Biologically-Inspired Visual Landmark Learning and Navigation for Mobile Robots*

Giovanni Bianco

Dept. of Electronics for Automation
University of Brescia
Via Branze 38, I-25123 Brescia
Italy
bianco@ing.unibs.it

Alexander Zelinsky

Dept. of Systems Engineering, RSISE
The Australian National University
Canberra, ACT 0200
Australia
alex@syseng.anu.edu.au

Abstract

This paper presents a biologically-inspired method for navigating using visual landmarks which have been self-selected within natural environments. A landmark is a region of the grabbed image which is chosen according to its reliability measured through a phase (Turn Back and Look - TBL) that mimics the behavior of some social insects. From the self-chosen landmarks suitable navigation information can be extracted following a well known model introduced in Biology to explain the bee's navigation behavior. The landmark selection phase affects the conservativeness of the navigation vector field thus allowing us to explain the navigation model in terms of a visual potential function which drives the navigation to the goal. The experiments have been performed using a Nomad200 mobile robot equipped with monocular color vision.

1 Introduction

Animals, including insects, are proficient in navigating and, in general, several biological approaches to solving navigational tasks seem to be promising for robotics applications. In Biology the different methods of navigating have been categorized as [14]: *guidance, place recognition - triggered response, topological* and *metric* navigation.

An agent performs *guidance* when it responds to *stimuli*. An improvement of this behavior being the selection of actions as soon as a place (or an environment) has been recognized so introducing the *place recognition - triggered response*. This represents the basic step for *topological* navigation, where places are connected by means of *qualitative* links and actions. The *quantitative* knowledge of those connections gives rise to the concept of *metric* navigation. In order to perform such tasks many animals usually deal with identifiable visual objects in the environment called *landmarks* [15].

The use of landmarks in robotics is well understood [13]. One of the crucial point deals with the nature of a landmark [3]: landmarks can be *artificial* or *natural*. Of course it is much easier to deal with artificial landmarks instead of dealing with natural ones, but the latter are more appealing because their use requires no engineering of the environment. However, a general method for dealing with natural landmarks still remains to be introduced. The main problem lies in the selection of the most suitable landmarks [13, 11].

Recently it has been discovered that wasps and bees perform specific flights during the first journeys to a new place to learn color, shape and distance of landmarks. Those flights are termed *Turn Back and Look (TBL)* [5]. Once the place has been learnt using landmarks, insects can then accomplish navigation actions accordingly. Cartwright and Collett introduced a model to explain the bee's behavior when she approaches pre-learnt places [4]. This algorithm could be suitably used for robotics applications to do visual guidance but it needs to be extended.

The aim of this paper is to introduce a biologically-inspired visual guidance model for the robot homing phase. The new system can deal with natural landmarks extracted from the environment by adopting a TBL phase (section 2). Using these landmarks an extension of the Cartwright and Collett model provides us with the necessary navigation information for visual homing (section 3). Furthermore, we show that TBL

*This research was performed at the Australian National University

affects the conservativeness of the navigation vector field thus allowing us to compute a (unique) potential function which drives the landmark navigation behavior (section 4).

2 Learning Landmarks

A landmark must be reliable for accomplishing navigation tasks, those which appear to be appropriate for human beings are not necessarily appropriate for robots because of the completely different sensor apparatus and matching systems [13]. If the meaning of reliability can be established then the problem of selecting landmarks can be automatically solved.

To effectively use visual landmarks real-time performance is needed, this can lead to the use of dedicated hardware. The robot *Nomad200* (Figure 1) that was used to accomplish the tests uses the *Fujitsu Tracking Card (TRV)* which performs real-time tracking of full color templates at a NTSC frame rate (30Hz). A template is a region of the grabbed image identified by two parameters m_x and m_y representing the sizes along X and Y axes.

