Map Building and Localization of a Robot Using Omnidirectional Image Sequences

Zoltán Vámossy

Institute of Software Technology
John von Neumann Faculty of Informatics
Budapest Tech
Bécsi út 96/B, H-1034 Budapest, Hungary
vamossy.zoltan@nik.bmf.hu

Abstract: The paper describes a map building module, where the image sequences of the omnidirectional camera are transformed into virtual top-view ones and melted into the global dynamic map. After learning the environment from training images, a current image is compared to the training set by appearance-based matching. Appropriate classification strategies yield an estimate of the robot's current position.

Keywords: PAL (Panoramic Annular Lens), omnidirectional vision, appearance-based navigation

1 Introduction

The goal of this project is to create a wheeled robot equipped with a panoramic annual lens (PAL)-optics, which is capable of autonomous navigation and collision avoidance within a weakly textured environment. The long-term goal of this project is the ability of autonomous mapping of the environment; finding, and navigating through user-specified path; and searching for a predefined object within an unknown environment. The digitalized picture of a camera serves as base information, which is filtered with different image processing algorithms. One of the elements of the investigation is the analysis of these basic algorithms and their collective effect. Another utilization of the camera image is to scan the position of objects in the robot’s environment, and the mechanism of the map-building. In the following, the article summarizes the theoretical background, the main components, the applied techniques and the results of the system.

2 Omnidirectional Imaging [1]

Imaging in general means the effort to portray the three-dimensionality of space conveyed by signal bearing waves on an Euclidean flat surface. Omnidirectional imaging shows ‘panoramic vision’, which means that there is pictorial representation encircling the spectator having real objects as a foreground.

The problem, however, is how to achieve that appearance of things at a given place and time conceived will be recorded and displayed as something actually existing at the place and time, even when no perceiver is present and a constituent of the object whose appearance it is, i.e. a scenic picture giving an effect of extension of the vista i.e. three dimensionality results. With other words, the problem of imaging and displaying is connected to the means how to obtain the optical imprint of the three-dimensional world in such a way that the cortex space in our mind matches the physical space, which is mostly experienced by touch, muscle tension and movement.

In general, our imaging strategy assumes that the person (or imaging system) stands facing the 3D environment on a level plane and looks through the picture plane – represented as an upright transparent surface – at a space chunk, which contains the natural horizon (the distant line where the earth or sea apparently meets the sky) where parallel lines are going to merge in a point, the vanishing point.

The emphasis is on ‘space chunk’, since the metric relation of a sphere cannot be mapped onto the metric relations of a plane, and, as a result, the sphere of vision can be perceived only in discrete chunks, called visual field.

It follows from this so-called See-Through-Window (STW) imaging strategy that our visual apparatus, and any imaging system based on its analogy, is not capable of perceiving and/or recording 360° panoramic view of space at once, i.e., it cannot render a ‘pictorial representation of space and time data encircling the spectator and often having real objects as a foreground’.

2.1 Omnidirectional Imaging is the Result of Centric Minded Thinking

If one assumes that the geometric structure of space encircling us is cylindrical, rather than spherical, centric minded imaging (CMI) can be established. This way of thinking may be backed up by the observation that the visual signal processing of Mother Nature seems to operate according to a similar philosophy. The presence of vertical parallax is less important for us than the horizontal one, further, stereopsis exists only horizontally, and, therefore, it is more appropriate to speak of cylinder of vision, instead of sphere of vision.

If now it is assumed that the radius of the circumscribed cylinder is equal to the vision distance, a panoramic view of the image volume shows up on the wall of this imaginary cylinder. However, the result of this course of thoughts is only a 360° panoramic view, but not an omnidirectional panoramic image in the sense of image definition, since it is not an intensity pattern displayed on an Euclidean flat surface yet.

It can be shown that by using special stretching maneuvers one can transform this panoramic view image projection onto a plane surface perpendicular to the axis of the imaginary cylinder. As a result, a panoramic annular image of the three-dimensional environment is formed, where points in the cylindrical space seen at constant field angles perpendicular to the axis of the cylinder of vision are located on concentric rings in the image plane. The geometric relation of the three-dimensional environment remains represented in polar coordinates and provides an image in which the points retain the same 1:1 relation to each others as in reality. This allows a distortion-free omnidirectional display of the imaged scene.

Analyzing further the annular image one can find that this optical imprint displays the 2D skeleton of the encircling 3D environment in such a way that one may get data on the place and time position of object points, since the width of the ring shaped image corresponds to the viewing angle in the direction of the axis of the cylinder of vision.

