge the information obtained from multiple sensors, and describes an instance of this approach for the challenge task of detecting and localizing mobile obstacles in a dynamic environment. It makes the following significant contributions: (a) an efficient approach for a robot to use the environmental structure to model the expected performance of the algorithms that process the sensory inputs, and (b) a probabilistic approach that uses the learned models to robustly combine the information obtained from the different sources. Furthermore, the robot is able to better exploit the rich information encoded in camera images. All algorithms are implemented and tested on a humanoid robot platform (Aldebaran Naos [7]).

The remainder of the paper is organized as follows. Section II describes the test domain and the proposed approach, while Section III describes the experimental results. Section IV provides a brief overview of some related methods, and the paper concludes with Section V.

II. Test Platform and Proposed Approach

This section first describes the experimental domain and the challenge task chosen to evaluate the proposed approach. This is followed by a description of the available information sources, and the algorithm that effectively merges the information obtained from these sources.

A. Test Platform

RoboCup is a research initiative with the stated goal of creating, by the year 2050, a team of humanoid robots that can beat the champion human team in a game of soccer on an outdoor soccer field [17]. The Standard Platform League

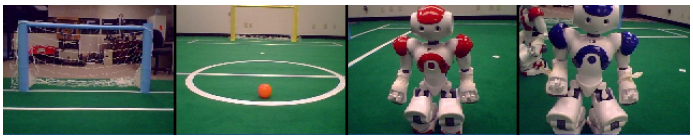

Fig. 1: Images of the Nao [7] and the soccer field.

of RoboCup has a team of humanoid robots (three robots per team) playing a competitive game of soccer on a $6m \times 4m$ indoor soccer field [18]. Figure 1 shows some images of the humanoid robot and the soccer field. The robot used as the common platform in this league is the Aldebaran Nao [7], a 58cm tall robot with 23 degrees of freedom—five in each arm and leg, two in the head, and one at the pelvis. The primary sensors are the two monocular color cameras, one in the forehead and one in the nose. The test platform imposes the constraint that only one camera can be used at a given instant, i.e. stereo capabilities do not exist. Each camera has a 58^o diagonal field of view, and provides images at a maximum resolution of 640×480—the 320×240 or the 160×120 images can be used for faster processing. There are two ultrasound sensors in the chest, one each on the left and the right with a 60^o field of view. Other sensors include accelerometers and bump sensors, in addition to microphones, loudspeakers and LEDs. The robot is equipped with Wi-Fi to communicate with other robots or an off-board PC. However, all processing for vision, locomotion, localization and team coordination is to be performed in real-time (30Hz) on board the robot, using the x86 AMD GEODE 500MHz CPU that runs embedded Linux.

The robot soccer framework presents many of the challenges faced while deploying a humanoid robot in the real-world (e.g. autonomous vision, motion, localization, team coordination). At the same time, it provides a moderate amount of structure that makes the domain tractable to solutions. It is therefore an ideal platform for our experiments.

B. Proposed Approach

The goal is to enable the robot to learn models that can predict the response of the information processing schemes, and to use these models to robustly fuse the information provided by the individual processing schemes. In this work, the processing schemes under consideration are:

- *Ultrasound (US)*: each ultrasound sensor provides a reading of object distance within a 60^o cone, up to a maximum distance of $\approx$ 150cm. The bearing information is limited to object presence on the left and/or the right.
- *Vision–Color (VC)*: Since the main objects in the domain (ball, robots, goals, field etc) are color-coded, color segmented regions in the input images are used to detect objects of interest based on heuristic constraints.
- *Vision–SIFT (VS)*: In order to extract maximum information from the images, we incorporate the popular SIFT (Scale Invariant Feature Transform) algorithm [19] that characterizes objects using image gradient features.

The task we choose to address is that of localizing mobile obstacles in the humanoid robot's environment. In this work, localization refers to the relative distance and bearing (angle with respect to the axis pointing straight ahead) of the obstacles with respect to the robot. In the robot soccer scenario, the major "obstacles"are the other robots (opponents and teammates) on the field. Collision with other robots can cause physical damage and is likely to provide the opponents with an advantage, since the rules of the game penalize robots that collide with each other. We include teammates despite the availability of wireless communication because the communication is typically delayed and noisy.