Figure 1: The *Nomad200*

For each template the Fujitsu system performs a match in a sub-area of the actual video frame adopting the block matching method. This introduces the concept of *correlation* given by the sum of the absolute differences between the values of the pixels. To track a template it is necessary to calculate the correlation between the template and a frame not only at one point on the frame but at a number of points within a *searching area*. The searching area is composed of 16×16 positions in the frame usually taken around a specific point. The whole set of correlations is referred to as the *correlation matrix*.

The matching system provides as an output the coordinates of the position which represents the global minimum in the correlation matrix. Different templates usually have different correlation matrixes (see Figure 2). This approach strongly resembles the *region-based optical flow* techniques [10]. There, the flow is defined as the shift that yields the best fit between the image regions at different times. But in order to select the best landmarks only static images are considered in our approach.

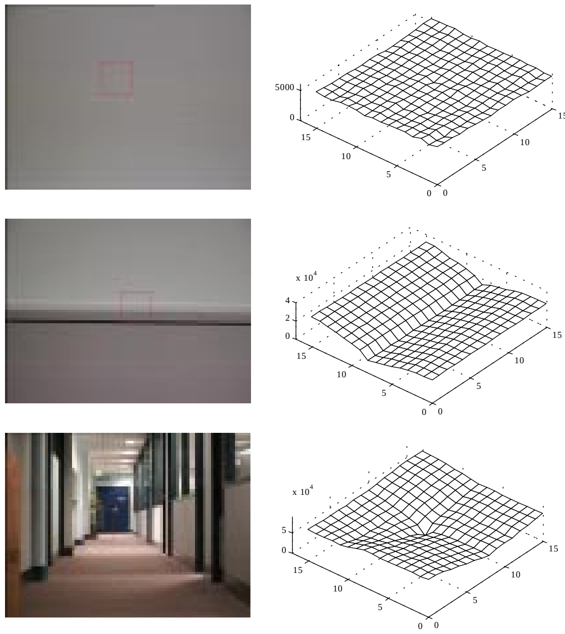

Figure 2: Various examples of correlation matrixes around the templates (box shaped)

Mori *et al.* [11] have taken advantage of the correlation matrix to generate attention tokens from scenes by using what they called the *valley method*. Extending the concept we introduce a measure given by:

$$r = 1 - \frac{g}{g'} \tag{1}$$

which for our landmarks represents the *reliability* of a template whose global minimum is g and local minimum is g'. The latter is computed in a circle around g. Therefore, we define *reliable landmarks* as *templates which are uniquely identifiable* in a search area.

There are several degrees of freedom in searching for the best landmarks within a video frame but some simplifications can be made to speed up the whole search process [1]. For example, only squared templates can be considered, thus $m_x = m_y = m_{xy}$.

So far, we use landmarks that have been *statically* selected from the goal image. We found that it is necessary to *test* them in real situations in order to verify whether they represent good guides for real navigation tasks.

TBL helps to verify the usefulness of landmarks by testing whether during the motion the statically chosen landmarks still remain robustly identifiable [1, 5]. Through a series of stereotyped movements small perturbations (local lighting conditions, changes in camera heading, different perspectives and so on) can influence the reliability of the statically chosen landmarks. Figure 3 shows the robot moving away from the goal and the camera (arrows pointing to the top) is continuously pointing towards the goal.