From the described stretching and transformation maneuvers it follows that the technical realization of such an optical system must be of catadioptric type.

Several patents have been filed all around the world, claiming better performing optical systems for CMI. All these endeavors can be classified into two main groups: either they are based on multiplexed element design using several optical elements such as lenses and/or cones and/or prisms and/or mirrors with coinciding optical axes, or the others use a single glass block with sophisticatedly shaped refracting and reflecting surfaces.

2.2 PAL, the Omnidirectional Imaging Block

The omnidirectional imaging block PAL (Panoramic Annular Lens) consists of a single glass block with reflective and/or refractive plane, concave and convex surfaces. This means that already in the simplest case the number of possible shapes of the imaging block amounts to the number of the iterative variations of the fourth class which can be formed of three elements, i.e., to 81. However, it has to be emphasized that only a few of them produce good quality images. A well designed PAL-optic

1) is almost afocal, and both a virtual and a real image are formed inside the optic;

2) it renders, via a relay lens, a sharp image from right up against the lens surface out to infinity;

3) its center region around the optical axis does not take part in the forming of the panoramic annular image; it serves only to ensure undisturbed passing through of the image forming rays;

4) objects to the front of the optic are imaged to the interior of the annular image, and objects to the rear of the optic appear on the outer rim of the annular image;

5) a collimated light beam entering the PAL-optic through its plane surface, after passing through the lens, leaves it in form of a light cylinder that evenly illuminates the surrounding. The height of the light cylinder at a given distance from the optical axis depends upon the refractive index of the material the CMI block has been made of.

The interpretation of the resulting panoramic images may cause some confusion, since we are not accustomed to see our three-dimensional world in front and behind us simultaneously, only in discrete chunks successively. Since relations can be established between the polar coordinates of this centric minded imaging (CMI) and the rectangular coordinates of the STW image plane, the annular images can be displayed in Cartesian coordinates. Using appropriate software, the ring shaped image can be ‘straightened out’, i.e., the 360° panoramic image displayed in polar coordinates can be converted into Cartesian coordinates, and the mentioned discomfort immediately disappears.

3 Experimental Mobile Robot System

A remote controlled Model RC is used as the base of the mobile robot [2]. The Model RC is capable of precision controlling: both the direction and the speed can be set. The range of the remote control is about 30-40 meters and the maximal speed is about 20 km/hours.

The PC is connected to the remote controller with an extra electronics, and then controls both controller transistors with three impulses. The impulses are generated by an 18F1320 PIC micro controller. An interrupt is generated in the PIC program with a timer every 15ms. On every interrupt, the PC is sending three signals: the first signal sets the beginning of the periodical time, the second impulse controls the direction of the car; the third control-data sets the speed of the car.

The wireless camera with PAL optics is mounted on the top of the mobile car (Figure 1), looking at the floor; therefore, it is capable of observing the immediate environment only. The result of the camera is relayed using a TV tuner to the main PC program, which controls the robot automatically using the images gained from the camera as its only input.


Figure 1
The mounted PAL on the Model RC

4 Map Building from the Image Sequences

The image flow arrives from the input module, which is responsible for either capturing images from a camera, or play back a test video file. It forwards the images to both the decision maker, and the map builder module. The decision maker analyses the image, and sends a direction/speed signal to the navigation module, which, in turn, forwards it to the controller PIC.

4.1 Image Preprocessing

In order to make a valid control decision, the image is preprocessed by three filters:

1) Using a HSL filter, the program segments the image to Hue, Saturation, and Luminance components. The Hue component is between 0, and 360, the Saturation, and the Luminance will fall between 0, and 100. The HSL filter is used in two experiments: in line following mode, when the predefined track is homogeneous, or in object-following mode when the object to be followed is significantly differing in color from the rest of the environment.

2) The RGB filter is almost as efficient as the HSL, but the algorithm is significantly faster. Using this filter, the three color channels is analyzed using a minimum and a maximum values; if the color of the pixel is within these values, then it remains intact on the resulting image; otherwise the filter will make it black.

3) Using the threshold filter, image binarization can achieved; the result image will contain only the pixels needed for navigation.

