Technical Report

Appearance-based heading estimation: the visual compass

Author: Frédéric Labrosse
Report Ref.: UWA-DCS-06-048
Date: 28 February 2006
Version: 2.1
Status: Release

Department of Computer Science,
University of Wales,
Aberystwyth,
Ceredigion SY23 3DB,
U.K.

Please note that most figures in this report are better viewed in colour.

Abstract

In this report we present an algorithm to estimate the heading of a robot relative to a heading specified at the beginning of the process. This is done by computing the rotation of the robot between successive panoramic images, grabbed on the robot while it moves, using a sub-symbolic method to match the images. The context of the work is Simultaneous Localisation And Mapping (SLAM) in unstructured and unmodified environments. As such, very little assumptions are made about the environment; the few made are much more reasonable and less constraining than the ones usually made in such work.

The algorithm's performance depends on the value of a number of parameters, values being determined to provide overall good performance of the system. The performance is evaluated in different situations (trajectories and environments) with the same parameters and the results show that the method performs adequately for its intended use. In particular, the error is shown to be drifting slowly, in fact much slower than un-processed inertial sensors, thus only requiring unfrequent re-alignment, for example when re-localising in a topological map.

1 Introduction

One of the difficult tasks a mobile robot faces when autonomously navigating is localisation, whether in a known or unknown environment (for example in the context of Simultaneous Localisation And Mapping, SLAM). Localisation is often limited to finding the position of the robot, assuming that the orientation, or heading, is known from another source. For example, the heading can be estimated using a known magnetic field, known landmarks such as the sun or stars, or by integrating rotational information either provided by odometers or accelerometers.

Using external, fixed information such as a magnetic field or stars is attractive since this does not require integration but provides an immediate estimate of the orientation. However, such data is not always available. For example, there is no reliable magnetic field on Mars and Earth's magnetic field can vary enormously when close to other sources of magnetic fields (such as electrical equipment or ferrous ore bodies) or when close to the Earth's poles. Stars can also not be visible for example when indoors, underground, under an overcast sky or during day time.

Integrating rotational information has the major drawback of generally becoming less and less accurate as integration introduces additive errors at each step. In this report we describe a method to integrate rotational information to estimate the heading of a robot. The rotational information is provided by a system that uses panoramic views of the environment of the robot grabbed by an omni-directional camera on the robot. The system only performs simple computations on the data, namely pixel shifts and distances between images. In particular, high-level (such as feature) extraction is avoided. Rather, only the appearance of the environment is used; the method is purely sub-symbolic. To this end, an algorithm will be described that uses these simple comparisons between successive images to offer a good compromise between robustness, reliability and portability.

Early results of this work have been presented elsewhere [Labrosse, 2004]. The visual compass proposed in this paper is one of the building blocks of a large project on SLAM [Mitchell and Labrosse, 2004, Neal and Labrosse, 2004].

Section 2 introduces related work and describes where our work stands and can be used. Section 3 describes the theory behind the visual compass, presents experiments designed to evaluate it and quantify its performance, and in particular the dependency of the performance on various parameters of the method. Based on this, an algorithm is devised, Section 3.3, and then evaluated with various datasets acquired in a range of situations covering both indoors and outdoors environments, Section 3.4. Finally the method and presented results are discussed and future avenues of research are introduced. A conclusion is presented in Section 4.

2 Related work

Visual navigation increasingly relies on local methods: paths are specified in terms of intermediate targets that need to be reached in succession to perform the navigation task [Vassallo *et al.*, 2002, Neal and Labrosse, 2004, Gourichon, 2004]. This task can thus be reduced to a succession of homing steps [Mitchell and Labrosse, 2004].

Many of these homing methods that use vision require the heading of the robot to be constant or at least known. This is for example the case of most methods derived from the *snapshot model* [Cartwright and Collett, 1983, Cartwright and Collett, 1987] (see [Gourichon, 2004] for a review and "genealogy" tree of such methods and [Ruchti, 2000] for links between biology and computational models). A snapshot is a representation of the environment at the homing position, often a one-dimensional black and white image of landmarks and gaps between landmarks (e.g. [Röfer, 1995, Möller *et al.*, 1999]), but also two-dimensional images of landmarks such as corners (e.g. [Vardy and Oppacher, 2003]). Most of these methods use panoramic snapshots.

Two notable exceptions to the heading requirement estimate the change in heading between the current and target orientations using balancing of the optical flow [Röfer, 1997] or by performing a search in a parameter space containing the change in heading [Franz *et al.*,

1997].

Vision-based localisation [Thompson *et al.*, 1993] usually uses some transformation of the current view of the environment from the robot and matches it (often using a search in the pose space) with similar information synthesised from a map of the environment. For example, features of mountain images (such as peaks and ridges) can be extracted from range images (seen from the air) [Shaw and Barnes, 2003] or skyline [Naval Jr. *et al.*, 1997, Cozman *et al.*, 2000]. In some cases, the heading is determined at the same time as the position, which sometimes can lead to wrong matches [Neal and Labrosse, 2004], while in other cases, the heading is determined using an external reference.

Such external reference can be obtained using a variety of sensors (and often a combination of these sensors). For example, a magnetic field can be used on Earth, while sun light polarisation can be used outdoors (both being cues used by insects, e.g. [Frier *et al.*, 1996]). The position of the sun can also be used [Cozman and Krotkov, 1995], or more generally alignment with landmarks or beacons. For example, methods using alignment with a remote landmark[1] or linearly transforming the retinal position of three landmarks in panoramic images is proposed in [Gourichon, 2004]. The Global Positioning System (GPS) can also provide heading information by estimating the derivative of motion information. On the contrary, integrative methods can also be used. This is for example odometry that measures wheel turn and infers the position and orientation of the robot with respect to an initial reference. This is extremely unreliable, especially in outdoors situations and/or with skid-steering (the type of robot used in the outdoors experiments reported here). A more reliable way is to use inertial sensors: accelerometers for position estimation and gyroscopes for angular estimation. These work by integrating twice the signal to provide the desired information. The main drawback of these integrative methods is accumulation of error. However, there is evidence that insects, in particular ants, use odometry and dead-reckoning to compute a homeward vector [Wehner *et al.*, 1996, Åkesson and Wehner, 2002] and such sensors have been used with success in robotic applications (e.g. [Hogg *et al.*, 2002]).

Using vision to compute the heading is attractive, especially when it is also used to control the navigation. Indeed, this implies that less sensorial modalities are needed. Moreover, vision works everywhere (given the right camera) while other modalities often fail.

Almost all methods using vision for navigation extract features from the images and use these features to perform some matching. The features can be the landmarks of the snapshot and derived models, skyline in [Cozman *et al.*, 2000], etc. Moreover, landmarks can be characterised with a set of parameters such as size, area and contour length [Möller, 2001]. However, extracting landmarks and possibly their parameters can be expensive and certainly implies assumptions about the world, in particular on its structure [Gonzales-Barbosa and Lacroix, 2002]. Indeed, natural environments often present no obvious visual landmarks or when these exist, they are not necessarily easy to distinguish from their surroundings. Moreover, matching (or recognising) landmarks can be difficult if not impossible and tends to be expensive, in particular in terms of neural implementation [Möller *et al.*, 1999].

Instead, we propose to use the images as they are, with as little pre-processing as possible (none here, see Section 3.1.1 for more details); this is the *appearance-based* approach [Labrosse, 2004, Mitchell and Labrosse, 2004, Neal and Labrosse, 2004] (or signal-based approach in [Cozman *et al.*, 2000]). Using whole two-dimensional images rather than few landmarks extracted from images reduces aliasing problems; indeed, different places can look similar, especially if "seen" using only a few elements of their appearance. Finally, note that using the whole image in some cases can be equivalent to explicitly using a sub-set of the information. This is for example the case of images taken in either completely homogeneous or extremely unstructured outdoors environments; in that case, the skyline does not constitute much less information than the whole image.

Not many published papers propose to use raw images; a few examples follow. A one-dimensional panoramic image is used in [Röfer, 1995], from which the optic flow is extracted between successive images to control the robot. An array of light sensitive sensors (typically eight) is used in [Bisset *et al.*, 2003] to represent and recognise places; a process similar to the one described in [Neal and Labrosse, 2004] is used to provide rotation independence. In [Gonzales-Barbosa and Lacroix, 2002], histograms of Gaussian derivative filtered images are used[2]. A detailed study of the pixel-wise comparison (Euclidean distance) between panoramic images captured in outdoors environments is done in [Zeil *et al.*, 2003]. Finally, [Franz *et al.*, 1998] mentions the possibility of using a method similar to the one we present in this report to compute the change in heading between current and target position but discards it because it becomes unreliable when the two compared images are too different. Although we do not necessarily agree with this[3], we propose here an incre-

[1] Note that this is a capability that insects have, e.g. [Zeil *et al.*, 1996, Graham *et al.*, 2003].

[2] Note that the histograms constitute a somewhat reduced amount of information compared to using the whole images because the structure or spatial organisation of visual elements is lost. However, because the histograms are computed over rings of the images, some of the structure is preserved. Other publications mention different types of regions, e.g. [Gaspar *et al.*, 2000, Jogan and Leonardis, 2000].

[3] The quantity "too different" obviously needs to be characterised. However, experiments we have done show that if

mental method that estimates the change in heading by comparing successive images carefully chosen while the robot moves. The comparison is also designed for the task at hand but is a subset of a more general comparison performed for mapping [Neal and Labrosse, 2004] and homing [Mitchell and Labrosse, 2004]. The method is described in Section 3.

The template (mostly based on a set of visual landmarks[4]) *vs* parameter hypothesis in insects is the subject of many publications (see [Ruchti, 2000] for a review). There seem to be some evidence that desert ants use the latter [Möller, 2001], although the former clearly dominates amongst the implemented methods. There is also plenty of evidence that insects use magnetic and light polarisation cues [Frier *et al.*, 1996, Wehner *et al.*, 1996], but visual cues also seem important [Zeil *et al.*, 1996, Frier *et al.*, 1996, Graham *et al.*, 2004]. However, we emphasise the fact that we do not seek a biologically plausible solution in this work. As the next section shows, the described method uses retinal shifts of the images while there is plenty of evidence that insects do not perform such a shift but rather use fixed retinotopic images, e.g. [Wehner *et al.*, 1996], and that several snapshots are actually stored for each location [Judd and Collett, 1998]. Rather, we approach the problem with an engineering perspective.

