# An Architecture of Sensor Fusion for Spatial Location of Objects in Mobile Robotics

Luciano Oliveira[1], Augusto Costa[1], Leizer Schnitman[1], J. Felippe de Souza[2]

[1] Universidade Federal da Bahia, Brasil
lrebouca@ufba.br , augusto.loureiro@ufba.br , leizer@ufba.br

[2] Universidade da Beira Interior, Portugal,
felippe@ubi.pt

Abstract.
Each part of a mobile robot has particular aspects of its own, which must be integrated in order to successfully conclude a specific task. Among these parts, sensing enables to construct a representation of landmarks of the surroundings with the goal of supplying relevant information for the robot's navigation. The present work describes the architecture of a sensing system based on data fusion from a CMOS camera and distance sensors. The aim of the proposed architecture is the spatial location of objects on a soccer field. An SVM is used for both recognition and object location and the process of fusion is made by means of a fuzzy system, using a TSK model.

## 1 Introduction

The field of mobile robotics has evidenced an enormous potential for research and real experiments. Mobile robots with intelligent behavior are constructed in various parts which by themselves show potential for study [1].

In particular, the sensing is responsible for supplying the robot with the necessary information for the construction of a representation of its surroundings, where the robot is placed, thereby allowing a dynamic description of the obstacles and useful landmarks for the orientation of the robot. Such a task must be made by computational methods whose objective is to reduce the inaccurate nature of the sensors. In order to do this, sensor fusion techniques have been applied successfully, providing a more suitable description of the surroundings due to both redundancy and complementation of data.

To evaluate an artificial vision system and its architecture, it must be sub- mitted to some specific task. For the proposed system, robot soccer is chosen and the Robocup rules for small size robots are used for the local vision system constraints. For a team of robots to participate in a soccer match, various technologies must be present: principles of autonomous agent design, multi-agent collaboration, robotic and sensor fusion, among others. The main objective of the application of the mobile robot in a soccer environment is, therefore, the analysis of multiple areas of knowledge which serves to support socially signifi- cant problems as well as industry.

For the artificial vision system proposed in this work, a CMOS camera and infrared distance sensors are used. An SVM (Support Vector Machine) is applied for the classification of objects by a single color, without any image processing. Another SVM is used for obtaining the polar coordinates of the objects, by regression, from the image; for the sensor fusion, a fuzzy system using the TSK (Tagaki-Sugeno-Kang) zero order model integrates the information of angles of the objects with the data from the distance sensors, in order to refine the information obtained.

In Section 2, aspects of the classifier used are shown. Section 3 describes some architectures for sensor fusion. Section 4 describes the communication between the vision system and the robot. The architecture proposed, as well as its model, are presented in Section 5, while Section 6 presents general results obtained at each stage. To finish, Section 7 presents some conclusions.

## 2 Support Vector Machine

SVM is a hybrid technique of statistical and deterministic approaches. This means that to find the best space for classification hypothesis, a probability dis- tribution is determined from the input space. The technique originated from the work of Vapnik on the Principle of Risk Minimization, in the area of Statistical Learning [2].

The technique is applied in the following way: in the case of linear space, determine the hyperplanes of separation by an optimization problem; in the case of non-linear space, a kernel function is applied and the new space obtained is denominated the feature space – now linearly separable, of a dimension greater than the original.

A Fig. 1 illustrates the application of a kernel in the input space. In the feature space, a hyperplanes is obtained for separation.

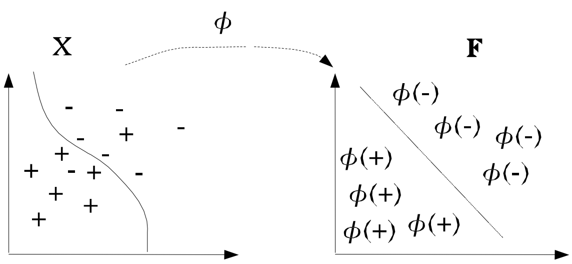



Fig. 1. Mapping of an input space non-linearly separable for a feature space

Another important aspect is the relationship between dimension VC (Vapnik- Chervonenkis), number of support vectors and generalization of the classifica- tion. To determine the dimension VC of a space, it is necessary to also determine the number of support vectors; in this case to find the oriented hyperplanes, within the smaller space of hypothesis defined by the function of risk estima- tion, which determines the limit of classified data. These ratios are determined by the probability of classification error of test data, after finding the support vectors through Eq. (1):

$$E_{\ell}[P(error)] \leq \frac{E[\text{nVS}]}{\ell} \qquad (1)$$

where $E_{\ell}$ denotes the expectation for the input data set size $\ell$ and nVS is the number of support vectors of the SVM. In this way, the fewer the number of SVM support vectors, the greater will be its degree of generalization, obviously respecting a minimum limit, which will be determined empirically.