4.2 Feature Tracking and Selection

Let I and J be two 2D gray scaled images [2]. The image I will be referenced as the first image, and the image J as the second image. Consider an image point u on the first image. The goal of feature tracking is to find the location v on the second image J such as the intensities are 'similar'. The vector d describes the relative motion is the image velocity at the point, also known as the optical flow. It is essential to define the notion of similarity in a 2D neighborhood sense. During the optical flow calculation the problem is to define the velocity d as being the vector that minimizes the summed squared differences of the intensities in I and J for a given integration window. To provide solution to that problem, the pyramidal implementation of the Lucas–Kanade algorithm is used. An iterative implementation of the Lucas-Kanade optical flow computation provides sufficient local tracking accuracy.

So far the tracking procedure is described that takes care of following a point u on an image I to another location v on another image J. However it is not described what means to select the point u on I in the first place. This step is called feature selection. It is very intuitive to approach the problem of feature selection once the mathematical ground for tracking is led out. At that step, the spatial gradient matrix is required to be invertible, or in other words, the minimum eigenvalue must be large enough (larger than a threshold). This characterizes pixels that are easy to track. Therefore, the process of selection goes as follows:

1) Compute the matrix and its minimum eigenvalue λ at every pixel in the image I.

2) Call λmax the maximum value of λ over the whole image.

3) Retain the image pixels that have a λ value larger than a percentage of λmax. This percentage can be 10% or 5%.

4) From those pixels, retain the local maximum pixels (a pixel is kept if its λ value is larger than that of any other pixel in its neighborhood).

5) Keep the subset of those pixels so that the minimum distance between any pair of pixels is larger than a given threshold distance (e.g. 10 or 5 pixels).

After that process, the remaining pixels are typically 'good to track'. They are the selected feature points that are fed to the tracker.

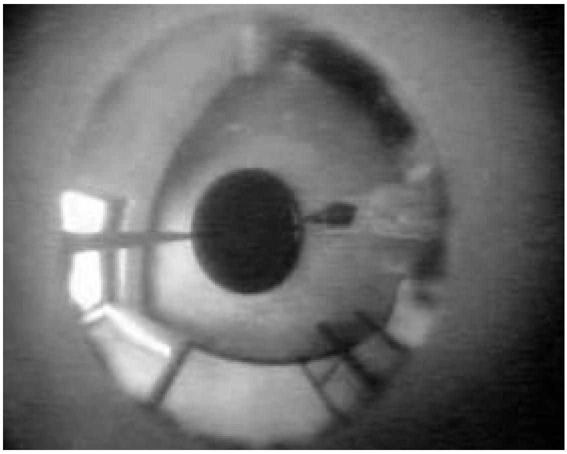

Figure 2
The PAL image

4.3 Map Building

The map builder module prepares the user-defined navigation: after the robot builds up the map of the environment, the user sets some checkpoints, and the robot tries to find the shortest path, and navigate through them, while avoiding any obstacles. The specific purposes of this module are localization of the robot, and maintaining a virtual top-view map of the environment.

To achieve this goal, the image (Figure 2) sequences received from the PAL-optics is mapped to a 'virtual top-view' (Figure 3). The result of this transformation is the following: the straight edges lies on the floor near to the robot become straight on the transformed image also, and this step later allows the insertion of the transformed part in the global map. To apply this transformation the algorithm assumes that the PAL-image is a regular annulus. Thus, transforming the distance of the pixels from the PAL-center will result in a top-view, map-like image. To determine the parameter of the transformation function, during a calibration process the relation is measured between the real distance and the pixel based PAL-distance on several points from the center of the image; and a cubic spline interpolation is applied on the measured data.

To increase the performance of the algorithm a transformation matrix is generated, which determines the source for every pixel on the resulting image. Once this matrix is generated, it can be used for every image, with real-time speed. After the top-view transformation, the module uses a static mask to cut out segments from the image, which has no information-content (for example, the central blind-spot), or the segment shows object lies far away from the robot.

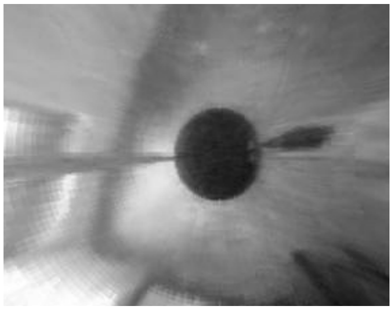

Figure 3
Virtual top-view image

The localization process uses the combination of two techniques: the program searches and tracks characteristic features on the resulting image, and parallel it calculates a summed histogram value in radial directions in every 3 degrees. The feature points and the summed histogram values are used to determine the relative location, and angle of the robot. After localizing the robot, the module will rotate the image to reflect the initial direction of the robot; the resulting image is melted into the global map also.