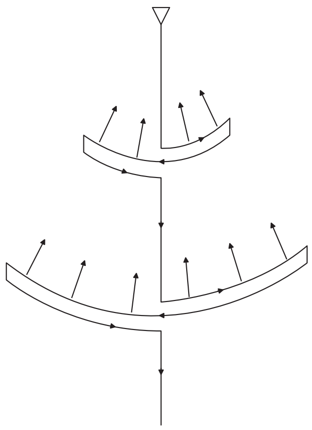

Figure 3: The TBL phase

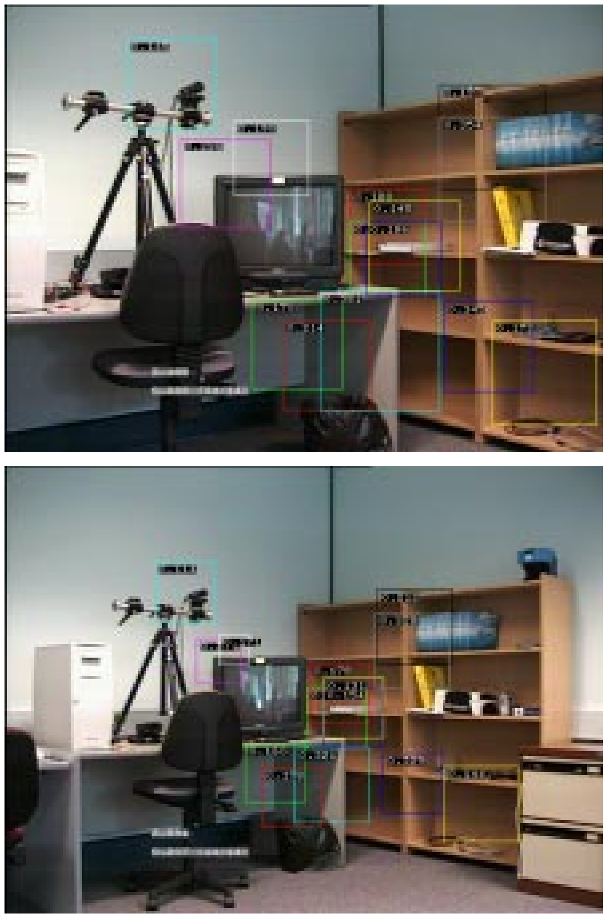

Figure 4: Two images taken during a TBL phase

The robot tries to *learn* which landmarks can be suitably used in real navigation tasks by simulating the conditions the robot will encounter along various possible paths to the goal.

At the end of the TBL process only those landmarks whose reliability r_l is above a threshold ϵ are classified as suitable for navigation tasks. This allows us to deal with *reliable* landmarks. TBL is performed within a region two meters wide and three length and takes about 13 seconds to be performed. In Figure 4 two pictures taken by a TBL phase exploited by the robot are shown. The numbers associated with each landmark represent r_l at different times.

3 Navigation from landmarks

After reliable landmarks have been chosen then navigation information can be extracted from them. The underlying biological principle is that a real movement is represented by an *attraction* force. It is produced by taking into account that the agent tries to restore the original position and size of every landmarks [4]. The Cartwright and Collett model takes into account only mono dimensional information therefore an extension to the real 2D images is necessary. Two different kinds of information can be easily computed in real time: the actual displacement of a landmark from its position viewed from the goal place and its present size.

Let $\vec{d_l}$ the difference between the original and the present position of a landmark l and W_l a weight given by the ratio $\frac{M_{xy}}{m_{xy}}$ where M_{xy} is the original magnification value. The attraction strength given by landmark l is computed as the force produced by:

$$\vec{v_l} = \vec{d_l} \cdot W_l \tag{2}$$

All the data coming from different landmarks must be fused together and the *averaging by confidence* paradigm is applied in our context [14, 6]. The individual data are weighted by introducing a sigmoid function $s(r_l)$ ranging from 0 to 1. The confidence given by the sigmoid function is based on the actual value of the reliability factor r_l. The overall navigation vector can be thus calculated as:

$$\vec{V} = [V_x \; V_y] = \frac{\sum_{l=1}^{L} \vec{v}_l \cdot s(r_l)}{\sum_{l=1}^{L} s(r_l)} \tag{3}$$

where L is the number of landmarks chosen after TBL, r_l is the reliability value of landmark l and $\vec{v}_l$ is the attraction force *felt* by landmark l. Lastly, V_x and V_y represent an estimation of the distance (along x and y axes of the environment) from the actual position to the goal position. The vector $\vec{V}$ represents the next movement from the present robot position.