## 3 Sensor Fusion Architectures

Sensor data fusion represents the process of combining data or information from multiple sensors for the estimation or prediction of states of entities (objects for measurement) [3]. Some objectives for diverse scenarios for application of sensor fusion can be cited: detection of the presence of an object or environmental condition, object recognition, tracking of objects or monitoring of a continuous event, the combining of information to make intelligent decisions, among others. In the area of sensor fusion, the principal organization for regulating terminol- ogy is the Joint Directors of Laboratories (JDL) Data Fusion Workgroup. This organization defines a model of processes for data fusion, by levels of informa- tion, besides a dictionary of terms related to this area [3]. The JDL determined three levels for sensor fusion: data, features and decision.

An interesting work in the area of sensor fusion is presented in [4]. The work is dedicated to the question of temporal sensor fusion and four types are sug- gested: centralized, decentralized, sequential and statistical. Centralized fusion is suitable when the system has only one active sensor. When more than one sensor is present in the system, decentralized fusion (pure, sequential or statistical) is better applied through a timeline.

Another important work on a flexible architecture for sensor fusion is pro- posed in [5]. The author proposes six categories based on three levels of JDL, yet used as processing input/output modes.

## 4 Communication with the Mobile Robot

The architecture used for the mobile robot is illustrated in Fig. 2 and proposed in [6]. The processes which comprise the agent utilize an onboard multi-threaded program.

The Cognitive Level consists of a system based on symbolic knowledge which manipulates both the information received at the Instinctive Level as well as

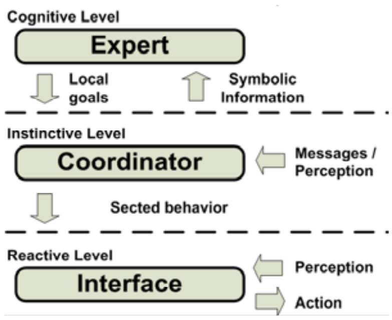



Fig. 2. Architecture for the robot proposed by [6]

asynchronous messages received from other agents, generating the symbolic in- formation for updating of the knowledge base at the Cognitive Level and imple- mented at an instinctive levels.

As the Reactive Level is responsible for processing of messages of perception from the mobile robot, this becomes a point of communication between the agent and the artificial vision system proposed. The Reactive Level is illustrated in Fig.
3 and is composed of fuzzy controllers, input filter, output filter and mailbox. The process responsible for the control of these elements is called the Interface and is orthonormal to Fig. 3.

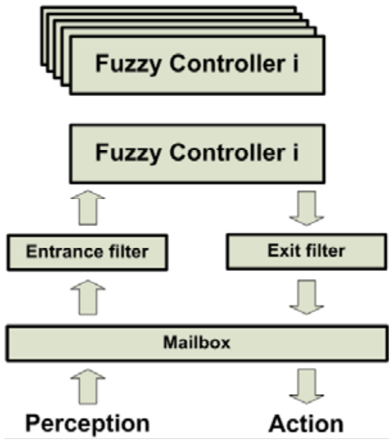



Fig. 3. Reactive level

All messages sent by the artificial vision system are stored in the mailbox. The input filter extracts the value of the linguistic variables from the fuzzy visual information used by the fuzzy controllers, while the output filter checks the fuzzy controller outputs and combines them to define the action for the actuators.

## 5 System Architecture

The system architecture, illustrated in Fig. 4, consists of four principle modules: data acquisition, feature extraction/classification, sensor data fusion and internal representation of object recognition.