3 The visual compass

In this section, the visual compass is presented. We start with the theory behind it and, based on that, devise experiments to assess its performance.

The approach used in this project is purely appearance-based, i.e. does not extract any high-level information from images about the world. This alleviates the need for complex feature extraction, which to be possible and efficient needs to make assumptions about the world in terms of the features it contains. Because such assumptions are not portable, we want to work at a sub-symbolic level, in other words at the level of the visual signal: the appearance of the world surrounding the robot.

One then needs to answer the following questions: *How can appearances be compared?* and *What information about the robot in its environment can be obtained from this comparison?* In the following sections, we will discuss appearance comparison and its use in extracting rotational information.

3.1 Theory

We define here the *appearance*, describe how appearances can be *compared* and finally how useful *information* can be extracted from appearances when comparing them.

Figure 1: An omni-directional image

3.1.1 Appearance

In this work, the appearance of the environment surrounding a robot from a given pose is an image taken by the robot of the environment. The idea is not new and has been used in recognition (e.g. [Bichsel and Pentland, 1994]) and inspections (e.g. [Nayar *et al.*, 1996]) tasks.

Directly using the appearance of the world by opposition to extracting features or the structure of the world is attractive because methods can be devised that do not need precise calibration steps as will be shown later.

Like others, e.g. [Jogan and Leonardis, 2000, Cozman *et al.*, 2000, Gaspar *et al.*, 2000, Gonzales-Barbosa and Lacroix, 2002, Goedemé *et al.*, 2005], we use panoramic images as they provide in one image everything that can be seen from the current position. More precisely, we use an omni-directional camera to grab omni-directional images, Figure 1. The camera is made of a "normal" camera pointed upwards and looking at a hyperbolic mirror linked by a perspex tube to the camera, Figure 2. Note that because of the projection on the mirror, right and left are inverted in omni-directional images.

For ease of processing, Section 3.1.3, the omni-directional image is unwrapped into a panoramic image that we define as the appearance of the world from that particular position, Figure 3. The unwrapping excludes the parts of the omni-directional image that correspond to either the robot or the outside of the mirror and only keeps the useful part of the image, i.e. the part corresponding to the surroundings of the robot. This *useful part of the image* as well as the unwrapping procedure will be discussed and experimented with later.

the camera remains "far", which again needs to be quantified, from obstacles or that the comparison is performed with care, then the method works, see Section 3.1.3.

[4]The authors are not aware of any work studying a more global matching method of images in insects.

Figure 3: The panoramic view corresponding to the omni-directional image on Figure 1: the appearance from the corresponding position of the robot

Figure 2: The omni-directional camera: the "normal" camera (bottom) and the hyperbolic mirror (top) linked by a perspex tube

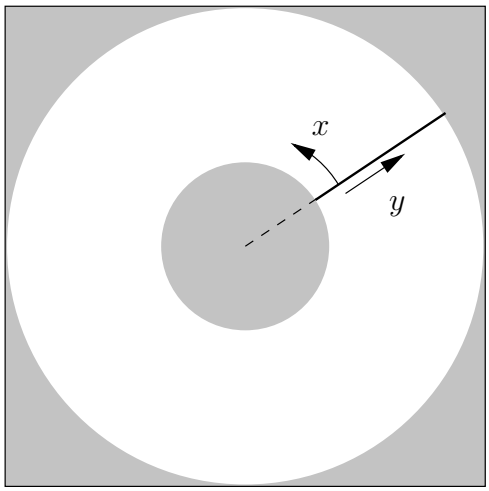


Figure 4: The unwrapping process. The grey areas correspond to unusable parts of the image (projection of the robot or outside of the mirror). x and y are the coordinates of pixels in the panoramic image.

The unwrapping is performed by scanning a line emanating from the centre of the omni-directional image and rotating it around the image by a fixed increment. Pixels of the panoramic image are taken along the line at regular intervals using the nearest pixel of the omni-directional image[5], Figure 4. The direction of the unwrapping is such that the front and back of the robot are respectively at columns $0.25 \times w$ and $0.75 \times w$ from the left of the image, where w is the width of the panoramic image.

The unwrapping process is controlled by several parameters. The height of the panoramic image depends on the thickness of the white "doughnut" on Figure 4 and the sampling along the rotating line. This thickness is determined by the size of the omni-directional images and the size of the robot in the images. The sampling along the line can be non-linear to give more importance to various parts of the images. The outside of the doughnut mostly corresponds to the parts of the environment that are far away from the robot and/or that are tall while the inside corresponds to the parts that are close to the robot. Here we chose to use a linear sampling of one pixel in the panoramic image for each pixel along the line. This is further discussed in Section 3.2.5. The width of the panoramic images is also of importance. For ease of computation, it should be a multiple of 360, each pixel then representing a portion of a degree of angle of the environment surrounding the robot. This is further discussed in Section 3.2.2.

This unwrapping method does not produce a completely accurate representation of the environment, as the geometry of the mirror used in the omni-directional camera is not taken into account in the above transformation and no calibration of the system has been done. In particular, the perspex tube used to link the camera and the mirror is not perfect and introduces many local distortions. Moreover, the process assumes that the optical axis of the camera projects at the centre of the image, which is very probably wrong. However, as long as the distortions do not change dramatically with time, calibration is not needed as will be discussed later, Section 3.1.3.

3.1.2 Image space and appearance comparison

The input to the system is thus made of appearances (images) that need to be compared. An $h \times w$ pixels image with c colour components per pixel is a point in the *image space*, a space having $h \times w \times c$ dimensions representing all possible images of the given size. Appearances of an object form a manifold in that space, i.e. a surface embedded in the image space but having a lower dimension than that of the image space (e.g. [Tenenbaum, 1998]). The parametrisation of the man-

[5] We have also used bi-linear interpolation in [Labrosse, 2004], but this is more expensive to compute and does not improve the performance of the system.

ifold is typically dictated by the degrees of freedom of the capture of the appearance of the object. This could include position and orientation of the camera but also factors such as illumination and geometrical configuration of the object, if variable.

Comparing appearances then becomes measuring the distance, using some distance metric, between points on the manifold along the surface defined by the manifold. It has been shown that the curvature of the manifold can be high [Lu *et al.*, 1998] and that linear interpolation between views is not necessarily enough [Bichsel and Pentland, 1994][6]. In this work however, the linear interpolation remains valid because compared appearances are not too distant in time and thus on the manifold; their linear interpolation thus remains close to the manifold.

The distance metric can be one of several. We have tried an L^1 norm (Manhattan distance) and an L^2 norm (Euclidean distance) in previous work [Mitchell and Labrosse, 2004] not showing any difference in the results obtained by either. Here we used the Euclidean distance. The distance between two images is thus defined as

$$d(\mathcal{I}_1, \mathcal{I}_2) = \sqrt{\sum_{i=1}^{h \times w} \sum_{j=1}^{c} (\mathcal{I}_2(i,j) - \mathcal{I}_1(i,j))^2}, \quad (1)$$

where $\mathcal{I}_1(i,j)$ and $\mathcal{I}_2(i,j)$ are the j^{th} colour component of the i^{th} pixel of images $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively. Pixels are enumerated, without loss of generality, in scan-line order from top-left corner to bottom-right corner. In all experiments presented here, we used the RGB (Red Green Blue) colour space, thus having three components per pixel.

The choice of the Euclidean distance is driven only by its simplicity and it not introducing any discontinuities. Similarly, using the RGB colour space is not optimum because different colours will contribute differently to the comparison, only because their RGB encoding is different, not because they are more prominent. For example, red $(1.0, 0.0, 0.0)$ will appear much less different from black $(0.0, 0.0, 0.0)$ than purple $(1.0, 0.0, 1.0)$ from black (1 against $\sqrt{2} = 1.4142$). This can create an inhomogeneous "pull" of the heading towards the brighter objects (or indeed "push" when the field of view is reduced, see Section 3.4.2). Some experiments described in this report show that RGB is adequate in most cases but does create problems in other cases. Moreover, RGB is very dependant on changes in illumination and other colour spaces would be better with that respect [Woodland and Labrosse, 2005]. Again, however, since we only compare successive images, chances are that illumination changes will not be dramatic. Moreover, RGB being the space used by the camera, using it means that no further computation is needed.

[6] This is however what is done when performing Principal Component Analysis on the image set to reduce its size and dimensionality, in other words to simplify the manifold, e.g. [Jogan and Leonardis, 2000] in a robot localisation context.

3.1.3 Extraction of rotational information

Using the appearance of the robot's environment is only useful if one can extract information from it about the environment that is useful to the robot. In this work, we are interested in rotational information.

Note that translational information is also present in changes of the appearance when the robot translates. This has been used for example in [Mitchell and Labrosse, 2004] and experimented with in [Zeil *et al.*, 2003].

The perfect case

We assume for now that the optical system is perfect. This would imply several constraining, even impossible, assumptions, namely that:

- the optical axis of the camera is aligned with the axis of the mirror,
- the camera assembly (including the mirror and perspex tube) is calibrated,
- the axis of the system is aligned with the axis of rotation of the robot and the robot **only** turns on the spot,
- the resolution of the images is infinite.