Figure 5 summarizes the situation where the picture represents a typical frame taken during a navigation test. In particular, the circle at the bottom-center represents the overall attraction exerted by the goal. Above the circle the variance of that attraction is reported and under the circle the attraction vector is broken down into a magnitude and an angle. In the circle on the right the single attraction exerted by each landmark is drawn. Each landmark has a number associated with it given by the value of the sigmoid function applied on its reliability measure. The arrows at the top-center of the figure represent the motion commands given to the robot. In the rectangle on the left the visual potential field profile which has been followed so far is drawn.

In Figure 6 the results of a navigation experiment are shown. Most of the navigation trials reach the goal position within an error range of 25 centimeters. Some starting points, namely those labeled 3, 4 and 7 are not attracted by the goal. These failures can be dealt with by introducing a potential field around the goal position.

4 The visual potential field

The whole set of attractions felt by the robot can be measured by placing the robot in different points of an environment are represented by a vector field, as shown in Figure 7.

A field of vectors $\vec{V}(q)$ is defined as a *potential field* when it is produced by a differentiable function U with:

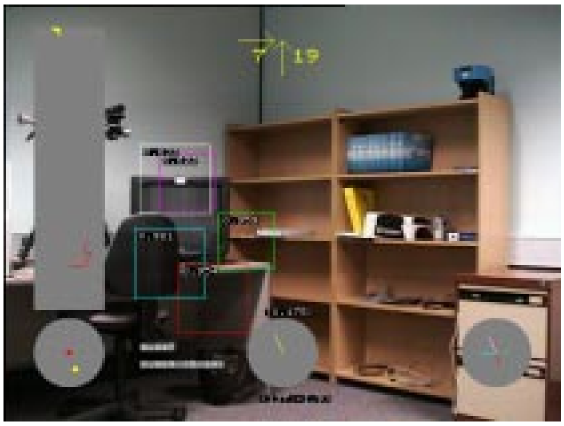

Figure 5: A frame taken by a real navigation video (see text on the left)

$$\vec{V}(q) = \vec{\nabla} U(q) \tag{4}$$

where q is the actual *configuration* of the robot [9]. Classically, both U and q are mathematically specified and researchers (see, e.g. [8, 7]) have addressed the problem of choosing the best choice for U given a set of sensors and the environment. In our case, the configuration q of the robot depends on the actual position of the robot: x, y and θ. Our approach consists of calculating the shape of $U(q)$ as *a posterior*, starting from the vector field $V(q)$.

To simplify the whole process of data collection a fixed θ is considered [2]. This means that only two dimensions (x and y) are considered instead of three, but the whole approach still remain valid.

Therefore, from equation 4 it can be stated that:

$$\vec{V}(x,y) = [V_x(x,y) \; V_y(x,y)] = \left[\frac{\partial U(x,y)}{\partial x} \; \frac{\partial U(x,y)}{\partial y} \right] \tag{5}$$

Now, in order to state that the vector field can be derived from a potential function U a necessary and sufficient condition is that the relation (*Cauchy-Riemann*) holds [12].

$$\frac{\partial V_x(x,y)}{\partial y} = \frac{\partial V_y(x,y)}{\partial x} \tag{6}$$

In other terms:

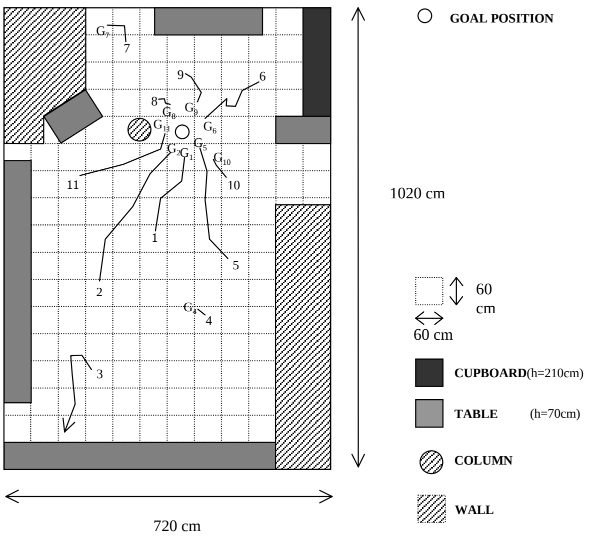

Figure 6: Navigation experiments

$$\frac{\partial V_x(x,y)}{\partial y} - \frac{\partial V_y(x,y)}{\partial x} = 0 \qquad (7)$$

