Image-based memory for robot navigation using properties of omnidirectional images

Emanuele Menegatti, Takeshi Maeda§ and Hiroshi Ishiguro†

Intelligent Autonomous Systems Laboratory
Department of Information Engineering (DEI)
Faculty of Engineering, The University of Padua
Via Gradenigo 6/a, 35131 Padova, Italy
§VStone Co. Ltd.
Shimaya 4-2-7, Konohana
Osaka 554-0024, Japan
†Department of Adaptive Machine Systems
Osaka University
Suita, Osaka, 565-0871 Japan

Abstract

This paper proposes a new technique for vision-based robot navigation. The basic framework is to localise the robot by comparing images taken at its current location with reference images stored in its memory. In this work, the only sensor mounted on the robot is an omnidirectional camera. The Fourier components of the omnidirectional image provide a signature for the views acquired by the robot and can be used to simplify the solution to the robot navigation problem. The proposed system can calculate the robot position with variable accuracy ("hierarchical localisation") saving computational time when the robot does not need a precise localisation (e.g. when it is travelling through a clear space). In addition, the system is able to self-organise its visual memory of the environment. The self-organisation of visual memory is essential to realise a fully autonomous robot that is able to navigate in an unexplored environment. Experimental evidence of the robustness of this system is given in unmodified office environments.

Key words: omnidirectional vision, image-based navigation, Fourier transform, hierarchical localisation, mobile robot
PACS:

1 Introduction

A mobile robot that moves from place to place in a large scale environment needs to know its position in the environment to successfully plan its path and its movements. The general approach to this problem is to provide the robot with a detailed description of the environment (usually a geometrical map) and to use some kind of sensors mounted on the robot to locate itself in its world representation. Unfortunately, the sensors used by the robots are noisy, and they are easily misled by the complexity of the environment. Nevertheless, several works successfully addressed this solution using high precision sensors like laser range scanners combined with very robust uncertainty management systems [19] [2]. Another solution, very popular in real-life robot applications, is the management of the environment. If artificial landmarks, such as stripes or reflecting dots, are added to the environment, the robot can use these objects, which are easy to spot and locate, to calculate its position on a geometrical map. An example of a successful application of this method is the work of Hu [8].

Unfortunately, these two approaches are not always feasible. There are situations in which an exact map of the environment is either unavailable or useless — for example, in old or unexplored buildings or in environments in which the configuration of objects in the space changes frequently. So, the robot needs to build its own representations of the world. This internal representation can be something different from a metrical map. As an example let us consider topological maps. These are representations of the environment that capture the topology of the environment and that have been successfully used for robot navigation and map building [4] [14] [18]. This means that in most cases a geometrical map contains more information than that needed by the robot to move in the environment. Often, this adds unnecessary complexity to the map building problem. Kuipers showed that is possible to construct a hierarchical description of the environment [13] by first building a topological map and then, on top of it, a metrical description of the environment. In a previous work we showed it is possible to implement these ideas in a real robot fitted with an omnidirectional vision system [15].

In addition to the capability of reasoning about the environment topology and geometry, humans show a capability for recalling memorised scenes that help themselves to navigate. This implies that humans have a sort of *visual memory* that can help them locate themselves in a large environment. There is also experimental evidence to suggest that very simple animals like bees and ants use visual memory to move in very large environments [5]. From these considerations, a new approach to the navigation and localisation problem developed, namely, *image-based navigation*. The robotic agent is provided with a set of *views* of the environment taken at various locations. These locations are

Fig. 1. The omnidirectional vision sensor used in the experiments.

called **reference locations** because the robot will refer to them to locate itself in the environment. The corresponding images are called **reference images**. When the robot moves in the environment, it can compare the current view with the reference images stored in its visual memory. When the robot finds which one of the reference images is more similar to the current view, it can infer its position in the environment. If the reference positions are organised in a metrical map, an approximate geometrical localisation can be derived. With this technique, the problem of finding the position of the robot in the environment is reduced to the problem of finding the best match for the current image among the reference images. The problem now is how to store and to compare the reference images, which for a wide environment can be a large number.