Each robot has a uniform of a specific color—all robots in one team are *red* while those on the other team are *blue*. As seen in Figure 1, each uniform is characterized by four large regions (head, shoulders, chest), and being able to see at least three of these regions arranged in a specific pattern, can be used to detect a robot in the image. However, the colored uniforms can be detected uniquely only from specific viewpoints, and up to a distance of $\approx$ 2m. In addition, given that the robot can only use one camera at a time, the distance to the object is computed based on a geometric comparison of the known object size and the detected size (in pixels) in the image. Since segmentation errors can affect the size of the detected image region, the computed distance can have significant errors. The bearing values are based on offsets (from the image center) of the segmented regions, and are more robust to such segmentation errors. The SIFT algorithm,

(a) (b)

Fig. 2: (a)-(b) Images with some gradient features superimposed.

on the other hand, characterizes objects of interest by local image gradient features that are known to be robust to scale, orientation and illumination changes [19]. Given such feature representations of the target object (in this case a robot),

recognition in test images can be achieved up to a distance of $\approx$ 4m. However, the distance to the object cannot be computed accurately, even though the bearing can be (once again) computed with reasonable accuracy. Figure 2 shows some images with the gradient features superimposed. Table I

Scheme	Distance		Bearing	
	FOV (cm)	Accuracy	FOV (deg)	Accuracy
Ultrasound (US)	20 to 150	high	80	low
Vision-Color (VC)	20 to $\approx$ 200	medium	190	high
Vision-SIFT (VS)	20 to $\approx$ 400	low	190	high

TABLE I: The field of view (FOV) and accuracy of distance and bearing computation using the available processing schemes. Vision has a larger angular FOV because the robot typically pans its head while moving forward.

compares the three processing schemes based on the field of view and accuracy. We see that the three available processing schemes have complementary characteristics. Typically, heuristic constraints are imposed (manually) on when (and how) the information from each of these sources should be used. Instead, our algorithm enables the robot to learn a model that can predict the performance of the individual schemes. Based on these learned models, Algorithm 1 assigns suitable "weights" to each source and robustly merges all the available information. The algorithm requires models of the expected

Algorithm 1 Multisensor Information Merging

Require: Learned models that predict the error in range distance and bearing for measurements from each information source.

Require: Learned SIFT model of the target object (in this case, robots).

1: **repeat**
2: $UpdateExistingEstimates()$
3: $\{d_{us}, dir\} = CurrentObstacles_{us}()$
4: $\{d_c, \theta_c\} = CurrentObstacles_{vc}()$
5: $\{\theta_s\} = CurrentObstacles_{vs}()$
6: $ResolveCurrentEstimates()$
7: $MergeWithExistingEstimates()$
8: **until** end of the game

error in each processing scheme, and the SIFT model of the target objects. Section III describes how these models can be learned, while the algorithm is described below.

Since the relative estimates of each obstacle need to be maintained and tracked across a sequence of frames, a Kalman filter [20] is associated with each estimate. The first step, the "time update" of the Kalman filters ($UpdateExistingEstimates()$, line 2), adjusts the existing obstacle estimates in response to the robot motion since the previous update. It also removes older estimates that correspond to obstacles that have not been seen for some time. Each processing scheme is then used to compute the distances and bearings of the obstacles in the current frame ($CurrentObstacles()$, lines 3–5). The ultrasound sensor provides limited bearing information (left, right or both), and *Vision-SIFT* does not provide distances to detected obstacles.

The next step ($ResolveCurrentEstimates()$, line 6) performs two tasks. It first groups the individual distances and bearings obtained from the current frame so that, for instance, the distances and bearings obtained from the ultrasound sensor are grouped with similar values computed using the colored regions. In the case of processing schemes that provide partial information (e.g. *Vision-SIFT* only provides bearing), grouping is done based on the available information. Since the number of detected obstacles in each frame is small, this grouping is accomplished using the expected errors in the measured values. For example, if the difference between the bearing computed using colored regions and the bearing computed using SIFT, is more than the expected error in the individual measurements, they are not grouped together. This "threshold" can be modified to detect obstacles at different resolutions, and more sophisticated data association can be performed [21], but it is not required in our domain.