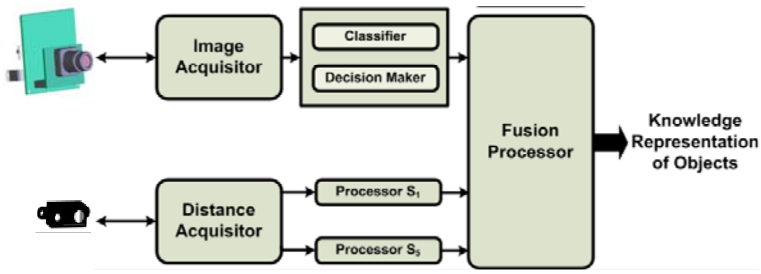



Fig. 4. Vision System Architecture

The goal of the system is to recognize and locate robots and the ball, by theis colors. The Aquisitor modules are responsible for the acquisition of data from the camera and distance sensors. Once an image frame is obtained, this is classified by an SVM. The centroid of each extracted object is calculated and sent to the Decision Maker to obtain information such as angle ($\theta$) relative to the robot and estimated distance ($d$) of each object. At the same time, the acquired data of all the distance sensors is separated and sent to each Processor $S_i$, where $i \in [1, 5]$, and a function of interpolation is applied with the objective of finding the distance from the decimalized value of each of these. Finally, all extracted information is sent to the Fusion Processor in order to generate knowledge representation of each object in the scene. The location of objects recognized in the image is given by its polar coordinates relative to the center of the robot ($\theta, d$).

The processors $S_i$, as well as, the Classifier and Decision Maker are imple- mented in two distinct threads and synchronized in the Fusion Processor, once the fusion is made from the active sensors.

### 5.1 Acquisition

The camera used is the CMUCam2, designed by the Carnegie-Mellon University and comprises a microprocessor SX52 made by Ubicom and a CMOS OV6620 detector.

In order to reduce as much as possible the acquisition time of a camera frame, without reduction in quality of recognition, the obtained image is re-dimensioned to 44x36 pixels. The acquisition time for this frame is 245 ms on average and is determined by the camera's hardware.

The distance sensors used are GP2D02 made by Sharp and use infrared technology and the time for acquisition of a distance value is 75 ms.

### 5.2 Object Recognition

The Classifier module for image pattern and Decision Making is responsible for processing of the output of the image Aquisitor module. The image pattern classification has the objective of identifying moving objects on the soccer field. These objects could be: own team robots, opponent's robots and the ball.

The CMUCam2 offers a color tracking function. The technique used is straight- forward thresholding and consists of tracking determined pixels whose colors belong to the interval $R_{min}$, $G_{min}$, $B_{min}$ and $R_{max}$ , $G_{max}$ , $B_{max}$ , in the RGB space. The disadvantage of this function is the high sensitivity to variation in lu- minosity of the environment. Because of this, an SVM is used with the objective of building a classifier which is more robust to variations in luminosity.

### 5.3 Object Location

From the features extracted from the image (area of the object, height, width, centroid) the location of objects is determined through its polar coordinates ($\theta, d$), where $\theta$ represents the angle and $d$ the distance relative to the center of the robot. To achieve this, two SVM responsible for effecting regression are used: one to determine the angle of the object and map each pixel to its corresponding angle; the other for distance and to map the height of each object in the image and its distance to the robot.

To obtain the pixel-angle function, rays and parallel lines are defined on a blank sheet of paper. The ball is put at the intersection between each ray and line, where the origin of the rays is the center of the robot. Fig. 5 illustrates this situation. A set of pairs (pixel, angle) is then obtained from the centroid of each object and the respective location of the angle relative to the center of the robot is determined.

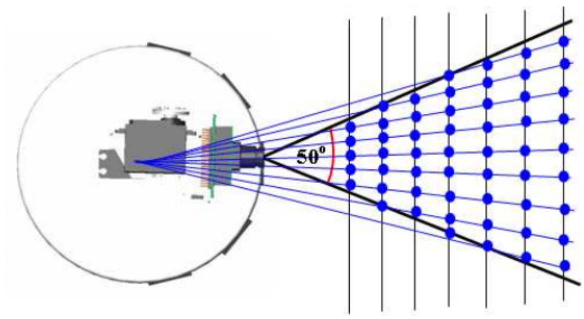



Fig. 5. Determination of the relationship between pixel and angle

The artificial vision head proposed is controlled by two servo-motors and, therefore, has two planes of movement (panoramic movement by rotation on the y axis in the camera system of coordinates and tilt movement by rotation through axis z). In order to reduce the effect of radial distortion of images, occurring when the artificial vision head is at different angles in relation to the z, a value of the vertical axis of the head is added to the pair (pixel, angle) previously obtained. Fig. 6 illustrates the angles used. For each angle (20º, 35º, 50º and 65º) a set of pairs (pixel, angle) is determined and the tuple (pixel, angle, angle of head) is submitted to the classifier.