These will be relaxed later! Rotating the robot on the spot would then result in a simple column-wise shift of the appearance in the opposite direction. The exact rotation angle could be retrieved by simply finding the best match between the first image (before rotation) and a column-wise (with column wrapping) shift of the second image (after rotation). The best shift would simply correspond to the rotation undertaken by the robot. Figure 5 shows two appearances different only in a perfect, i.e. (almost) respecting the above assumptions, rotation by 10° of the robot between the two appearances. The perfect rotation was obtained by rotating the first omni-directional image into the second omni-directional image, both images being then unwrapped to produce the appearances. This does not completely satisfy the assumptions because the centre of the omni-directional images is only approximately the projection of the axis of rotation of the optical system and because the resolution of the images is not infinite. Figure 6 show the Euclidean distance between the two images as a function of the column-wise shift (rotation) of the second image:

$$d(\mathcal{I}_1, \mathcal{I}_2, \alpha) = \sqrt{\sum_{i=1}^{h \times w} \sum_{j=1}^{c} (\mathcal{I}_2(\alpha,i,j) - \mathcal{I}_1(i,j))^2}, \quad (2)$$

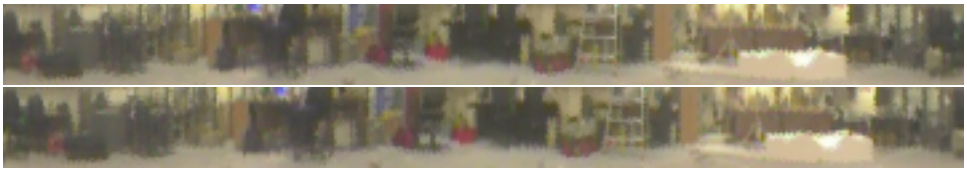


Figure 5: Two appearances from the same position but with a rotation of 10°

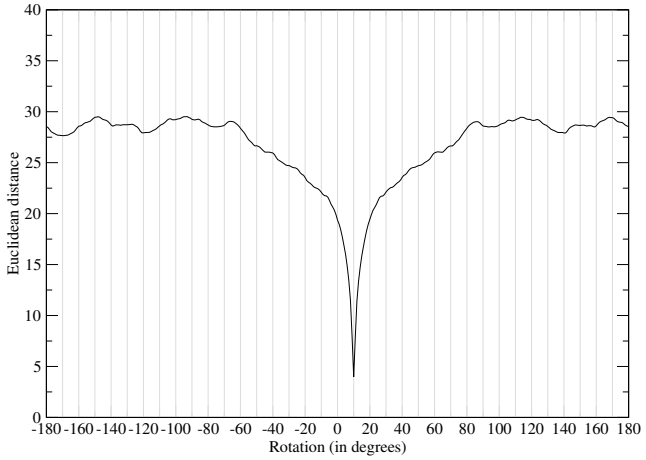


Figure 6: The Euclidean distance between the two images in Figure 5 as a function of the column-wise shift of the second image

where $\mathcal{I}_1(i,j)$ and $\mathcal{I}_2(\alpha,i,j)$ are the j^{th} colour component of the i^{th} pixel of images $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively, the latter being column-wise shifted (with column wrapping) by α pixels (corresponding to $\alpha°$ because the images are 360 pixels wide). The minimum of the function is clearly obtained at 10°. Note that the minimum distance is not 0 as both the manually performed rotation of the omni-directional image and the re-sampling done during the unwrapping process introduce errors (Section 3.1.1).

It can be seen that the function presents many local minima. However, if one assumes that the rotation between the two appearances is not too large, then finding a local minimum is enough when one starts from 0. Alternatively, if an estimation of the change in heading is available, again a local minimisation is enough. These aspects will be discussed in Section 3.3 where the minimisation algorithm will be described.

The finite resolution case

Obviously, we do not have an infinite angular resolution. This is because the omni-directional image has a finite resolution (a hardware constraint!) and the unwrapping procedure (Section 3.1.1) cannot extract more information than that present in the original image. This means that a quantisation error is made when evaluating the rotation using only column-wise shifts. However, the error is at most $\pm 0.5°$ if the angular resolution is 1° per pixel and statistically will compensate over a long run if the rotation is not systematic. Ways of improving this will be discussed later.

The general case

One of the advantages of appearance-based methods compared to full reconstruction is that they do not need any accurate calibration or even perfect optics.

The assumptions mentioned in the perfect case above are rather unrealistic and at least constraining. For example, such optical system as the omni-directional camera can be physically difficult to implement and to ensure proper reliable alignment would require it to be probably prohibitively heavy[7]. Moreover, calibration of the system would have to be performed regularly as such a system might be deformed/damaged in situations such as planetary exploration or during landing on remote places. Moreover, the robot usually does not rotate on the spot if only because it usually also translates while turning.

All these imperfections mean that a rotation of the robot (either on the spot or while moving) will result in more than a simple column-wise shift of pixels in the appearance. Indeed, the view-point will change introducing new "features" in the appearance while others will disappear. However, provided the change in view point is not too dramatic, the method above can still be used. For example, Figure 7 shows the appearances before and after a displacement of 20 cm followed by a rotation of 30° [8]. Figure 8 shows the Euclidean distance between the two appearances as a function of the column-wise shift of the second appearance. The minimum of the function is at $-30°$, indicating the correct rotation of the robot, despite the change in position resulting in a non-exact match (which is visible in the higher value of the minimum, compared to the perfect case).

To make the system even more robust, we will only consider the parts of the images that correspond to the front and back of the robot since they carry very little of the forward (or backward) translation information while the parts corresponding to the sides carry most of

[7] This is not completely true anymore as compact omni-directional cameras are readily available, although their suitability for this work hasn't been evaluated by the authors yet.

[8] The measurement of displacement and rotation is performed using the motion tracking system VICON 512, Section 3.2.1.

Figure 7: Two appearances before and after a displacement of 20 cm and rotation of 30°

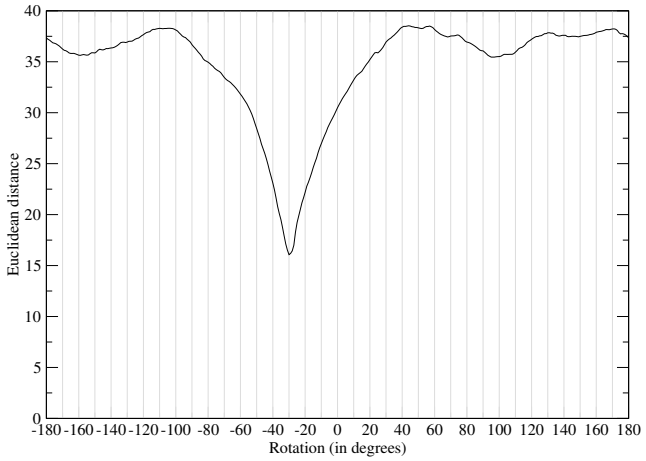


Figure 8: The Euclidean distance between the two images in Figure 7 as a function of the column-wise shift of the second image: the general case

it [Nelson and Aloimonos, 1988]. This will be discussed further in Section 3.2.3.

Dramatic changes in view point are created either by large displacements or by small displacements while the robot is close to (large) objects. Indeed, important changes in the appearance happen at occlusion points in the images. However, since we do not seek an exact match but only the best match, the method still works. The only constraint is that changes in appearance due to occlusions are small and distances in Cartesian space between the robot and obstacles are larger than the robot's displacement between successive images. This constraint is usually satisfied since robots tend to either navigate away from obstacles or, when near obstacles, move slowly.

Limitations

There are limitations in using the appearance for navigation related tasks. However, these are rather pathological and similar to human (or at least biological) limitations and are inherent to systems using vision.

The method needs visible features in the environment, despite not explicitly using them. Indeed, if the environment is completely featureless, then the distance in image space will remain constant (and null) and the method will fail. However, even low contrast in the appearance is enough to measure differences in images, while method based on explicit feature extraction will usually fail with low contrast images.

Checkerboard-type environments might also make the method fail, depending on the spatial sampling rate used: if the rate is too low, aliasing might occur, i.e. different places will produce similar (if not equal) images and the heading computation will fail. The aliasing problem however will not disappear if the rate is made higher, but the heading computation will succeed.

A more subtle problem is the "one-feature-syndrome": if only one feature is visible, for example when the robot passes a single tree in a desert, the un-rotation procedure will make that feature match at the same position in the succession of images, resulting in the system reporting a rotation even when none occurred or on the contrary no rotation when one did occur, depending on where the feature is — the unique, or dominating, feature "pulls" (or "pushes" in some cases) the stabilisation process. This however, is exactly what happens to human beings in similar situations, e.g. on board a boat at sea in the night when passing another boat. We thus have the constraint of having at least two distinguishable features in the environment that should be isotropically arranged (as is the case for most methods derived from the *snapshot model*). However, because we use the appearance, by opposition to landmarks, of the world, this usually is the case. The *equal distance assumption* [Franz *et al.*, 1998] states that the landmarks must be at the same distance from the position of the snapshot. This again is also an assumption made here because different distances create different amplitudes of the optic flow, leading to different apparent rotations. However, because we use only a global measure, this is not such a problem here. These two assumptions will be revised later when the field of view will be reduced, Section 3.2.4.

3.2 Evaluation of the theory

In this section, we show the result of experiments we conducted to assess the performance of the theory. These experiments were conducted in semi-controlled environments with specific robot trajectories. The result of these experiments are then used to devise an algorithm (presented in Section 3.3), which is then evaluated in various situations, Section 3.4).

3.2.1 Experimental setup

The experimental setup is as follows. The omnidirectional camera (Figure 2, page 4) is mounted on a

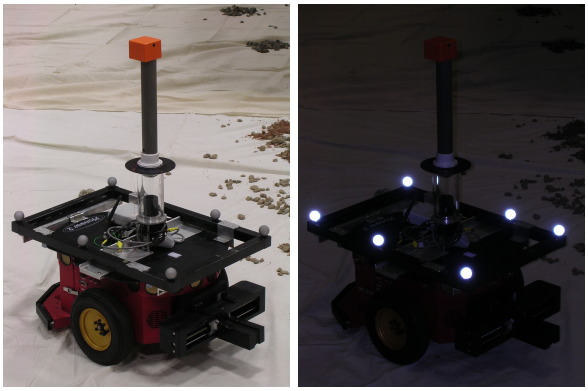


Figure 9: The indoors robot with the camera, the magnetic compass and the reflective markers

Figure 10: The outdoors robot with the camera and magnetic compass

mobile robot, either a Pioneer 2DXe for the indoor experiments (Figure 9) or a Pioneer 2AT for the outdoor experiments (Figure 10). The camera is mounted such that its axis is as close as possible to the axis of rotation of the robot. However, in the spirit of appearance-based methods, no proper calibration of this arrangement has been made. In particular, there is no guaranty that the proper alignment is obtained, nor that the axis of the camera is perpendicular to the plane defined by the top of the robot (or the one created by the wheels of the robot for that matter!). The outdoors robot uses skid-steering to turn, thus producing non-smooth rotations, especially on high-grip surfaces such as the car park used later.

Images are grabbed and can be processed on the robot. However, in all experiments reported here, the images are saved when grabbed and batch-processed later to be able to perform different experiments on the same data set with different values for the different parameters, thus providing comparable results.