that can be referred by with the term *conservativeness*.

In Figure 8 the degree of conservativeness for each considered point is plotted: large areas of the whole environment have $\frac{\partial V_x}{\partial y} - \frac{\partial V_y}{\partial x}$ roughly equal to 0.

Furthermore, when the TBL threshold approaches 1 the conservativeness of the vector field tends to become rapidly 0 everywhere [1]. This represents the condition for computing the potential function which drives the navigation process.

If the vector field is conservative then the calculation of the integral between two points is path-independent. This can be taken advantage for computing the potential function by allowing the goal position to have reference point co-ordinates. By setting 0 as the value of the potential at the goal, every other point is thus referred to in terms of potential in reference to the goal position. With a TBL threshold of 0.25 the potential function is represented in Figure 9.

There are large areas of the environment from which the goal position can be reached. Nevertheless, some starting points might lead to local minima where the robot can become trapped. The situations presented in the previous section can be clearly interpreted as attraction given by other minima that represent *false* goals. In [1] and [2] two different approaches

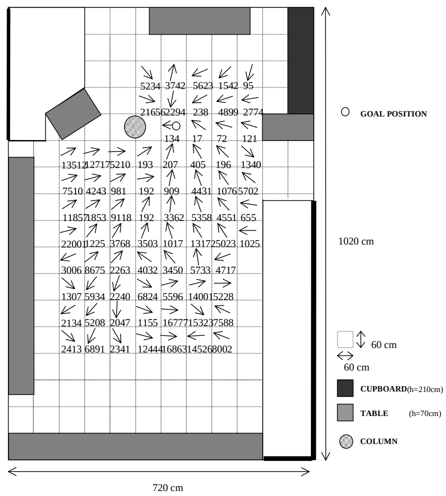

Figure 7: An example of a navigation field: directions and modules (numbers)

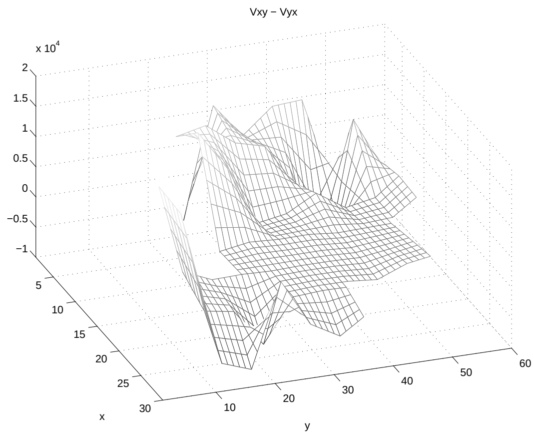

Figure 8: Conservativeness with ϵ set at 0.20. The goal position is in $(20, 25)$.

to deal with false goals have been tested: the use of a bigger threshold for TBL and the use of the whole snapshot instead of landmarks.