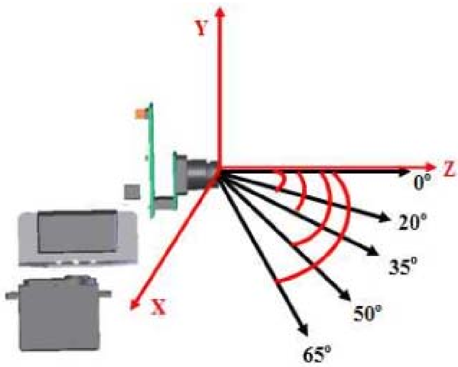



Fig. 6. Angles of the moving head used to determine the pixel–angle mapping function

The distance information extracted from the image is only used as a rough estimate, in case an object is in the shade of the distance sensors. In fact the estimate of distance only functions as a reference for the location of the robot in relation to moveable obstacles. When the robot moves, it can refine the distance information, rotating its body for example, to obtain a more reliable measure- ment.

## 5.4 Distance Sensor Processor

The distance sensor processor is responsible for separating the data obtained by the thread of data acquisition of the sensors and determination of the distance corresponding to the values decimalized from each sensor.

Once the sensor data is obtained, this is then submitted to a polynomial function of interpolation. The useful range of operation of these sensors is 10 to 80 cm. For an interpolation of greater precision, two polynomials are defined in two ranges: one for a range of 10 to 35cm and another for the range 35 to 80cm.

### 5.5 Fusion Processor

Once the values for decision for each sensor (camera and distance sensors) are selected, these are submitted to the Fusion Processor. The sensor data fusion is made by way of two distinct stages: angle and distance selection from the image, and fusion between these values of decision and the values obtained by the distance sensors. As the vision sensor is the slowest of the sensors used and having greater density of space-temporal data by acquisition, this is used as a synchronizer of fusion process.

The timeline of the classes of sensors used is illustrated in Fig. 7. The acqui- sition time of a distance value and conversion to centimeters is approximately 75 ms. This corresponds to approximately four readings of distance in the process- ing of an image and selection of variables of decision in the image (300 ms). In view of this the three first readings of the distance sensors are discarded and only the last represents the distances of the objects recognized. This is illustrated by the dotted lines in Fig. 7.

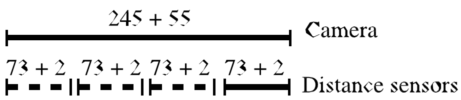



Fig. 7. Timeline of sensors involved in the fusion process (acquisition and processing)

The temporal alignment of the sensors information is necessary, as the sensors used are of the active type. After making this alignment, the values of decision are used for the sensor fusion. In order to do this, the fuzzy sets are determined from the physical disposition of the distance sensors, as per Fig. 8.

In Fig. 8, the three sensors (S3) are used as an illustrative reference of the fuzzy set, used for the fusion processing through the linguistic variable distance in all the sensors. The final distance and angle are determined by a fuzzy system using the TSK zero order model. The reason for this choice is determined by the performance of this system for processing data samplese.

Once the angle is determined by the camera its value can be refined by the distance sensors. The disposal of the distance sensors is known, and for reasons of simplicity, the angles are supplied from the imaginary line (dotted line in Fig.
8) which comes from the origin of the robot and passes through each central point of the distance sensors. It can be seen in Fig. 8, that the further the object is from the distance sensor, the more precise its angle is determined. This is verified in the following way: the spread of each distance sensor is made by way of a cylinder 3 cm wide (space between receptor and emitter) an object between

$A_1$ and $A_1$ will have an angle less precise than an object between $A_2$ and $A_2$. Therefore, from the geometry of the sensors, the fuzzy sets are obtained
illustrated in Fig. 9. 0 0