The orange box at the top of the grey mast on Figures 9 and 10 contains a magnetic compass used to assess the visual compass during outdoors experiments. The compass is mounted high enough to avoid magnetic interferences created by the robot. However, the compass is disturbed by the presence of metal in the floor of our lab as well as all electronic equipment in the lab and therefore has not been used indoors. Outdoors, the compass is also sensitive, to a lesser extend, to tilt and yaw of the platform. The performance of the magnetic compass will be characterised during outdoors experiments.

Indoors, the performance of the visual compass is evaluated using the real-time motion tracking system VICON 512. The system tracks and provides in real-time the position of reflective markers. Objects can be defined in the system as a set of markers shown to it. The objects can then be tracked in real-time (at between 70 and 120 frames per second on the computer used for the experiments reported here) and a software server can then provide real-time data about the position and orientation of the object to any software client connected to it. Figure 9 shows the frame on the indoors robot along with the reflective markers.

The standard deviation of the error on the position of the markers returned by the VICON system is of the order of the millimetre, with good calibrations of the system, as stated by VICON. The reflective markers on the robot form a rectangle roughly 510 mm long by 330 mm large. The angular error is thus of the order of $2/330$ radians, thus $12/11\pi \approx 0.35°$. Note that statisticians tell us that when measuring more than one value, then the standard error becomes that of the error for one divided by the square root of the number of points, here six, which would produce an error of the angle measurement of the order of $0.14°$. However, not knowing the exact algorithm used by the VICON system to compute the orientation of the rectangle, in particular whether it does exploit redundancy between markers and how it does it, we consider the more conservative error mentioned previously.

We experimentally verified the repeatability of the tracking by obtaining from the VICON system the orientation of the robot when standing still at five different positions of the area used during the experiments reported here (centre and four corners). Table 1 shows some statistics on the data of the five sets and shows good repeatability. However, the accuracy of the VICON system is probably not constant on the area used for the experiments, therefore introducing non-constant errors over space. This is however, very difficult to measure without a third means of measuring orientations. Moreover, for our experiments, the important properties of the heading as returned by the VICON system (or the magnetic compass) is that it is absolute and thus not drifting, property that needs to be evaluated for the proposed system. Moreover, the error in the heading provided by the VICON system is likely to be

Table 1: Repeatability/accuracy of the VICON system in measuring the heading of a static object (all angles in degrees)

Set	Min.	Max.	Mean	Std. dev.
1	344.961	345.086	345.050	0.0211748
2	290.200	290.291	290.241	0.0172916
3	251.590	251.712	251.661	0.0303238
4	140.331	140.506	140.372	0.0397026
5	47.1506	47.2113	47.1833	0.0163133

at least one order of magnitude lower than that of the proposed system.

To allow comparison, when an absolute heading is available, either from the VICON system of the magnetic compass, the visual compass is initialised with the absolute value at the beginning of the experiment.

The two robots are identical software-wise and use GNU/Linux as their operating system on a PC104 format Pentium III at 800 MHz. The robots are connected to the Internet via a slow wireless connection through a firewall protecting the main wired network from the wireless network. The VICON server is on the wired network side. The robots must thus get the VICON data from the network, which in itself introduces slight delays. The software implementing the visual compass and grabbing all the necessary data for its evaluation runs a tight loop without threading to provide maximum control over the timing of its execution. However, there are two important aspects over which we have no control. One is the access to disk when saving the images. The GNU/Linux operating system tends to buffer all disk accesses but still needs to flush the transactions now and then, which does take time. More importantly, the second operation over which we have no control is the access to the VICON server over the network, in particular through the firewall, as well as the performance of this server. We have observed a few delays in the timing that are partly due to the network, but mostly due to the VICON server stopping for a few frames. Figure 11 shows the timing of one of the conducted experiments: time of the beginning of the grabbing of each image from the start of the experiment. Arrowed are the most salient extraneous delays introduced during image grabbing. These will have an effect on the result we obtain and are discussed later.

3.2.2 Pure rotation: RotateLeft

The first experiment is that of a pure rotation: the robot and the camera rotate on the spot to the left. This means that the images only change by a column-wise shift, or at least this would be the case if the mechanical assembly was perfect. Table 2 gives some statistics about the experiment.

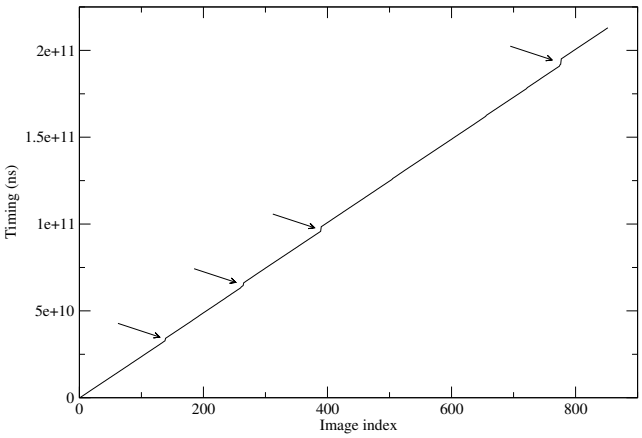


Figure 11: The timing of the pure rotation (RotateLeft) experiment, Section 3.2.2

Table 2: RotateLeft: statistics

Frames per second	4.00
Degrees per frame	1.29
Total rotation (deg)	1101.28

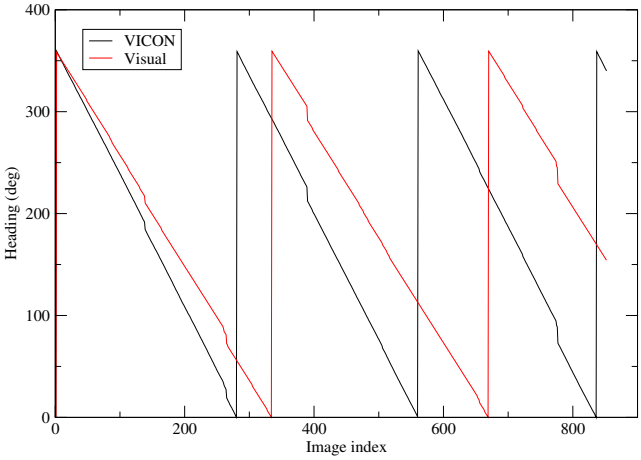


Figure 12: RotateLeft: VICON and visual heading with all the images

The theoretical method

We first run the method described in the theory (Section 3.1.3) using all the grabbed images. Figures 12 and 13 respectively show the headings as measured by the VICON system and the visual compass and the error between the two headings, all headings being measured in degrees, between 0 (inclusive) and 360 (exclusive), positive turning clockwise, as for a magnetic compass.

It is rather clear that the error increases linearly with the number of processed images. This obviously is not desirable!

As Table 2 shows, the robot rotated on average by 1.29° per frame. This, in most cases, would be rounded

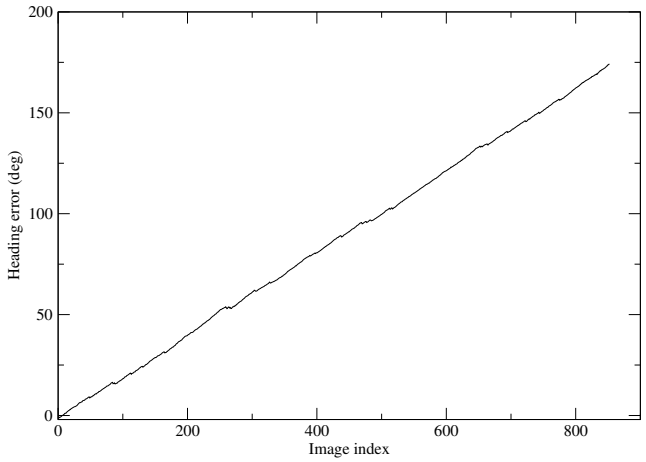


Figure 13: RotateLeft: error between VICON and visual headings with all the images

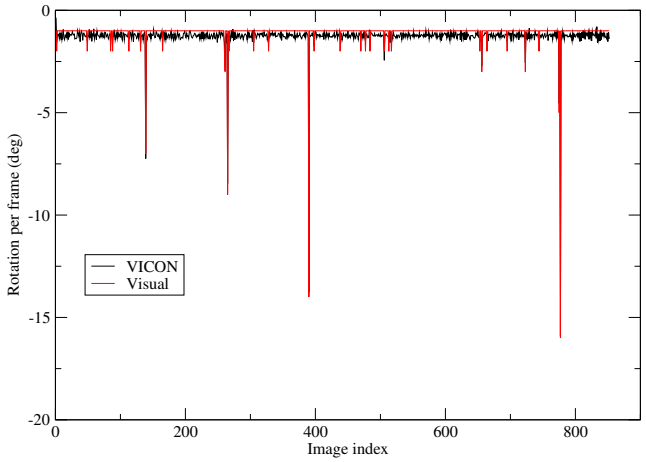


Figure 14: RotateLeft: the rotation per frame as given by the VICON system and the visual compass

by the system to a rotation of 1° between each frame, introducing an error of 0.29° per frame. In some cases however, errors cancel out, leading to a slope of the error function of 0.20, lower than 0.29. This is clearly visible on Figure 14, which shows the rotation per frame as given by the VICON system and the visual compass. Figure 15 shows different close-ups of Figure 14. They first show that indeed the rotation computed from the images is often smaller than the rotation given by the VICON system, but not always, and second that when the rotation is more important (two bottom close-ups), the visual compass follows closely the VICON system. These more important rotations are due to the timing anomalies already mentioned, Figure 11 and recapitulated in Table 3. They are clearly visible on Figure 12 in steps in the heading provided by both the VICON system and the visual compass and on Figure 14 in the high peaks of rotation. This was to be expected as the extra delay between the corresponding frames introduces a jump in the heading as a function of image

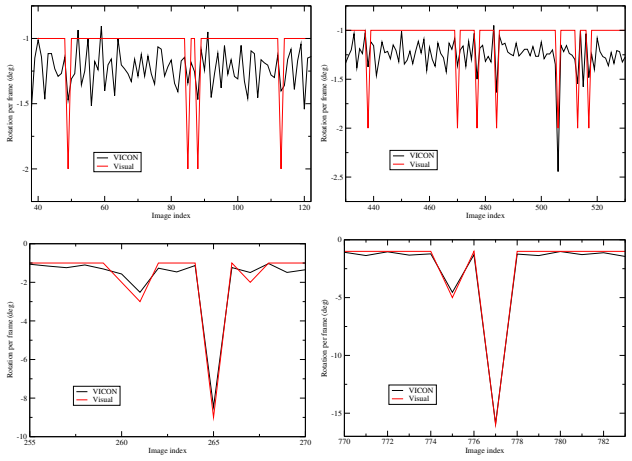


Figure 15: RotateLeft: close-ups of Figure 14

Table 3: RotateLeft: major anomalies in the timing

Image index	139	265	390	777

index. Note that in that case the effect of the rounding becomes negligible.

