30th International
Symposium on Automotive Technology & Automation

Florence, Italy 16th - 19th June '97

Robotics, Motion and Machine Vision in the Automotive Industries
Co-sponsored by CRIIF, France

Programme Committee Chair:-
Dr Reinhold C Mann, Head Intelligent Systems Section, Oak Ridge National Laboratory, USA
Co-Chairs:- Professor A
Goldenberg, University of Toronto,
Department of Mechanical Engineering, Robotics and Automation Laboratory, Canada, Dr Franz May,
Image Understanding Systems Department, Daimler-Benz AG,
Germany and Mr Y Ishiguro, Development Department, Toyota Motor Corporation, Japan

The Largest European Automotive Forum with International Participation

Over 700 high quality refereed papers in 15 simultaneous dedicated conferences
AVOIDING OBSTACLES WITH MOBILE ROBOTS - INTEGRATING VISION AND SONAR DATA IN CONNECTIONIST APPROACH

Mr A Silva, Dr P Menezes and Professor A Dias,
Universidade de Coimbra,
Portugal

97RO061

Abstract

To navigate a mobile robot in a cluttered environment, where obstacle presence and location are unknown, several sensing modalities can be cooperatively used. The aim of our work is the choosing of relevant sensing devices for this task, including the development of an integration scheme where sensing modalities can be mutually enhanced and validated. We also propose the implementation of some innovative sensing and control strategies, using the integrated sensorial information.

In our on-going work, visual data obtained by a binocular active vision system is integrated, together with ultrasonic range measurements, in the development of a obstacle detection and avoidance system based on a connectionist approach. In this system, two connectionist structures, a feed-forward artificial neural network and a cellular automata, are used to build a map of the environment surrounding the robot. This structures are used to navigate through this environment, avoiding collisions with unexpected obstacles. Since no a priori knowledge of the world is available, we rely on sensory data only in order to build a valid world representation.

All our experimental work is being done around a mobile platform (Figure 1), which provides us with the necessary testbed to our experiments. This mobile platform is also the support for the sensing devices described throughout this article.

* This work was partially financed by NATO Scientific Affairs Division on the framework of the NATO Science for Stability Program, PO-ROBOT project.
* Working under grant PRAXIS XXI BM/7255/95 from JNICT - Junta Nacional de Investigação Científica e Tecnologia.
1. Sensing Modalities

Besides odometry, two main sensing modalities are used in the process of environment mapping. These are ultrasonic range data and information acquired through binocular active vision.

1.1. Ultrasonic Sensors Range Measurements

The range measurements are obtained from a ring of 24 ultrasonic sensors present in our mobile platform. This kind of sensors, being cheap and easy to use, frequently present erroneous data, caused by multiple reflections and sonar scattering. However this information can be validated, in one hand, through the accumulation of readings, and, in the other hand, by including other modality of sensorial information [4].

For the ultrasonic sensors a simple model is used: a reading is represented by a cone where the arc points have the same probability of belonging to an obstacle, being that probability inversely proportional to the distance measure. A point inside the cone is considered to have a negative probability of belonging to an obstacle and this probability is proportional to the distance between the point and the cone's arc. An illustration of this principle is summarized in Figure 2.

![Figure 2: Model of an ultrasonic sensor reading.](image)

Using this technic we can speed up the update of this grid-based representation with very good experimental results and avoiding the usual gaussian representations - see Figure 1.b.

1.2. Obstacle Detection by Binocular Disparity

Vision is a powerful sensing modality which can be used, not only in low level tasks, like the ones of obstacle detection and avoidance presented here, but also in higher level problems, e.g. object identification and localization. In this work we are using stereoscopic vision as a mean of complement and validate data obtained by the ultrasonic sensors. To do this, we developed a simple algorithm for obstacle detection, using stereoscopic information. Its simplicity makes it useful for mobile robotics, where the need for real time implementation asks for algorithms with moderated computational demands.

The solution found allows us to detect obstacles in the field of view of the stereoscopic setup, both above and below the robot's motion plane. The algorithm follows this simple principle: if we have a stereo pair of images of the ground plane, in front of the robot, where Z is constant, and we can establish an affine transformation which relates each point in one image, with the correspondent point, in the other image. For any image pair with points not in the ground plane, the transformation will fail. This means that any obstacle in the field of view, above or below the ground plane, will result in a disparity region when the two images are matched using the transformation mentioned above. When the robot is moving, any relevant disparities between image pairs is identified as an obstacle.

The mathematics underlying this principle are very simple. A real world point with (X, Y, Z) coordinates is related to its image representation, with coordinates (u, v), by a 3×4 transformation matrix as expressed in (1).

\[
\begin{pmatrix}
X \\
Y \\
Z \\
1 
\end{pmatrix} =
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24} \\
a_{31} & a_{32} & a_{33} & a_{34} \\
0 & 0 & 0 & 1 
\end{bmatrix}
\begin{pmatrix}
u \\
v \\
s \\
1 
\end{pmatrix} 
\]

(1)

From our particular setup we have that Z=γ, since we want all points to lie in the ground plane. This way (1) can be simplified into (2).

\[
\begin{pmatrix}
u \\
v \\
s \\
1 
\end{pmatrix} = A_{3×4}
\begin{pmatrix}
x \\
y \\
z \\
1 
\end{pmatrix} 
\]

(2)

Since we have two images from the same scene, one from the right camera and another from the left camera, and we can write:

\[
A_{3×4}^{-1}
\begin{pmatrix}
u_x \\
v_x \\
s_x \\
1 
\end{pmatrix} = A_{3×4}^{-1}
\begin{pmatrix}
u_y \\
v_y \\
s_y \\
1 
\end{pmatrix} = \begin{pmatrix}
x \\
y \\
z \\
1 
\end{pmatrix} 
\]

