

Active Stereo Vision for Object Recognition and Cognitive Map Formation in a Virtual World – A Demonstration

Ilkay Ulusoy, Ugur Halici, Kemal Leblebicioglu
Computer Vision and Intelligent Systems Research Lab.
Middle East Technical University, Ankara, Turkey
Ph: +90-312-210 4558
Fax: +90-312-210 1261
{[ilkay, halici](mailto:ilkay.halici@metu.edu.tr)} @metu.edu.tr
<http://vision1.eee.metu.edu.tr/~halici/>

In very few mobile robotic applications stereovision based mapping and navigation is used because dealing with stereo images is very hard and time consuming. Despite all the problems, stereovision still becomes one of the most important resources of knowing the world for a mobile robot because imaging provides much more information of the world than most other sensors.

Real robotic applications are very complicated because besides the problems of finding how the robot should behave to complete the task at hand, the problems faced while controlling the robot's internal parameters bring high computational load. Thus, first working in a simulated environment in order to find the strategy to be followed by the robot and then applying this on real robotic applications is preferable.

In this study we describe a demonstration for object recognition and cognitive map formation using only stereo image data in a 3D virtual world where 3D objects and a robot with stereo imaging system are simulated. Stereo imaging system is simulated so that it has the actual human visual system properties. The objects are various trees and cottages composed of two parts with different texture, color, shape and size. For example a cylindrical body with wood texture and a conical top with leave texture is for a pine tree, similarly a cylindrical body and a spherical top is for an apple tree.

Only the stereo images obtained from this world are supplied to the virtual robot (agent). By applying our disparity algorithm on stereo image pairs, depth map for the current view is obtained. A cognitive map of the environment is updated gradually with the depth information extracted while the virtual agent representing the robot is exploring the environment. The agent explores its environment in an intelligent way using the current view and environmental map information obtained up to date. Also, during

exploration if a new object is observed, using its view from different directions, it is labelled with its shape such as a big size pine tree, a medium size apple tree, a small size house etc.

The Virtual Environment

The simulation software of the virtual environment is developed using C++ programming language and OpenGL graphics library [1]. As the user interface of the software is depicted in

Figure 1, it is composed of four panes. In the virtual world shown in this figure, there are trees of different sizes and types and a cylindrical shaped cottage among them. The left and right panes above are views from the left camera (eye) and right camera (eye) respectively. These stereo images are used in forming the map and recognizing the objects. The bottom panes render the scene from the top viewpoint, and front viewpoint. These top and front view information are not used in any algorithm but placed here just for visualization purposes.

The software allows specifying and rendering 3D arbitrary geometric shapes. The shapes are defined in a text file in a pre-defined format. The software supports a plug-in mechanism to be loaded and linked dynamically into the application at run-time. The plug-in accesses internal structure of the scene data and images rendered, so that it controls the navigation in the scene by locating the camera at any point in 3D and setting camera parameters. Also, gaze direction, focal length, field of view, base distance between the cameras are the main camera parameters which could be modified whenever desired via plug-in.

The plug-in also exports camera views. In this study, the stereo images obtained via the plug-in are processed for depth perception. After processing of the images according to the purposes, the camera location and parameters for the next image rendering

are updated in an intelligent way for the purpose of cognitive map construction and object recognition.

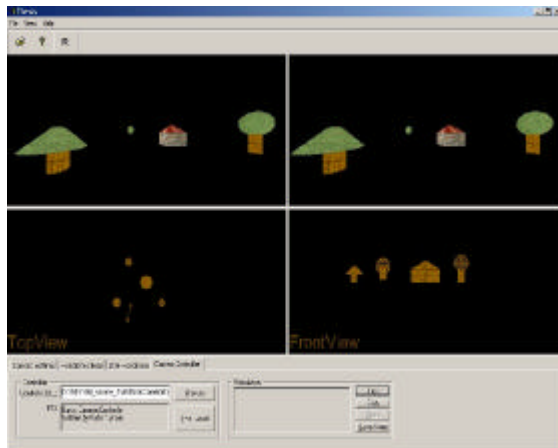


Figure 1 Screen shot from the software

Active Stereo Vision for Cognitive Map Construction and Object Recognition

In this study, we aim to construct a 3D, world-centric, grid-based cognitive map because of its simplicity and applicability. Each grid codes a 1x1x1 unit cube of world unit.

While the agent explores its environment in an intelligent way, a cognitive map is constructed. The agent is aware of its camera parameters and viewing direction, and knows its position in the environment. After each movement of the agent in the environment, stereo images are obtained and processed in order to extract depth information. For this purpose, feature points, which are high contrast edges, are found on the images using steerable filters as been done in [2]. Corresponding pairs are found by matching the feature points which have similar orientation, phase, and magnitude [3]. After performing the stereo triangulation, depth for the current view is extracted. Knowing the camera parameters and locations, the cognitive map is filled with the current depth information.

During the exploration, if an object which has not been seen before is in the current view, the agent goes around that object and extracts feature points of it in 3D. After segmenting the parts of the object using color and texture information, each part of the object is classified as sphere, cone or cylinder using the general quadratic equation of three parameters. For example in Figure 2, a spherical tree-top is shown. After classifying the object parts, the object is recognized. For example a spherical top with a cylindrical body is an apple tree.

If other new objects are seen while the agent goes around an object, the agent goes to the closest of the newly observed objects directly after finishing the recognition of the currently observed object. If nothing new is seen then the agent navigates to the most unexplored direction.

In the end a 3D map of the environment is obtained where almost all of the region is explored. If 3D map is viewed from the top, occupancy is obtained as in Figure 3.

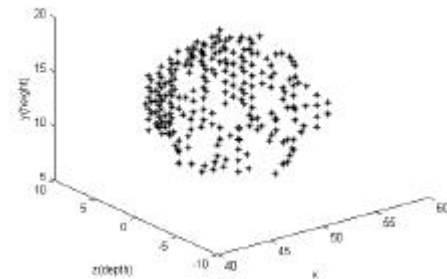


Figure 3 Feature point locations for an apple tree top (i.e. sphere) in the 3D cognitive map.

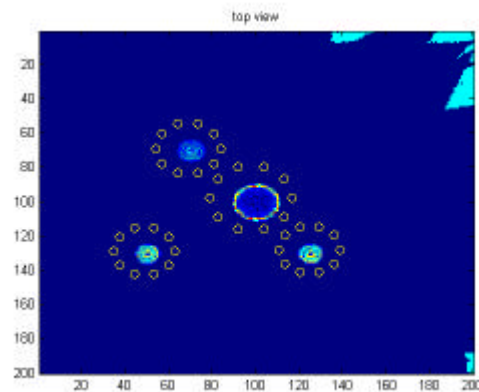


Figure 4 The occupancy (i.e. top view) of the environment obtained from 3D cognitive map by summing the grids in y axis. Circles show the cameralocations while turning around the objects.

References

- [1]. Fahri Tunçer, Image Synthesis for Depth Estimation Using Stereo and Focus Analysis. *M.Sc. Thesis, Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, 2002.*
- [2]. Ali Erol, Automatic Fingerprint Recognition, *Ph. D. Thesis, Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, 2001.*
- [3]. Ilkay Ulusoy, Edwin R. Hancock, Ugur Halici, Disparity Using Feature Points in Multi Scale. *SSPR/SPR 2002: 320-328.*