As we will see in Section 2.1, different methods have been proposed. In this paper, we have fully developed a method we proposed in a previous work [10]. The robot is equipped with an omnidirectional camera and takes a set of omnidirectional images at the reference locations, then it compares the current omnidirectional image with the reference images. In order to store and match a large number of images efficiently, we transform each omnidirectional view into a compact representation by expanding it into its Fourier series. The agent memorises each view by storing the Fourier coefficients of the low frequency components, that we call the "**Fourier signature**" of the image. This drastically reduces the amount of memory required to store a view at a reference location. Matching the current view against the visual memory is computationally inexpensive with this approach. Details on how to calculate the *Fourier signature* from the original image are given in Section 2.1. In Section 2.2, we will describe the process of matching the current view against the visual memory. This process is derived from calculating the degree of

Fig. 2. An omnidirectional image taken at a reference location.

Fig. 3. The panoramic cylinder created by the omnidirectional image of Fig. 2.

similarity between two omnidirectional images using the signatures associated to them. In Section 2.3, we will show experimental evidence of what we called *hierarchical localisation* in a complex real-world environment in which many objects are present. In Section 3.1, we will show experiments in which the robot explores a new environment and memorises the local views at many locations. When the exploration phase is finished, it organises the memorised views so that they reflect the geometry of the environment. In Section 3.5, we explain how the robot plans its navigation toward a destination in a reactive manner by using its self-organised memory.

2 Image-based Localisation

As we pointed out in the introduction, the first problem to tackle when building an image-based localisation system is to find a manageable way of storing and comparing the reference images. The aim is to have a data set that fully describes the environment and enables the system to reliably associate the current view with the reference view taken at a nearby location, while keeping the dataset small enough to be easily stored and quickly processed.

The first step, in order to lower the number of required reference images, is to use an omnidirectional camera. In fact, if the robot is fitted with a standard perspective camera, the view of the environment from a certain location changes with the direction of gaze. To be able to recognise this point regardless of the instantaneous heading, the robot needs to take several pictures in different directions. The amount of memory required to store and retrieve such a large number of images can rapidly grow to an unmanageable size. A solution can be to constrain the movements of the robot in order to keep the camera pointing at the same location [3], but this greatly limits the motion of the robot. Another solution can be to extract from the images some features that reduce the amount of required memory while retaining a unambiguous description of the image [20]. Nevertheless, working with a perspective camera, collecting such a large number of images is tedious and time consuming. Therefore, we used the omnidirectional camera depicted in Fig. 1. This camera mounts an hyperbolic mirror with a black needle at the apex of the mirror to avoid internal reflections on the glass cylinder [9]. A single omnidirectional image gives a 360° view of the environment from a certain location, see Fig. 2.

One might object that omnidirectional images have a low resolution, but this usually is not a limitation in tasks like navigation and localisation. In fact, we are more interested on the position of the objects than in the details on their surfaces. Actually, to some extent, the low resolution can be an advantage, because it lowers the number of pixels to be processed to extract the desired information. We will show that the relatively low-resolution images we used contain enough information for the localisation and navigation task.

2.1 Image signature

Let us come to the second step, the comparison of the current image with the reference images. The simplest approach might appear to be some sort of direct comparison of two images pixel by pixel, but this will force us to store the whole image using much memory. We propose to use what we call a **Fourier signature** to represent the omnidirectional images. The Fourier signature is computed in three steps. First, the omnidirectional image is transformed in a **panoramic cylinder**, this is a new image obtained unwarping the original omnidirectional image, as depicted in Fig. 3. Second, we calculate the 1-D Fourier transform of every line of the panoramic cylinder and we store in a matrix the Fourier coefficients line by line. Third, we keep only a subset of the Fourier coefficients, those corresponding to the lower spatial frequencies, as signature for the image.