Once the grouping has been completed, the values within a group are merged to arrive at a single estimate for each obstacle sensed in the current frame. The merged values are the weighted average of the individual values:

$$d^j = \sum_i w^j_{d,i} d^j_i \qquad (1)$$
$$\theta^j = \sum_i w^j_{\theta,i} \theta^j_i$$

where d^j and θ^j represent the computed distance and bearing to the jth obstacle in the current frame—they are the weighted averages of the values provided by the individual schemes ($i \in \{us, vc, vs\}$). The symbols $w_{d,i}$ and $w_{\theta,i}$ are the weights associated with the distance and bearing values obtained from the ith source. However, since SIFT cannot provide a distance, and the ultrasound only provides directional (left or right) information:

$$\theta^j = w_{\theta,us}\{w^j_{\theta,vc}\theta^j_{vc} + w^j_{\theta,vs}\theta^j_{vs}\} \qquad (2)$$
$$w_{d,vs} = 0, \quad w_{\theta,us} = \begin{cases} -1 & \text{if only right US triggers} \\ +1 & \text{otherwise} \end{cases}$$

The weights are a measure of the degree of trust associated with each processing scheme. They are obtained by normalizing the degree of certainty associated with each scheme:

$$w_i = \frac{p_i I_i}{\sum_j p_j I_j} \qquad (3)$$
$$p_{d,i} \propto 1/f_{d,i}, \quad p_{\theta,i} \propto 1/f_{\theta,i}$$
$$f(x) = a_0 + \sum_{k=1}^{N} a_k x^k : N \in [1,3],\ x = d \text{ or } \theta$$

where $p_i \in [0,1]$ is the certainty associated with a measurement from the ith source (us, c, s). It is a function of the expected error, which in turn is a function of the measured distance or bearing. The errors in the measured bearing values are mostly a function of the distance (except for *Vision-SIFT* where a distance value is not available). In situations

where several sensors are able to provide good distance and bearing measurements, the function can be changed to a joint function of distance and bearing, and the proposed functional form can be automatically inferred. The symbol I is the indicator function that indicates the presence or absence of a measurement from each source.

The individual and merged estimates generated from the current frame are matched with the existing estimates (from prior frames), using the same grouping process used in line 6. The grouped current and prior estimates are then merged ($MergeWithExistingEstimates()$, line 7) through the "measurement-update" step of the Kalman filters. The prior estimates without matches in the current frame are retained until they are eventually removed in the "time update" of line 2. It is possible to directly input the measured values to the Kalman filters, but that would still require that the current measurements be matched with the existing estimates. We also found that using the estimated errors (with the different processing schemes) in the Kalman filter noise models did not provide the desired accuracy. We show (below) that the proposed algorithm enables the robot to robustly detect and localize obstacles in its environment.

The experimental results in this paper will consider other mobile robots as the obstacles, and will use the proposed approach to robustly find the relative distance and bearing to these obstacles (i.e. localize the obstacles). In general, the challenge tasks, processing schemes, sensors and robot platforms may change. However, robots operating in a dynamic environment will still need to fully utilize the available information. With suitable models for the individual processing schemes and the target object representations (all of which can be learned as described in the next section), an approach similar to the proposed algorithm can be used to perform robust information fusion.

III. Experimental Setup and Results

This section describes the approach to learn the models required in Algorithm 1, and then describes the experiments conducted to evaluate the proposed approach.

A. Error Models and Visual Representation

The vision system on the humanoid robot follows an established sequence: input images are color segmented using a *color map* that maps image pixels to numerical color labels. Contiguous segmented regions of the same color are grouped into regions that are then used to detect objects based on heuristic constraints—see [15] for details.

Most mobile robot environments have a moderate amount of structure, which can be used to automate tasks that usually require extensive manual supervision. The robot soccer domain has objects of known color at known positions. Based on prior work where this knowledge was exploited to learn the color map [10] autonomously, we model the expected performance of the information processing schemes.