The systematic error occurs when the rotation is at constant speed and with a value that is not a multiple of 0.5° per frame (with an angular resolution of the appearance of 1° per pixel, see later). When it is the case, slight variations in the rotational speed mean that overall, the error accumulated will cancel out.

Several solutions to the space/time sampling problem can be envisaged, which we will discuss next: higher angular resolution of the panoramic images, interpolation of the distance measurement and better space sampling.

Angular resolution

A possible way of reducing systematic error introduced in the system is to increase the angular resolution of the appearances. This however is only possible if the resolution of the omni-directional image is high enough to indeed contain enough information for the desired angular resolution.

For example, if the omni-directional image is 400 pixels wide (and high), then the highest angular resolution that can be attained is 3.49 pixels per degree. Indeed, a circle of diameter 400 pixels provides $1,256$ pixels on its circumference, thus at the top row of the appearance image. The other rows will still have the same number of columns; their angular resolution will not be as high and will thus contain redundant information.

Figure 16 shows the error in heading between that returned by the VICON system and that computed by the visual compass, this for different angular resolutions of the panoramic images. The omni-directional images used have 400×400 pixels, thus providing at best an

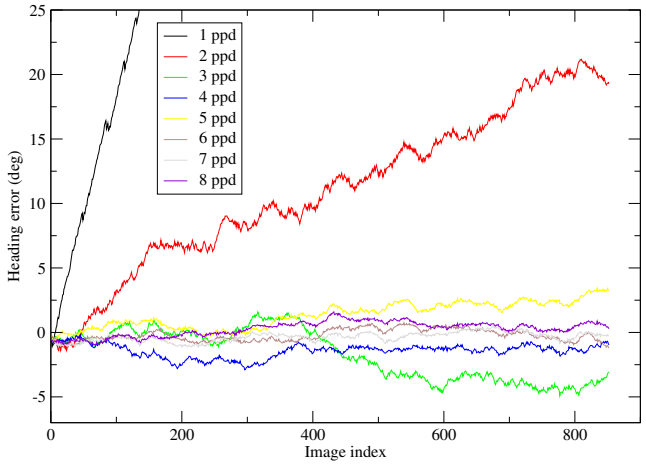


Figure 16: RotateLeft: error between VICON and visual headings with all the images and different angular resolutions (degrees per pixel)

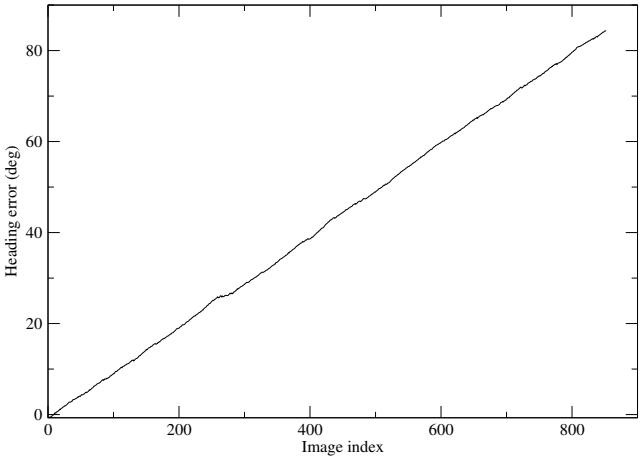


Figure 17: RotateLeft: error between VICON and visual headings with all the images and interpolation of the distance function

angular resolution of 3.49 pixels per degree. A clear improvement is visible. In particular, with resolutions approaching maximum use of the available information (from 3 pixel per degree), the heading error does not seem to diverge anymore, at least on this data set (see later for more).

A systematic error would happen only if the rotation between each frame was constant and of a value that is not a multiple of the angular resolution divided by 2. Increasing the angular resolution means that this is almost never possible, especially since the robot will physically not be able to undergo such a rotation at a **constant** speed, necessary condition for the systematic error to happen. Moreover, even if such a rotation was possible, the error would be small and thus only of consequence for long runs.

However, increasing the angular resolution also increases the computation time and is thus not very satisfying.

Interpolation of the distance measurement

To provide a sub-pixel estimation of the best match between compared images, the distance measurements can be interpolated, provided we have the right model for the distance function (as a function of rotation of one of the images against the other). Since we are only interested in improving the behaviour of the function at its minimum, we will concentrate on modelling it around this position.

Close examination of the data shows that a parabolic function can be used to interpolate the distance function around its minimum. Such function is simple to compute and we found that the position of the minimum is generally close to the minimum of the distance function with higher angular resolution images (this is not necessarily the case for the **value** of the minimum, as we see in Section 3.3, where a better model will be described).

For three values of the distance function $d(\theta)$ at angle α, β and γ (in that order), the sub-pixel minimum (where the parabola's derivative is 0) of the function is at position μ:

$$\mu = \beta - \frac{(\beta-\alpha)^2(d(\beta)-d(\gamma)) - (\beta-\gamma)^2(d(\beta)-d(\alpha))}{2[(\beta-\alpha)(d(\beta)-d(\gamma)) - (\beta-\gamma)(d(\beta)-d(\alpha))]}. \quad (3)$$

Here, β will be the pixel-accurate value for which the distance function is minimum and we have $\alpha = \beta - 1$ and $\gamma = \beta + 1$, which significantly simplifies the equation.

Figure 17 shows the error between the VICON and visual headings with interpolation. Compared to Figure 13, a significant improvement is achieved, but still the error grows linearly. However, this improvement is at very low cost compared to increasing the angular resolution.

Space sampling

Space sampling, i.e. by how much the robot moves (or only rotates in this experiment) between two consecutive images, is also of importance. There is indeed a trade-off between the frequency of the sampling on the one hand and the error introduced at each frame, the speed of processing and accuracy of the result on the other hand. Let us develop this further.

A high sampling rate will lead to very similar successive images. This implies that:

- the minimisation of the distance function will be fast and will not be trapped in a local minimum since the absolute minimum will be close to the current position;

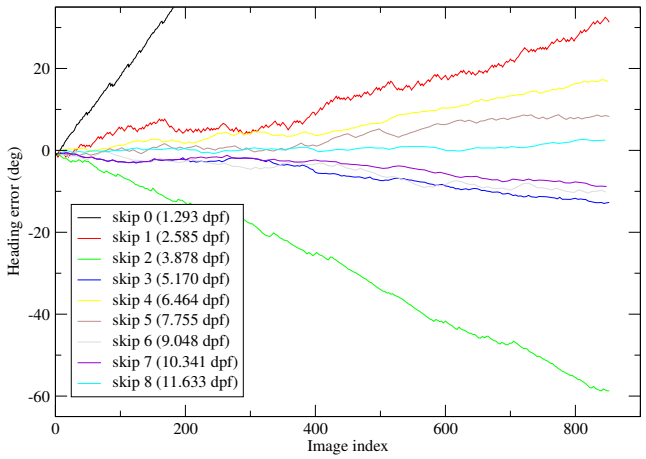


Figure 18: RotateLeft: error between VICON and visual headings for different constant spatial sampling rates. The average rotation per frame (in degrees) is given in parenthesis.

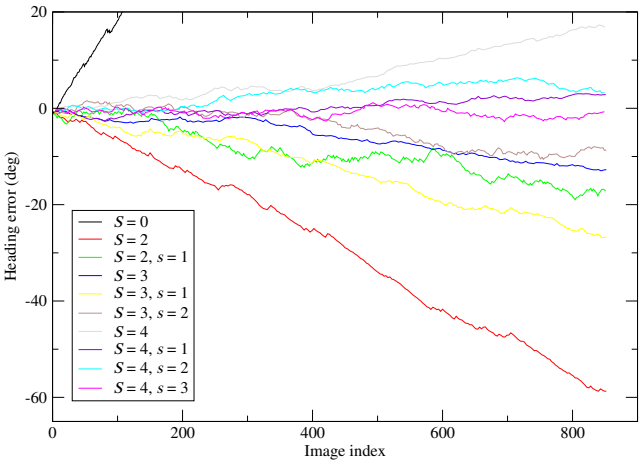


Figure 19: RotateLeft: error between VICON and visual headings for different variable (random) spatial sampling rates

- the rotational information extracted by the minimisation of the distance between the two images will be as accurate as possible since there will be little parallax error due to change in view point.

These are obviously desirable properties. However, a high sampling rate also means that:

- more images need to be processed, which in itself limits the sampling rate;
- more error due to limited angular resolution is introduced since more images are processed. This is especially true if the robot undergoes a rotation at constant speed that is not a multiple of 0.5° per rotation, with an angular resolution of 1 pixel per degree in the appearance (see the first result reported in this section).

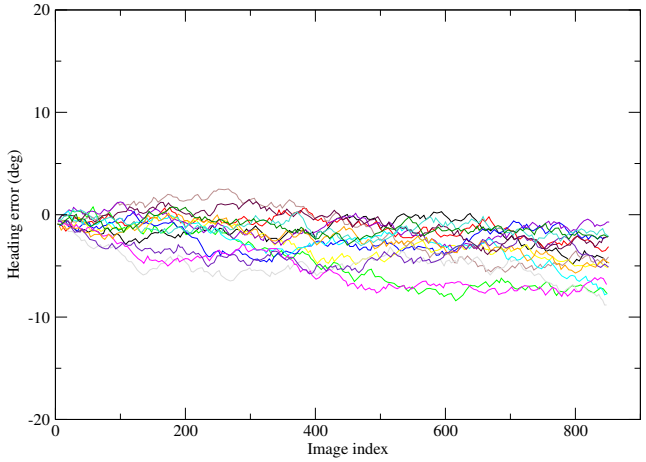
This latter problem is shown in the following experiments. The importance of the change in view point will be shown with another data set, this one not having any such change since the motion is a pure rotation.