(3)
From (3) we easily obtain the relation between matching points in the left and right image:

$$\begin{bmatrix} s_l u_l \\ s_l v_l \end{bmatrix} = A_l A_r^t \begin{bmatrix} s_r u_r \\ s_r v_r \end{bmatrix} = F_{3,3} \begin{bmatrix} s_l \gamma \\ s_l \gamma \end{bmatrix} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{bmatrix} \begin{bmatrix} s_l \gamma \\ s_l \gamma \end{bmatrix}$$  \hspace{1cm} (4)

Manipulating this equations and making $\alpha_{3,3} = 0$ we can easily obtain (5):

$$\begin{bmatrix} u_l \\ v_l \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -u_l u_r & -v_l v_r \\ u_l u_r & v_l v_r \\ -u_l v_r & v_l u_r \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} u_r \\ v_r \end{bmatrix}$$  \hspace{1cm} (5)

This relation is a set of linear equations relating the parameters $\alpha_{ij}$ with the coordinates of correspondent points in the left and right images. Since we have more unknowns than equations, we need a set of correspondent point pairs, at least 8, in order to make possible the computation of all $\alpha_{ij}$.

If we admit that there is some amount of noise in point coordinates, to enable robust results, it is advisable the use of a larger set of points and a minimization criterion, such as the least-squares criterion. (5) can then be solved by using the Single Value Decomposition (SVD) method.

A hard problem to solve in the approach described here is the noise introduced in several stages along the process (e.g. image acquisition, correspondence errors, etc). Some mechanisms were implemented to remove most of the noise while trying to avoid turning the process too computational intensive for real time implementation.

Sample results obtained using the described method are presented in Figure 3.

---

To make the obtained results useful for integration with the ultrasonic readings, a further step must be taken: a simple calibration is done in the $Z=\gamma$ plane, relating points with image coordinates $(u,v)$ with points with $(X,Y,\gamma)$ coordinates in the real world. Again SVD is used to guarantee the calibration robustness. After this calibration step, we can compute the distance from the robot to a detected obstacle by finding the coordinates in the $Z=\gamma$ plane of image points where disparity starts to be significant. This range information can then be easily integrated with data coming from the ultrasonic sensors in an uniform representation.

2. Visual and Ultrasonic Sensor Data Integration

The use of several sensing modalities takes us to the problem of integrating the various data flows in an unified world representation. In our approach the integration is supported by a cellular automata. This can be pictured as a cellular lattice, where each cell is connected to its nearest neighbors (Figure 4). The cells' values are updated iteratively, and depend only on the values of each cell and its neighbors, and the connections between them, accordingly with the automata dynamic rules.

![Figure 4: Connections between cells in our cellular automata.](image)

In our representation, each cell maps a $\delta$ side square of the environment and its value is proportional to the certainty of the presence of an obstacle there. This certainty value is the result of data incoming from the sensing devices and from the dynamics of the cellular automata itself. This will lead to the following dynamic rules for our cellular automata:
3. Controlling the Robot

A three layer feed-forward artificial neural network has been trained to navigate the robot, as information about the environment is being gathered and processed in the cellular automata, using that information to safely avoid obstacles and arrive at preset goals [1].

This neural network type has been often used [8], [9], [7] in mobile robotics to enable the robots to reproduce simple, but intelligent, behaviours. We used it to teach the robot simple heuristics, enabling it to navigate through its environment avoiding the obstacles in his way. To do this, the network inputs were fed with information collected in the environmental model being created by integration of incoming sensorial flows, and its output was a free direction that would allow the robot to get nearer to its goal. The information fed to the network inputs, emulated a ring of virtual sensors around the robot. When the controller asks for sensorial information, this virtual sensors obtain their range readings from the world model being constructed, thus being much more reliable than real sensors (e.g. just using the ultrasonic sensors).

This simple controller is being used with success, in cooperation with the model building system, to navigate the robot in rooms with unknown structure. In Figure 6 we can see the environmental model in construction while the robot is being navigated by the neural controller.

4. Conclusions and Future Work

The main conclusion of our work is that different sources of sensorial information can be integrated in a common structure to allow the robust modeling of the environment in autonomous robot navigation. Experiments with an isolated source of information shown us that most of the encountered problems can be overcome by the inclusion of more sensing devices.

In our particular setup it was easy to conclude that the use of ultrasonic ranging information alone was not reliable (e.g. in very short distances) in result of the sensors performances. Including vision we believe that robustness and flexibility were added, while validation between different sources of information became available. In the other hand, ultrasonic information is very fast to acquire, while image processing is always problematic to implement in real time. This way, with the use of ultrasonic sensors together with binocular active vision we can be less demanding in processing power to deal with stereo image pairs.

Other side of our work which is worth mentioning is the implementation of both the integration of sensorial information and the control mechanism using connectionist structures. Neural networks...
and cellular automata both rely on the same principle of many simple processing units densely interconnected. Implementation of this kind of technics can allow the fast processing of large maps and the development of higher level behaviors.

At this stage our prospects of future work are mainly focused on the exploration of the richness of possibilities given by the use of binocular active vision in mobile robotics [2], [3], [5]. Our initial setup only gives low level obstacle detection information, but we hope to expand the range of information obtained from the flow of stereo image pairs (e.g. 3D info and optical flow).

In the near future we propose to extend the 2-D model to a 3-D one, again using the possibilities of stereo vision. Acquiring higher level information than simple obstacle presence from the model (e.g. obstacle structure) is another issue being considered.

References