5 Concluding remarks

This paper has shown biologically-inspired learning and navigation systems based on natural visual landmarks. In particular the method for selecting landmarks (TBL) can affect the conservativeness of the navigation vector field thus allowing us to explain the navigation method in terms of a potential function. A panoramic field of view is thought to be necessary to produce well-shaped potential basins where the goal

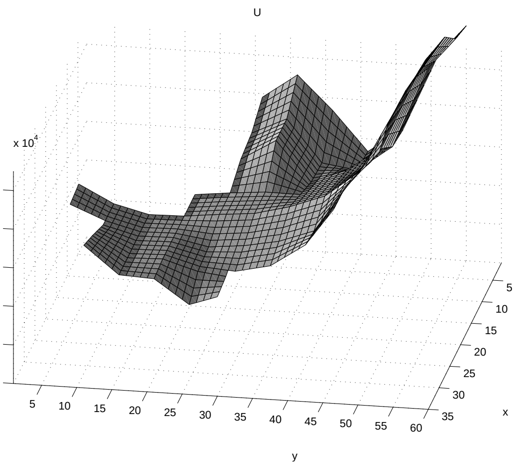

Figure 9: Potential function: TBL threshold set at 0.25

is the only global minimum. Work in this direction is necessary.

The presence of a navigation field is considered to be important for further extending the method: subgoals can be automatically placed at the boundaries of the basin of attraction. This can represent the basic step for topological navigations. Lastly, how to deal with obstacles either static or dynamic represent a mandatory development for the method.

Aknowlegments

This research would not be possible without the suggestions and the efforts of Dr. Yoshio Matsumoto, Dr. Jochen Zeil, Dr. Gordon Cheng, Mr. Jochen Heinzmann and Prof. M.V. Srinivasan.

References

[1] G. Bianco and R. Cassinis. Biologically-inspired visual landmark learning for mobile robots. In *Proc. of the European Workshop on Learning Robots (EWRL 8)*, Lausanne (Switzerland), September 18 1999.

[2] G. Bianco, A. Rizzi, R. Cassinis, and N. Adami. Guidance principle and robustness issues for a biologically-inspired visual homing. In *Proc. of the Third Workshop on Advanced Mobile Robots (EUROBOT '99)*, Zurich (Switzerland), September 1999.

[3] J. Borenstein, H.R. Everett, and L. Feng. *Where am I? Sensors and Methods for Mobile Robot Positioning*. The University of Michigan, April 1996.

[4] B.A. Cartwright and T.S. Collett. Landmark learning in bees. *J. of Comp. Physiology*, A(151):521–543, 1983.

[5] T.S. Collett and J. Zeil. Flights of learning. *J. of the American Psychological Society*, pages 149–155, 1996.

[6] S. Harmon, G. Bianchini, and B. Pinz. Sensor data fusion through a distributed blackboard. In *Proc. of the IEEE Int. Conf. of Robotics and Autom.*, pages 1449–1454, San Francisco (California), March 1986.

[7] M. Khatib and R. Chatila. An extended potential field approach for mobile robot sensor-based motions. *Intell. Aut. Systems*, 4:490–496, 1995.

[8] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. of Robotics Research*, 4(1):90–98, 1986.

[9] J.C. Latombe. *Robot motion planning*. Kluwer Academic Publisher, Boston/Dodrecht/London, 1991.

[10] J.J. Little and A. Verri. Analysis of differential and matching methods for optical flow. In *Proc. of the IEEE workshop on visual motion*, pages 173–180, Irvine, CA, 1989.

[11] T. Mori, Y. Matsumoto, T. Shibata, M. Inaba, and H. Inoue. Trackable attention point generation based on classification of correlation value distribution. In *JSME Annual Conf. on Robotics and Mechatronics (ROBOMEC 95)*, pages 1076–1079, Kavasaki (Japan), 1995.

[12] G. Strang. *Introduction to Applied Mathematics*. Wensley-Cambridge Press, 1986.

[13] S. Thrun. A Bayesian approach to landmark discovery and active perception in mobile robot navigation. Tech. rep., School of Comp. Science Carnegie Mellon University, 1996.

[14] O. Trullier, S.I. Wiener, A. Berthoz, and J.A. Meyer. Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology*, 51:483–544, 1997.

[15] R. Wehner. Arthropods. In F. Papi, editor, *Animal Homing*, pages 45–144. Chapman and Hall, London, 1992.