Using the same data set, with an angular resolution of 1 pixel per degree, we skip several images in turn to simulate various constant space sampling rates[9]. Figure 18 shows the error between VICON and visual headings for different spatial sampling rates. The reduction in heading error is clearly visible. However, the error still has the annoying property of drifting, if only by 5° over three complete rotations. The improvement is actually mostly due to the fact that less images are processed, thus introducing less error overall, but not necessarily less for each image. Indeed, the rotation per frame is in most cases far from a multiple of 0.5° per frame, thus leading to a systematic error being introduced. This is particularly visible when comparing the results when skipping one frame at a time, producing an average rotation of 2.585° per frame and skipping two frames at a time, producing an average rotation of 3.878° per frame. The error is lower in the former case, close to a multiple of 0.5°, than in the latter case, far from a multiple of 0.5°.

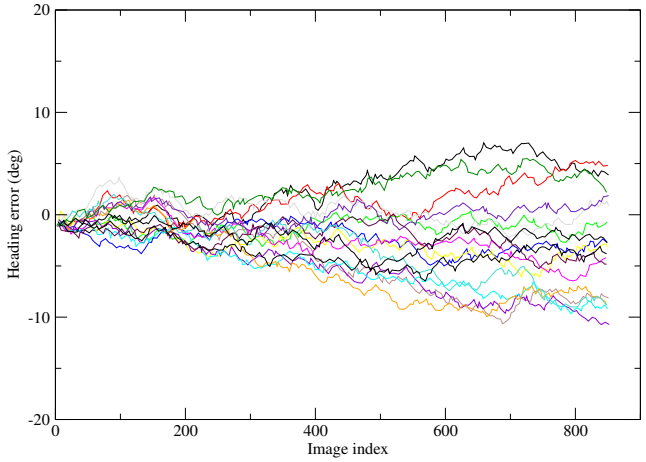
One way of verifying this is to simulate a non constant space sampling by skipping a random number of frames. To do that, a uniformly distributed random number is picked in the interval $[S - s; S + s]$. Figures 19 and 20 show results for various combinations of S and s. It is visible that a variable sampling rate usually improves the performance, compared to using a constant rate. This is however not always the case: for example $(S = 3, s = 0)$ (constant sampling) is better than $(S = 3, s = 1)$. In this case however, the constant sampling was providing a rotation close to a multiple of 0.5° per frame (Figure 18). However, Figure 20 also shows that the results obtained depend greatly on the actual frames that are randomly selected since different runs produce significantly different results, even more so for potentially larger skips. It is also clear that the error remains mostly drifting, although not systematically.

Space sampling is thus an important factor in the performance of the method. In particular, constant sampling should be avoided in the case of pure rotation, unless we can control the rate to be a multiple of 0.5°, which we cannot do without knowing the actual rotation. Random rate performs better but is still not the ideal solution. What would be more interesting is to determine when to use a new appearance of the world by looking at the evolution with time of the distance between appearances. Indeed the motion of the robot will introduce parallax errors at a rate that depends both

[9] The sampling is not exactly constant for reasons previously mentioned, Figure 11 and Table 3.



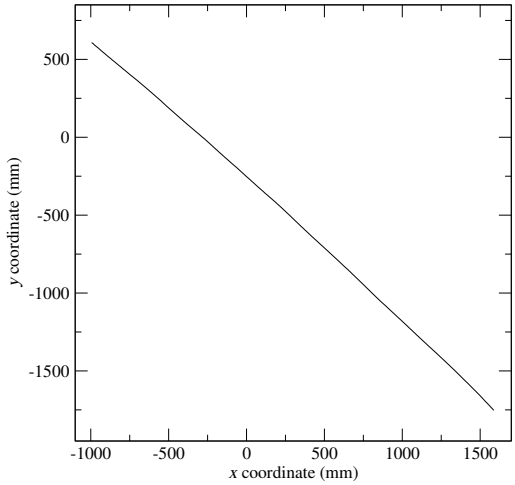
(a) $S = 4, s = 2$



(b) $S = 4, s = 3$

Figure 20: RotateLeft: error between VICON and visual headings for different variable (random) spatial sampling rates for $S = 4$ and two values of s

(a) A view of the trajectory and environment



(b) The actual trajectory

Figure 21: Straight1: the environment and trajectory (start at bottom right)

on the actual motion and on the environment. These parallax errors are what will make the computation of the change in heading unreliable. However, making any sort of assumption about these errors is not realistic and not desirable given our aim. Rather, we must study the evolution of the distance function between appearances and infer from that study when the parallax error becomes too important and thus when a new appearance must be used to compute the change in heading. This is done in Section 3.3.

But before that, we need to study a motion that will introduce parallax errors. We do that with the following data set.

3.2.3 Pure translation: Straight1

In this experiment, the robot is instructed to drive along a straight line from one end of the experimentation area to the other. Figure 21 shows the trajectory and the environment of the robot during the experience. As can be seen, obstacles were put close to the trajectory to create parallax error during the motion of the robot. Two appearances are shown on Figure 22. Due to mechanical imperfections of the robot and irregularities of the floor, the heading of the robot actually changed by $-5.19°$ during the run.

Figure 23 shows the error between the headings captured by the VICON system and computed by the visual compass using the method supporting the theory, i.e. without interpolation, 1 pixel per degree and considering all images. As with the previous case, the error diverges, but to a much lesser extent than previously. This is not surprising since the overall change in heading is also much lower. Looking at the actual rotation undergone between each frame, Figure 24, we can see that the error is due to the fact that the rotation as given by the VICON is always far from a multiple of $0.5°$ and always rounded to 0 by the visual method.

As previously seen, a good way of improving the performance of the method would be to increase the angular resolution of the appearance. However, as Figure 25 shows, the distance between successive appearances presents a sharp minimum that is not widened

Figure 22: STRAIGHT1: two appearances of the environment

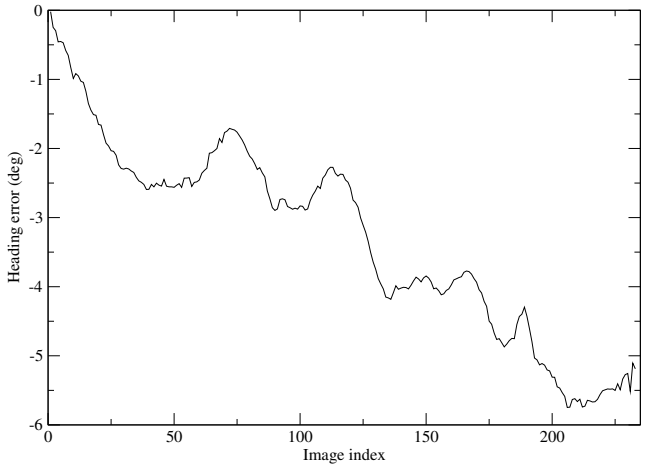


Figure 23: STRAIGHT1: error between VICON and visual headings with all the images

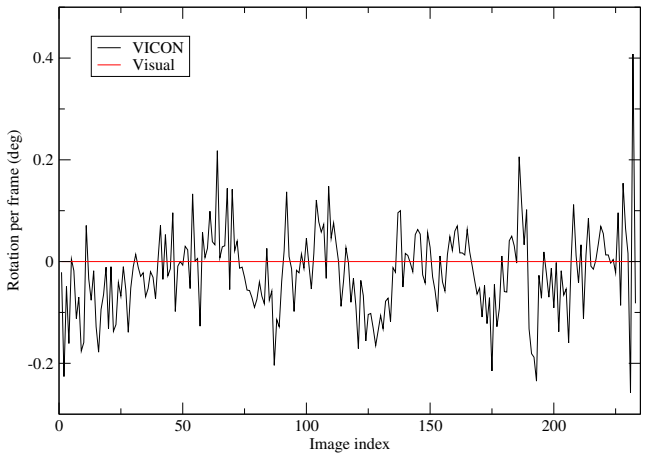


Figure 24: STRAIGHT1: the rotation per frame as given by the VICON system and the visual compass

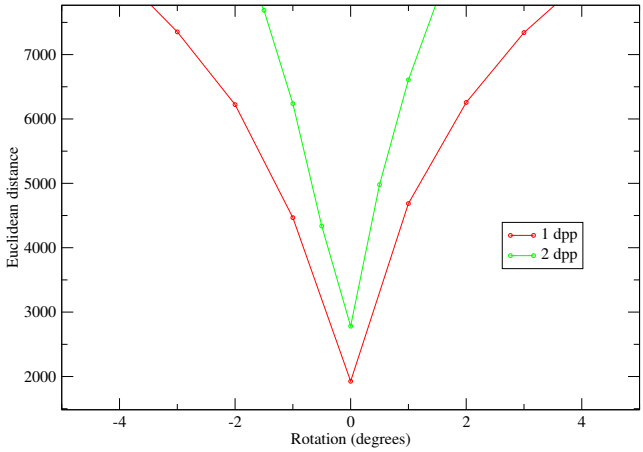


Figure 25: STRAIGHT1: close-up around the minimum of the Euclidean distance between appearances 187 and 188 as a function of the column-wise shift of the second image for two different angular resolutions

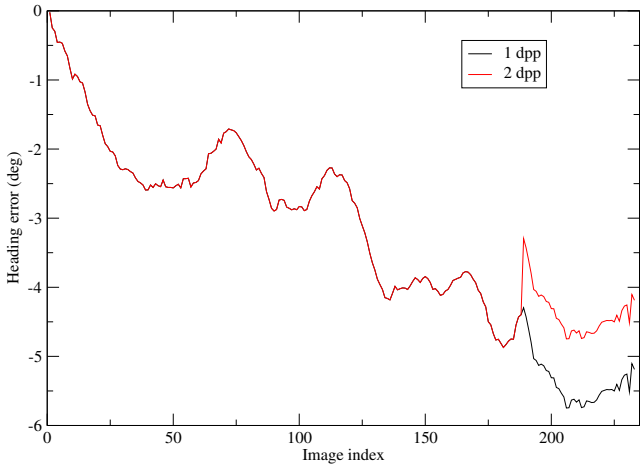


Figure 26: STRAIGHT1: error between VICON and visual headings with all the images for two different angular resolutions of the appearance

by the increase in angular resolution and stays at 0 for the two resolutions. This is because the actual rotation is very close to 0. This implies that increasing the angular resolution will not improve the method, as is confirmed by Figure 26. The only difference between the two curves occurs between frames 188 and 189 because they correspond to images taken just before and after one of the timing glitches of the experiment, which resulted in a more important difference (both in image space and Cartesian space) and more importantly a less marked distance function between the two images, Figure 27. The improvement seen between these two frames is pure coincidence (see below).