The information obtained from the different processing schemes can be merged autonomously if it were possible to compute the expected error in each measurement obtained from the individual schemes—a measurement with a lower error will automatically be assigned a larger weight in Equation 3 above. In order to learn such models of the expected error, obstacles (other robots in the current example) are placed at fixed positions on the field that are known to the robot. The robot is asked to move slowly through a sequence of poses (position+orientation) that it can reach with very high accuracy using cues from the standard visual processing sequence described above—typical examples include points on the center line of the field, and the white "dot" a fixed distance from each goal (that is used for penalty shootouts). At each such pose, the robot compares the actual distance and bearing values against the measured values to compute the measurement errors. The error values are collected and used to train a function approximator that models the measurement error as a function of the measured distance (or bearing), and then computes a number $p \in [0, 1]$ as a measure of the certainty of the measured values (Equation 3). Polynomial regression functions are used to approximate these errors (see Section II-B), and parameters of these functions (degree, coefficients) are learned using the collected statistics. Similar performance is achieved using more popular function approximators (e.g. neural networks [22]) but the polynomial functions are easier to estimate.

At each pose, the robot also projects the known positions of the obstacles within the field of view of the camera, to the image. The image gradient (SIFT) features extracted from the corresponding image regions are used to generate a training database of features that represents the robot, and a similar database is created for the background i.e. the environment. During testing, features extracted from the test images are compared with those in the training database, and a Nearest Neighbor classifier [22] is used to classify features and detect obstacles (> 3 feature matches).

B. Experimental Results

Given the learned models, the humanoid robot can now effectively localize the obstacles (i.e. other robots) on the field. The hypothesis we aim to test is that the merged estimate (using Algorithm 1) is more robust than the estimates obtained in the absence of this algorithm.

Kalman filters are used in Algorithm 1 to track the obstacles across frames. However, we are primarily interested in evaluating the detection and localization accuracy, and the corresponding experiments are described first. Though the different sensors have different field of view, we tried to make the test cases as challenging as possible.

The experiments consisted of the robot moving through a fixed sequence of poses with the obstacles placed at different points on the field. In our experimental scenario, a detection accuracy $< 100\%$ reflects the inability of the robot to find the obstacles (i.e. there are no false positives). The localization errors were hence computed only when the obstacles were detected correctly. When an obstacle is detected (using the processing scheme being evaluated), the robot stopped and

performed additional trials at that point. In each such trial, the relative distance and bearing of the detected obstacles were measured. The corresponding ground truth values were provided manually, except when the robot is well-localized and knows the global position of the obstacle it has detected. The difference between the estimated and ground truth values provide the error values that are documented in Table II. The first three rows of Table II correspond to the individual processing schemes: ultrasound (US), vision-color (VC) and vision-SIFT (VS). The last row corresponds to the results obtained with our algorithm, i.e. US+VC+VS. Each entry in the distance-error and bearing-error columns was computed over ≈ 20 different points (over the possible range of distance and bearing values), with ≈ 15 trials at each point.

The entries in the last column (labeled "Accuracy") in Table II were computed by capturing several images as the robot moved through the sequence of poses to compute the distance and bearing errors. The robot logged ≈ 400 images for each processing scheme, with more than half the images containing the obstacles. Some of these images corresponded to situations where the obstacles were outside the angular and/or distance-based field of view of one or more schemes. These images were hand-labeled to provide the ground truth (presence or absence of obstacles), and used to compute the detection accuracy of each processing scheme. The accuracy results reported in Table II are statistically significant. For the

Scheme	Error		Accuracy(%)
	Distance (cm)	Bearing (deg)	
Ultrasound (US)	6.5 ± 3.6	$--$	70
Vision-Color (VC)	17.5 ± 8.7	8.5 ± 4.0	81.5
Vision-Sift (VS)	$--$	9.1 ± 4.5	85.5
$US + VC + VS$	9.2 ± 5.1	8.8 ± 4.3	91.5

TABLE II: The distance and bearing errors, and the detection accuracy of the processing schemes. Proposed approach is more robust than the individual processing schemes.

challenge task under consideration (moving obstacles in the robot soccer domain), an average error of 10cm in distance and 10^o in bearing, along with a detection accuracy $> 90\%$ would be more than sufficient to operate robustly. The results in Table II are analyzed by comparing the measured values to these desired (i.e. target) values.

The processing scheme based on ultrasound information (row 1 in Table II) computes the distances very accurately but the bearings cannot be estimated. Though the sensor is very sensitive to obstacles, it cannot detect obstacles beyond a certain distance ($\leq$ 150cm) and outside the 60^o cone (for each sensor). Hence, though the detection accuracy is almost 100% within its limited detection zone, the overall accuracy over the wide range of test cases is only 70%.