In order to dynamically extend the map as the robot gathers more and more information, a static bitmap would not be sufficient; instead, the map is divided it into several, small images, and the module stores the two dimensional ordering between these segments. The 'melting' of one image into the global map is used in the navigation process.

5 Results

Although the tests started recently, some early results are already available. The developed system was tested indoor and outdoor environment also. Both cases the free work space were weakly-textured and significantly lighter then the obstacles. In the first experiments some simple navigation tasks were tested: collision avoidance, line-, and object following. One complex algorithm is used for all three of the navigation types.

First, the properties of the PAL-optics needs to determined. For line following, the algorithm determines a direction line, which converges to the center of the image, and begins inner radius distance away from the center point. The optimal length of the line of sight can be set using the distance value.

The algorithm determines the pixel intensity under the line of sight starting from 90 degrees (forward), scanning at each iteration first left, then right, until it reaches ±90° scanning degree value. For every line of sight, it determines the sum of the pixel intensities, whether it exceeds the value set by threshold minimum; the maximum of these values will be used as a navigation direction. This value is later verified, whether the robot fits on the given path, or not. If there is no appropriate path on the top part of the image, the bottom part will be analyzed by the same method, and the robot will either reverse, or stop.

For free fall navigation, the Canny edge-detection algorithm is used: the contour on the image will be interpreted as obstacle to avoid. The algorithm will potentially avoid these obstacles by selecting the longest clear direction.

For object following, the algorithm uses the RGB, and HSL filter described above, to distinct the object from the rest of the environment. After applying binarization, the controlling line will head towards the object.

The developed map building system was tested in outdoor environment located near the building of dormitory. After building the map from the virtual top-view image sequences (Figure 4), binarization was used to determine the free working space of the robot. After this step, path was planned with a wave propagation based method to achieve the goal.

Conclusions

As conclusion the system has met its current specified criteria: it is capable of collision avoided line-, and object following, within a weakly textured environment and the robot capable to build up appearance based map. To improve the skills of the robot the method can be found in [4] will be implemented.

Acknowledgement

The research has been supported by the OM TDK grant under the terms of grant OM FPO 245540/2005.

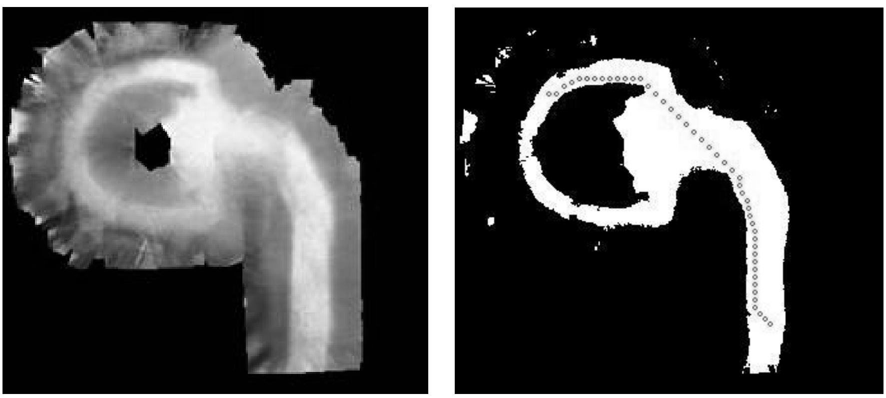

Figure 4

Constructed map from the melted image sequences (left) and the path (gray dots) generated with the wave propagation technique (right)

References

[1] P. Greguss, F. Alpek, M. Patko, "Changing from Guided Robocars to Autonomous Robocars by Using Humanoid Machine Vision System", in. *Proc. of 7th International Workshop on RAAD'98*, Bratislava, Slovakia, 1998, pp. 127-131

[2] J. Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker", Intel Corporation, Microprocessor Research Labs, 2000, http://www.intel.com/research/mrl/research/opencv/

[3] L. Mornailla, T. Pekár, Cs. Solymosi, Z. Vámossy, "Mobile Robot Navigation Using Omnidirectional Vision", in *Proc. 15th Int. Workshop on Robotics in Alpe-Adria-Danube Region*, June 15-17, 2006, Balatonfüred, Hungary, CD, ISBN 963 7154 48 5

[4] J. Tick: "Workflow Model Representation Concepts" in *Proc. of 7th International Symposium of Hungarian Researchers on Computational Intelligence*, HUCI 2006, Budapest, Hungary, November 24-25, 2006, pp. 329-337