The sharpness of the distance function is due to the fact that the differences between successive images are almost exclusively created by pure translation, transformation that potentially renders the images very dif-

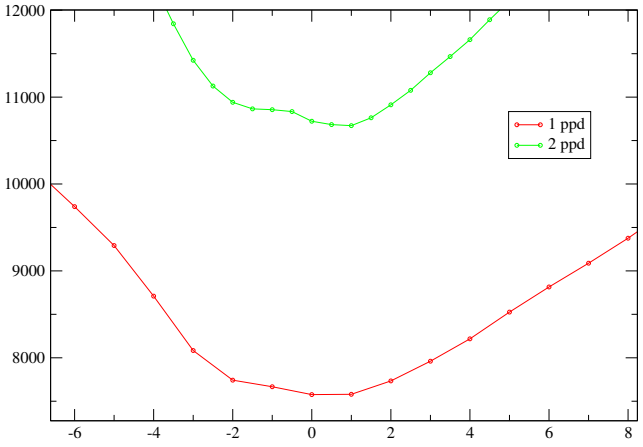


Figure 27: STRAIGHT1: close-up around the minimum of the Euclidean distance between appearances 188 and 189 as a function of the column-wise shift of the second image for two different angular resolutions

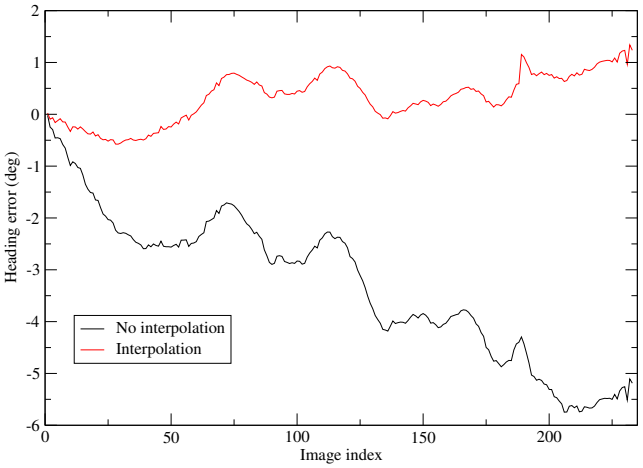


Figure 28: STRAIGHT1: error between VICON and visual headings with all the images and without and with interpolation of the distance function

ferent by introducing parallax error (appearance and dis-appearance of parts of the world are not recoverable by any transformation of the images). However, in this experiment, most translations are small (25 mm on average between successive images), making the images not too dis-similar, hence the sharp distance between them.

As with the previous dataset, another way to improve the algorithm is by interpolating the distance function. As can be seen, interpolation does improve the performance of the method. However, the error still drifts since the measured rotations are systematically within the envelop of the actual rotations, Figure 29. It is interesting to see that the most important difference between the rotations per frame computed without and with interpolation is for frames 188-189 where the rotation computed by the system is opposite to the rotation given by the VICON system. This is a case where the displacement of the robot was larger than normal (16.64 mm between frames 187 and 188 and 183.23 mm between frames 188 and 189) and shows that in the more general case of large displacements of the robot, interpolation is not a good idea. This is obviously dependent on how important the parallax error is, thus of the proximity and/or prominence of objects (in this experiment, the robot was close to large objects around frame 189, Figures 22 (bottom) and 21). This will be discussed further in Section 3.3.

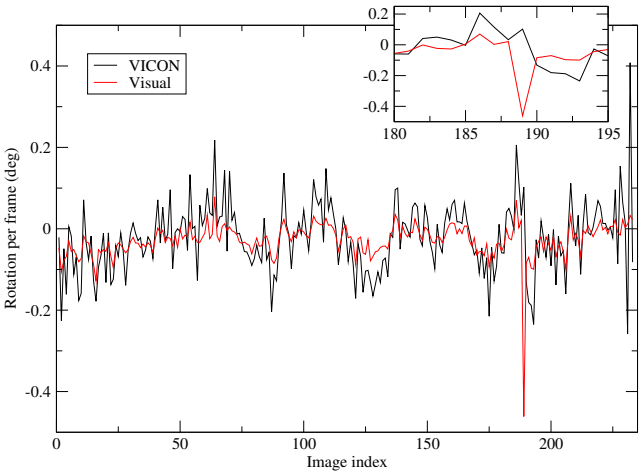


Figure 29: STRAIGHT1: the rotation per frame as given by the VICON system and the visual compass with interpolation of the distance function. Top-right is a close-up of the area around frames 188 and 189, see text.

In the case of an essentially straight trajectory (the case of the dataset presently discussed), one could avoid measuring changes in heading, or at lest lower the space sampling rate. In other words, only images taken while turning could be considered by the method. However, this is not a good idea for two reasons. The first is, as this experiment shows, that one does not necessarily know when and if the robot turns. It is not because the robot has been instructed to move on a straight line that the resulting trajectory is indeed a straight line. The second reason is that even if the robot moves on a straight line, parallax errors introduced by the motion will make (now distant) consecutive images not match well enough to extract meaningful information about the change in heading. It is clear however, as with the previous dataset, that if we lower (compared to what was used for Figure 23) the space sampling rate, the performance of the method should increase. This is first because less measurements will be made, thus integrating a lower number of errors and second because, in the case of this data set, small rotations will become more important and thus measurable.

Figure 30 shows the error between VICON and visual headings for different spatial sampling rates. As

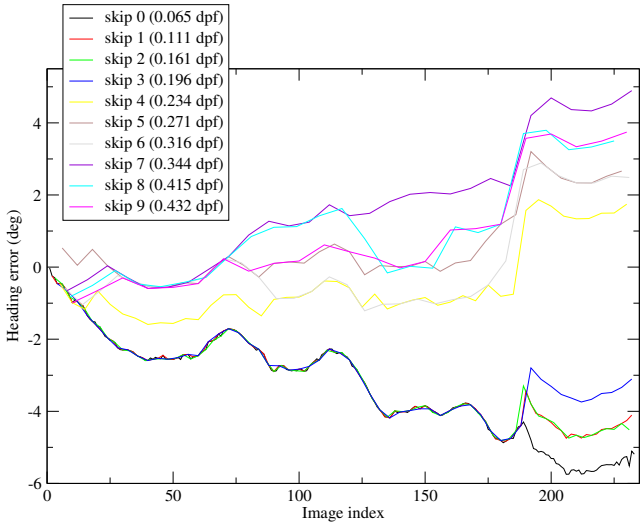


Figure 30: STRAIGHT1: error between VICON and visual headings for different constant spatial sampling rates

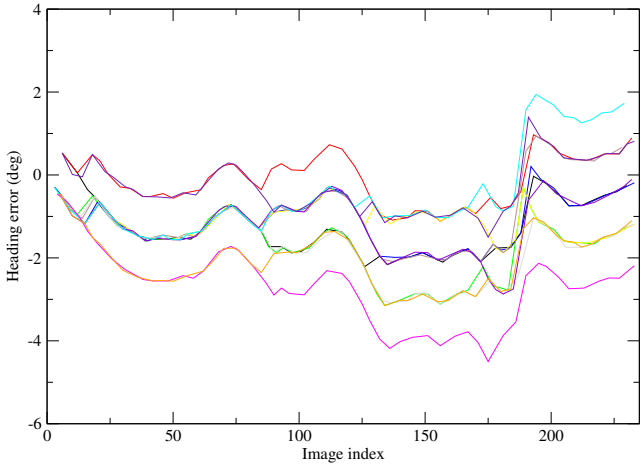


Figure 31: STRAIGHT1: error between VICON and visual headings for different variable (random) spatial sampling rates for $S = 4$ and $s = 2$

can be seen, skipping four images results in a significant improvement (at least before frame 189) over skipping less images. On the other hand, the performance deteriorates significantly when skipping seven or more. This shows that there is a range of sampling rates in which the performance is better and stable, which is in accordance with what we said earlier:

- below the range, too many errors are integrated;
- above the range, the errors become too large.

Obviously, the performance depends on the sampling and randomising it produces variable results, Figure 31.

It is interesting to note that all curves have the same overall shape on Figures 30 and 31 and that they mostly

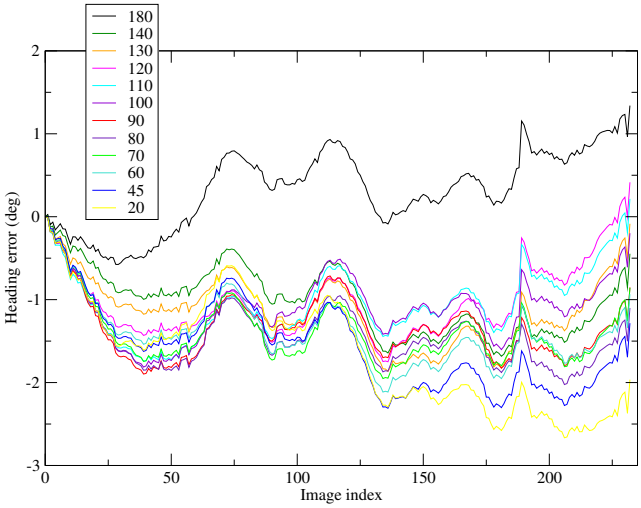


Figure 32: STRAIGHT1: error between VICON and visual headings with all the images and interpolation of the distance function for different widths of the front and back field of view

differ is a few distinct places. This shows that it is important to consider carefully which exact images to use. We will do this in Section 3.3.

Another improvement that can be done over the theoretical method is to only consider parts of the images. Indeed, as noted in [Nelson and Aloimonos, 1988], the contribution to the optic flow by the motion of the robot is not homogeneous in omni-directional images; the forward/backward translation mostly contributes in the regions corresponding to the sides of the robot and very little in the parts corresponding to the front and back of the robot while the rotation contributes equally everywhere.

Because we are interested in extracting the rotation information, only considering the regions of the images corresponding to the front and back of the robot allows us to discard most of the problems introduced by the translation, in particular sudden changes in appearance (parallax).