When the obstacles are detected using just the color information (row 2 in Table II), the error in distance estimates is substantially higher because the distance computation based on image region sizes is noisy. The bearing estimates are however reasonably accurate. The overall detection accuracy is not good because the color-based detection is not valid at all viewpoints (the uniform on the obstacle robots is not visible at all viewpoints) and at large distances ($>$ 200cm).

When the obstacles are detected using just the SIFT-based processing scheme (row 3 in Table II), the bearing estimates are statistically similar to those obtained using color. The expected error in bearing does not change much as a function of the measured bearing, and the scale and orientation invariance results in a higher detection accuracy than the color-based scheme. The robots can now be detected at different viewpoints and up to a distance of $\approx$ 400cm. The scheme fails when the obstacles are a significant distance away from the robot, or if very few SIFT features are detected on the obstacles (e.g. due to strong highlights). However, distances to obstacles cannot be computed.

Though not included in Table II, combining ultrasound with one of the vision-based techniques does provide an improvement, but either the detection accuracy is low (e.g. US+VC), or the distance computation is inaccurate or infeasible (e.g. US+VS with the obstacle outside the FOV of the ultrasound sensors). However, when all three processing schemes are used together (final row in Table II), the system is able to exploit the complementary features of the individual schemes. The localization errors are within the desired limits, and the detection accuracy is above the desired value. The distance errors are higher than those obtained with just the US scheme because the ultrasound sensors can help reduce distance errors only when the obstacle is within its field of view. Furthermore, the proposed approach is better than an ad-hoc information fusion approach, where extensive tuning over a few days results in a distance error of $\approx 14.1 \pm 6.6$cm and detection accuracy of $\approx 85\%$. These results show that the proposed approach better exploits the features of the individual schemes. Next, Table III shows the computation time associated with the

Scheme	Time/frame (msec)
US	33.3
VC	33.3
VS	125 ± 52
US + VC + VS	39.2

TABLE III: Computation time per frame for each processing scheme. With some approximations, the combined scheme can function at close to frame rate.

processing schemes under consideration. The ultrasound-based and color-based processing schemes individually take very little computational effort. When these schemes are executed in conjunction with the existing modules (vision, localization, team coordination etc) the robot can operate at frame rate (30Hz $=$ 33msec/frame). Though our SIFT implementation only processes specific regions of the low resolution images (e.g. only the pixels below the horizon), resulting in processing times that are much smaller than that of the standard SIFT implementation, the approach is computationally expensive. However, if the SIFT-based approach is applied at a reasonable frequency (e.g once every 10-20 frames) and over appropriate image regions, the combined approach can operate at close to frame rate. Kalman filter-based tracking of obstacles provides

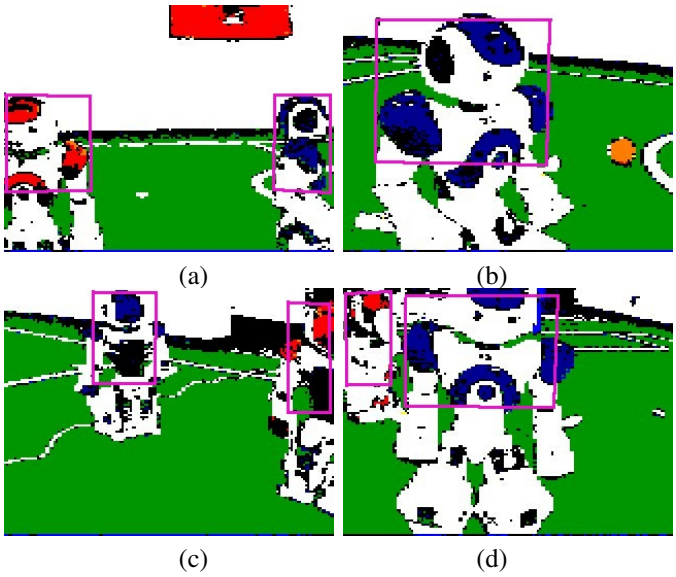

Fig. 3: (a)-(d) Image results: detected obstacles are enveloped in *pink* boxes superimposed on segmented images.

two benefits: (a) it is easy to incorporate the velocity of moving obstacles and account for the relative motion between the robot and the obstacle, and (b) it propagates belief for a few frames after the last sighting of an obstacle, and is robust to some intermittent noisy measurements and mis-classifications. The performance is observed to be statistically similar to that reported in Table II. Some qualitative results of the robot's performance are shown in Figure 3. The detected obstacles are enveloped in rectangular bounding boxes (in *pink* when viewed in color) superimposed on the segmented images.