From these fuzzy sets, eight rules are proposed:

$R_1$: IF distance = SHADOW THEN distance = d(camera)

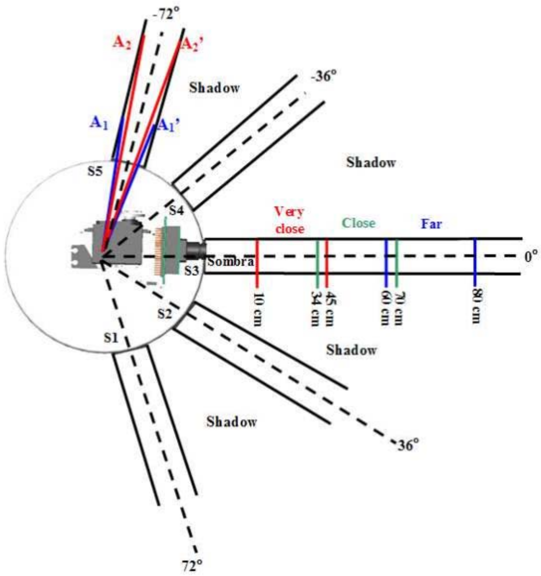



Fig. 8. Determination of fuzzy sets from distance sensors

$R_2$: IF distance = SHADOW THEN angle = a(camera)
$R_3$: IF distance = VERYCLOSE THEN distance = d($S_i$) $R_4$: IF distance = VERYCLOSE THEN angle = a(camera) $R_5$: IF distance = CLOSE THEN distance = d($S_i$)
$R_6$: IF distance = CLOSE THEN angle = a(camera)*0,5 + a($S_i$)*0,5
$R_7$: IF distance = FAR THEN distance = d($S_i$)
$R_8$: IF distance = FAR THEN angle = a($S_i$)

The functions $a(.)$ e $d(.)$ represent respectively the angle and distance ob- tained by the camera and distance sensors. The real values of distances and angles, after evaluation of the rules, are determined through Eq. (2).

$$S = \frac{\sum \psi_i z_i}{\sum \psi_i} \quad (2)$$

where $\psi_i$ is the T-norm of each antecedent and $z_i$ is the result of function $f(x, y)$, responsible for describing the relationship between the fuzzy sets of the an- tecedent.

At the end of the fusion process, each object identified and located is repre- sented internally by the following structure. For each image frame $h$numFrame$i$, all objects located identified with an $h$idObject$i$ and with three characteristics supplied: angle relative to the center of the base of the robot $h$angleObject$i$, distance of object to the front of the robot $h$distanceObject$i$ and fuzzy velocity of each object $h$fuzzyVel$i$.

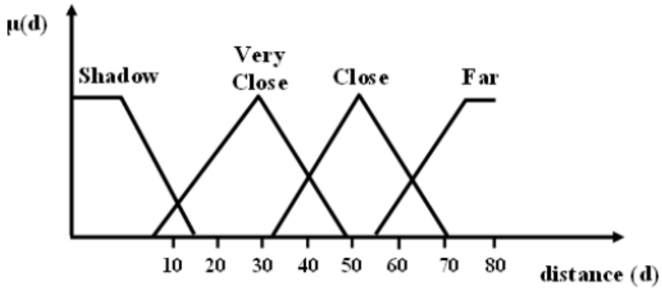



Fig. 9. Fuzzy sets for the linguistic variable distance

( frame hnumFrame*i* ( see htimePerception*i* )
( object name hidObject*i* )
( angle hangleObject*i* )
( deist hdistanceObject*i* )
( vel hfuzzyVel*i* ) )

The fuzzy velocity hfuzzyVel*i* is determined by Eq. (3). Fig. 10 shows the fuzzy sets used.

$$hfuzzyVel\boldsymbol{i} = [\mu_l(difP), \mu_m(difP), \mu_h(difP)] \qquad (3)$$

where $\mu_i(difP)$ are membership functions, and $i$ represents each of fuzzy sets of linguistic variable velocity (low, medium and high). $difP$ is the difference between centroid location of an object in relation of frames $n$ and $n-1$.