Note we do not calculate the Fourier transform of the original omnidirectional image, but we calculate the Fourier transform of the panoramic cylinder. This simplifies the problem of calculating the image similarity. First of all, the panoramic cylinder is a periodic function along the x-axis which, firstly, simplifies the calculation of the Fourier transform and secondly, is the natural

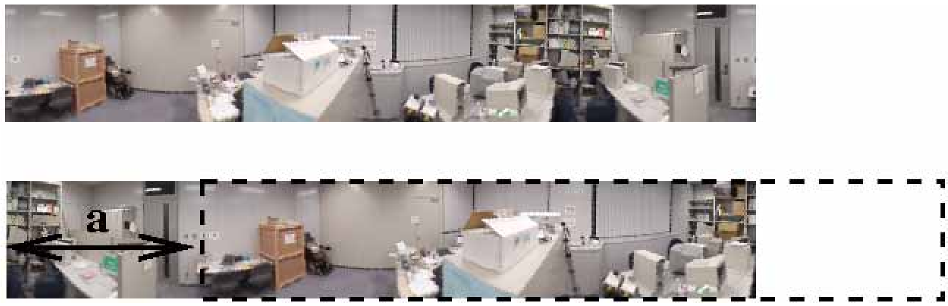


Fig. 4. Two panoramic cylinder acquired at the same location before and after a rotation on the spot. The dashed box indicates the spatial shift a between the two images.

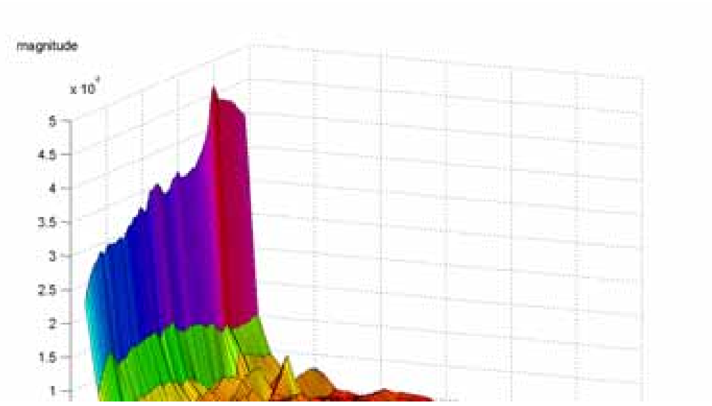
representation for implementing a *rotational invariance*. As we said, the robot must be able to match the current view with the corresponding reference image regardless of the current heading. So, we need to introduce a rotational invariance in the calculation of the similarity between two images. Using the Fourier coefficients as a signature for the images, this problem is also addressed. Let us explain how it works.

If the robot grabs two omnidirectional images at the same location but with different headings, these two images are actually the same omnidirectional image rotated about its centre. The amount of rotation corresponds exactly to the number of degrees the robot rotated. This means the two panoramic cylinders created by unwarping the omnidirectional image are actually the same image just shifted along the x-axis, like in Fig. 4. Let see how this consideration affects the Fourier transform of the two images. If $f(x)$ is one row of the first panoramic cylinder, $f(x-a)$ is the corresponding row of the shifted panoramic cylinder and by applying the **Shift Theorem**, we can write:

$$\mathcal{F}\{f(x-a)\} = e^{-j2\pi as}\mathcal{F}\{f(x)\} \qquad (1)$$

where $\mathcal{F}\{f(x)\}$ is the Fourier transform of $f(x)$. In other words, the Fourier transform of a shifted signal is equal to the Fourier transform of the original signal multiplied by the unit magnitude complex coefficient $e^{-j2\pi as}$. This property is valid for every row of the panoramic cylinder. This means that the amplitude of the Fourier transform of the shifted image is not changed and there is only a phase change, proportional to the amount of shift a.

Coming back to our panoramic images, we can then associate the magnitude of the Fourier transform to the appearance of the environment from a particular place and the phase of the Fourier transform to the heading of the robot. In such a way, when the robot is turning on the spot and the apparency of the environment is not changing, the magnitude of the Fourier transform does not change. What is changing is the phase of the Fourier transform and the amount of change is proportional to the change in the heading. Associating the


magnitude
x 10⁴
5
4.5
4
3.5
3
2.5
2
1.5
1