IV. Related Work

Humanoid robots, and mobile robots in general, are increasingly being equipped with multiple sensors and used in practical applications [3]–[6]. Since the sensors have different capabilities, and the environment changes dynamically, it is essential that all the available information be used effectively.

In recent times, the DARPA challenges [4]–[6] have had robot vehicles equipped with several sensors (lasers, cameras, GPS etc) navigating autonomously in the real-world domains such as deserts and urban roads. However, most of the decision-making was based on input from the range-finders and GPS, and visual input was predominantly used only for close-range detection or obstacle avoidance. In related work, Wellington et al. [14] used lasers and cameras to find the true ground height and hence traversability of vegetation-covered regions. Rankin et al. [16], on the other hand, merged stereo vision and thermal signatures to detect drop-offs at night. Murarka et al. [23] utilized stereo and range information to detect drop-offs on a robot wheelchair. However, these methods fail to model and use the errors associated with each information source—information fusion is typically based on manually specified heuristics.

Sensor fusion has been extensively studied in the field of networks and multiagent systems [24]. Several approaches have also been proposed for specific tasks such as image registration, using established state-estimation methods such as Kalman filters or Bayesian networks [21], [22]. However, most of these strategies use manually-specified heuristics that require supervision when applied to mobile robot domains. Some methods are also computationally expensive.

On humanoid robots, localization and object tracking is accomplished using the same probabilistic state estimation techniques used on other mobile robot domains (e.g. Kalman filters and Monte Carlo methods) [25]. Research in the RoboCup framework [26] and the humanoid robotics community has resulted in several innovative techniques for challenges specific to humanoid robot platforms. For instance, robust techniques have been developed to address the challenges of robot control and balance [12], [13]. Approaches have also been proposed for sensor-based navigation, for instance using stereo-vision [27]. Humanoid robots are also being used extensively for human robot interaction studies [28]. However, as with other mobile robot platforms, there is a need for an efficient strategy to fully utilize the available information to operate reliably in dynamic environments. This paper presents an approach that addresses an instance of this information fusion challenge in the robot soccer scenario.

V. Conclusions and Future Work

Developments in sensor technology have resulted in the deployment of mobile robots in applications such as medicine and autonomous navigation [3], [4]. A major challenge for a robot equipped with multiple sensors, is the ability to efficiently merge the information obtained from each sensor through different processing schemes, in order to operate robustly in dynamic environments.

In this paper we have presented an instance of such multi-sensor information fusion using range and visual information. The robot is able to autonomously learn models that predict the performance of the different schemes that process the visual input (from a color camera) and range input (from ultrasound sensors). The learned models are used in a probabilistic approach that effectively merges the information obtained from the different sources. In the robot soccer domain, we have shown that a humanoid robot is able to detect and localize obstacles significantly more robustly than what could have been accomplished in the absence of such an information fusion scheme.

In multirobot settings (e.g. robot soccer, disaster rescue), information merging can have other advantages. Information communicated by teammates can be merged to obtain robust estimates about areas that are hidden from the robot's field of view, which would prove very useful in surveillance scenarios [9]. In robot soccer, for instance, if the robot knows the global position of one of its teammates (e.g. the teammate communicates its pose with high certainty), relative distance and bearing to this teammate can help the robot localize itself in the global frame of reference. Furthermore, in applications

where the communication is delayed or noisy, the robots in a team may be able to coordinate their actions better using information fusion schemes.

Currently the robot learns its errors models (and object models) in a separate training phase. However, the robot can bootstrap such that the learned models are updated over time in response to environmental changes. Obstacles and other objects that are found to be stationary can even be used as "fixed markers" that enable a robot to localize when the initial set of field markers (e.g. goals) are not visible. Another direction of further research is to use a combination of existing gradient feature detectors such that the robot can detect the desired objects more efficiently and reliably.

This paper shows the feasibility of effectively using the available information for robust performance on a humanoid robot. The long-term goal is to enable robots to autonomously learn environmental models, effectively merge information obtained from different sources, and operate robustly in real-world application domains.