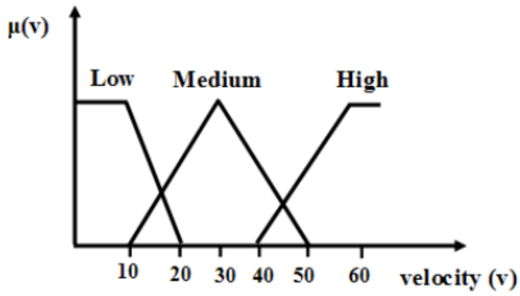



Fig. 10. Fuzzy sets for linguistic variable velocity

## 6 Results Analysis

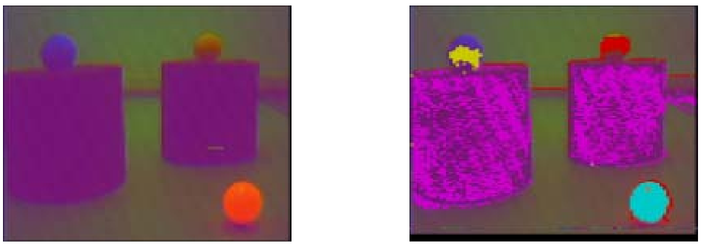
Three parameters are analyzed with the objective of evaluating the general per- formance of the system: for object recognition, the application of classifier in different ranges of lights; for the location of objects recognized, the square error and standard deviation of angles found.

For evaluation of the classifier, Table 1 illustrates the rate of precision for different luminosities. The range of luminosity tested was between 570 and 980 lux and is in accordance with the Robocup rules for small robots competition, which define an illumination of between 700 and 1000 lux, for each tournament.

Table 1. Results for different ranges of light

| Illumination (lux) | Precision |
|---|---|
| 570 | 87.75% |
| 660 | 84.08% |
| 780 | 84.87% |
| 800 | 86.36% |
| 920 | 87.72% |
| 980 | 90.00% |

Figs. 11(a) and 11(b) illustrate the results of classification of robots and ball image.



(a) Image obtained (b) Objects classified

Fig. 11. Classification results

For evaluation of determination method of angle measurement of an object relative to the robot, the standard deviation between the measured and real angle for each position chosen for the vision head is calculated. The values for standard deviation of six random measures (determined and real) are illustrated in Table 2.

The parameters analyzed previously point, therefore, to a robustness of recog- nition of an object and efficiency in their location. Through the fusion process, the values obtained can be refined at each instance of robot actuation.

Table 2. Standard deviation

| Head angle | Standard deviation |
|---|---|
| 20 | 2,34 |
| 35 | 2,15 |
| 50 | 2,00 |
| 65 | 2,18 |

## 7 Conclusions

A system of artificial vision applied in a mobile robot, based on sensor fusion, is here presented. The SVM technique is applied to either recognize patterns and to extract the decision attributes for objects in the image. For the sensor fusion, fuzzy logic is used.

A critical point of the system is the acquisition time of an image frame and it must be evaluated and changed through the use of dedicated circuits and a faster image processor.

The contribution of this work is: a system with modular architecture, making easier to repeatedly use modules in the system and also their inclusion; robust pattern recognition with luminosity variation and only one training sample and an efficient time classification through an SVM applied to the color space YCrCb; an adaptive system of information from sensors using Fuzzy Logic applied in the Fusion Processor.

Finally, one can state that the results in general are pointing to a good performance of the system.

## References


1. D. Kortenkamp, R. Bonasso, and R. Murphy. *Artificial Intelligence and Mobile Robots*. Menlo Park, 1998.
2. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.
3. D. Hall and J. Llinas. *Handbook of multisensor data fusion*. 2001.
4. S. Bruder, M. Faronq, and M. Bayoumi. Robotic heterogenous multi-sensor fusion with spatial and temporal alignment. In *Proceedings of the 30th IEEE Conference on Decision and Controle*, pages 506–511. IEEE, 1991.
5. B. Dasarathy. Sensor fusion potential exploitation – innovative and illustrative applications. In *Proceedings of the IEEE Special Issue on Sensor Fusion*, volume 85, pages 24–38. IEEE, 1997.
6. A. Costa and G. Bittencourt. From a concurrent architecture to a concurrent autonomous agentsarchitecture. In *IJCAI'99, Third International Workshop in RoboCup*, pages 85–90, agosto 1999. Springer, Lecture Notes in Artificial Inteligence.