ACKNOWLEDGMENTS

The authors would like to thank Michael Quinlan, Todd Hester and other members of the TT-UT AustinVilla robot soccer team comprising Texas Tech University and The University of Texas at Austin (in the Standard Platform League of RoboCup).

REFERENCES

[1] Hokuyo, "Hokuyo Laser," 2008, http://www.hokuyo-aut.jp/products/urg/urg.htm.

[2] "Videre Design Camera," http://www.videredesign.com/stereo_on_a_chip.htm.

[3] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards Robotic Assistants in Nursing Homes: Challenges and Results," *Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots*, vol. 42, no. 3-4, pp. 271–281, 2003.

[4] S. Thrun, "Stanley: The Robot that Won the DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[5] LAGR, "The DARPA Learning Applied to Ground Robots Challenge," 2005, www.darpa.mil/ipto/programs/lagr/.

[6] DARPA, "The DARPA Urban Robot Challenge," 2007, http://www.darpa.mil/grandchallenge/index.asp.

[7] Nao, "The Aldebaran Nao Robots," 2008, http://www.aldebaran-robotics.com/.

[8] J. Casper and R. R. Murphy, "Human-robot Interactions during the Robot-assisted Urban Search and Rescue Response at the WTC," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 33, no. 3, pp. 367–385, 2003.

[9] M. Ahmadi and P. Stone, "A multi-robot system for continuous area sweeping tasks," in *ICRA*, 2006.

[10] M. Sridharan and P. Stone, "Global Action Selection for Illumination Invariant Color Modeling," in *The IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[11] M. Sridharan, J. Wyatt, and R. Dearden, "HiPPo: Hierarchical POMDPs for Planning Information Processing and Sensing Actions on a Robot," in *International Conference on Automated Planning and Scheduling (ICAPS)*, September 14-18 2008.

[12] J. Rebula, F. Canas, J. Pratt, and A. Goswami, "Learning Capture Points for Humanoid Push Recovery," in *The IEEE International Conference on Humanoid Robots*, Pittsburgh, 2007.

[13] J. Pratt and B. Krupp, "Design of a Bipedal Walking Robot," in *Proceedings of the SPIE, volume 6962*, 2008.

[14] C. Wellington, A. Courville, and A. T. Stentz, "Interacting Markov Random Fields for Simultaneous Terrain Modeling and Obstacle Detection," in *Robotics: Science and Systems*, June 2005.

[15] P. Stone, M. Sridharan, D. Stronger, G. Kuhlmann, N. Kohl, P. Fidelman, and N. K. Jong, "From Pixels to Multi-Robot Decision-Making: A Study in Uncertainty," *Robotics and Autonomous Systems: Special issue on Planning Under Uncertainty in Robotics*, vol. 54, no. 11, pp. 933–943, 2006.

[16] A. Rankin, A. Huertas, and L. Matthies, "Nighttime negative obstacle detection for off-road autonomous navigation," in *SPIE*, 2007.

[17] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "Robocup:The Robot World Cup Initiative," in *ICRA*, February 1997, pp. 340–347.

[18] SPL, "The RoboSoccer Standard Platform League," 2008, http://www.tzi.de/spl/.

[19] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[20] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, March 1960.

[21] R. Brooks and S. Iyengar, *Multi-Sensor Fusion: Fundamentals and Application with Software*. Prentice Hall, 1998.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2008.

[23] A. Murarka, M. Sridharan, and B. Kuipers, "Detecting Obstacles and Drop-offs using Stereo and Motion Cues for Safe Local Motion," in *The IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2008.

[24] L. Panait and S. Luke, "Cooperative Multi-Agent Learning: The State of the Art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, November 2005.

[25] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, USA: MIT Press, 2005.

[26] L. Iocchi, H. Matsubara, A. Weitzenfeld, and C. Zhou, Eds., *RoboCup-2008: Robot Soccer World Cup XII*. Berlin: Springer Verlag, 2009.

[27] J.-S. Gutmann, M. Fukuchi, and M. Fujita, "Real-time path planning for humanoid robot navigation," in *IJCAI*, 2005.

[28] M. A. Goodrich and A. C. Schultz, "Human-Robot Interaction: A Survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2007.