Figure 32 shows the error between the VICON and visual headings when restricting the computation of the distance (with interpolation) between panoramic images to columns in $[90-r; 90+r]$ and $[270-r; 270+r]$ (for appearances with 1° per pixel), with $2r$ being the angular field of view of the front/back regions, for different values of the field of view. The performance of the system is not dramatically better in this case, if not worse, compared to only interpolating the distance measurement. Moreover, no single value is overall better than any other. It is however not surprising that the performance does not improve with this dataset because, as stated above, the measurement errors in rotation are large compared to the rotation itself[10]. We thus study

[10] Moreover, it is not certain that some of the measured

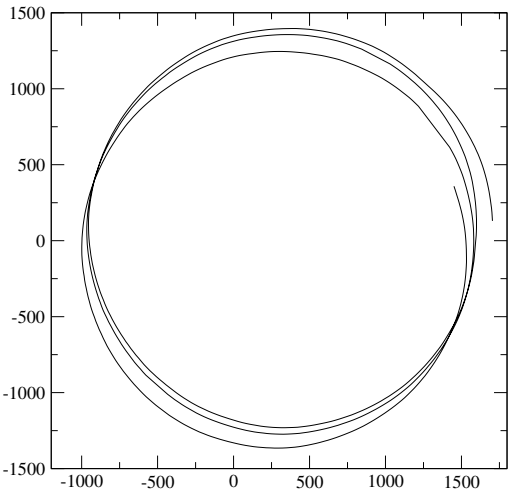


Figure 33: CIRCLE1: trajectory

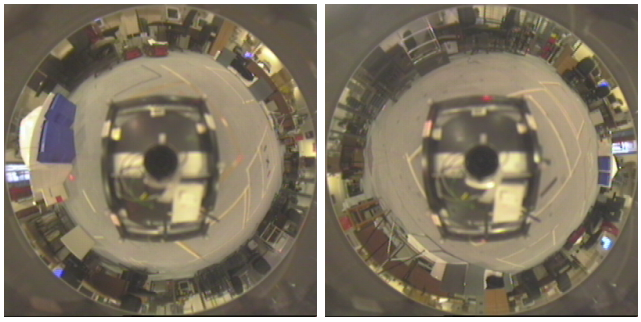


Figure 34: CIRCLE1: typical grabbed images

another dataset that introduces a different, more realistic type of motion.

3.2.4 Rotation and translation: CIRCLE1

This dataset was grabbed while the robot was translating and rotating at almost constant speed, thus moving along a roughly circular trajectory. Three complete revolutions were performed producing a total rotation of 1097.285°. Figure 33 shows the path followed by the robot while Figure 34 shows typical grabbed images of the area. In particular, it can be noticed that the robot was going across a mixture of fairly free areas and areas with a number of obstacles (approximately 1 m between the camera and the obstacles).

As for the other datasets, we plot the error in heading between the VICON system and the visual method using the variations on the theoretical method as explained above.

The first variation is on the angular resolution of the panoramic images. Figure 35 shows the error between the VICON and visual headings with all images, no interpolation, for various angular resolutions of the panoramic images. Increasing the resolution dramatically improves the performance, but only up to the maximum resolution the omni-directional images can support, i.e. 3.49 pixels per degree, as previously mentioned. However, this improvement is computationally expensive. Moreover, the error is clearly drifting.

error is not due to non-homogeneous coverage of the area by he VICON system, Section 3.2.1.

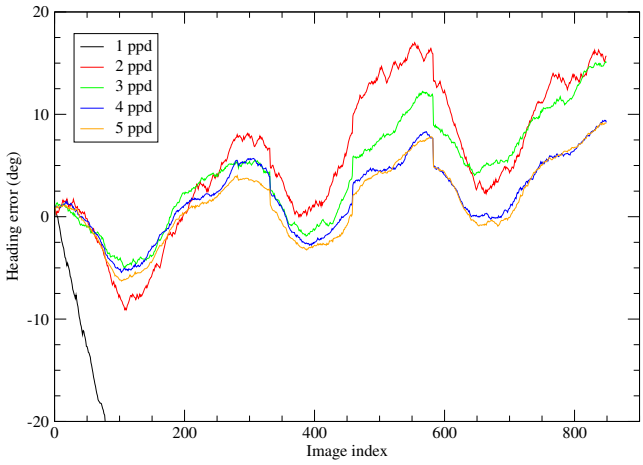


Figure 35: CIRCLE1: error between VICON and visual headings for different angular resolutions of the panoramic images

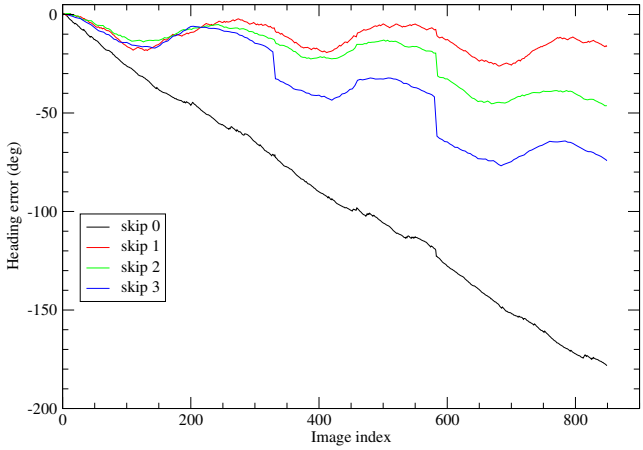


Figure 36: CIRCLE1: error between VICON and visual headings for different constant spatial sampling rates (without interpolation of the distance function)

Figure 36 shows that skipping some of the images is better than using them all but also that the performance deteriorates rapidly with the number of frames skipped. However, Figure 37 shows that the improvement between considering all frames and skipping one is not due to an improvement of individual measures (on the contrary) but rather to the fact that larger errors are integrated less often. This shows again the trade-off between better estimation of the rotation and frequency of estimation.

As before, we look at the effect of interpolating the distance function between images. Figure 38 shows the error between the VICON heading and the visual head-

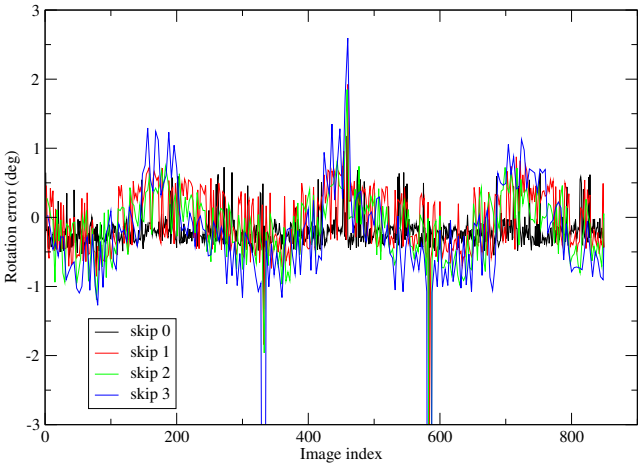


Figure 37: CIRCLE1: error between VICON and visual rotations for different constant spatial sampling rates

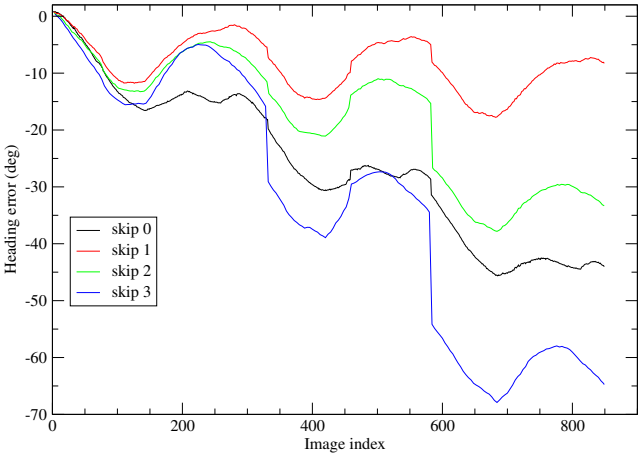


Figure 38: CIRCLE1: error between VICON and visual headings for different constant spatial sampling rates with interpolation of the distance function

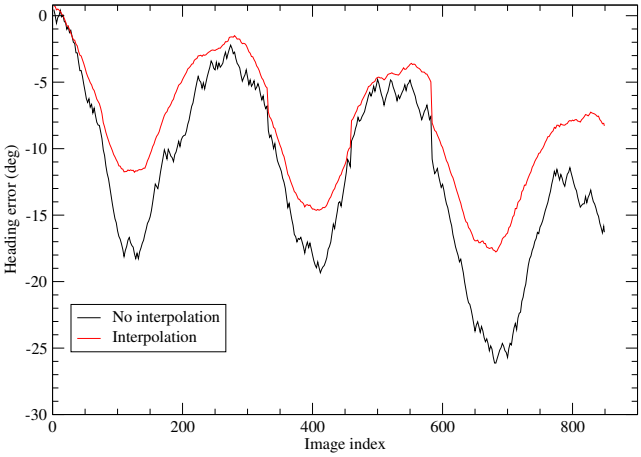


Figure 39: CIRCLE1: error between VICON and visual headings when skipping one frame, without and with interpolation of the distance function

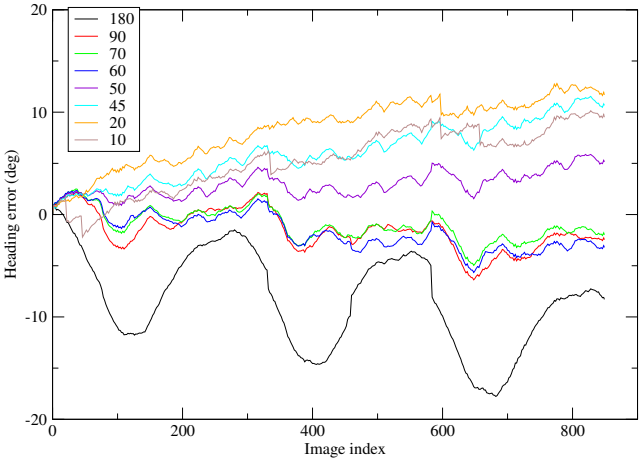


Figure 40: CIRCLE1: error between VICON and visual headings when skipping one frame with interpolation of the distance function for different widths of the front and back field of view

ing. It is clear that interpolation dramatically improves the performance, even over the best result obtained without interpolation, Figure 39.

The important oscillations visible on the error curves are due to several areas in the images (corresponding to boxes, an open door, monitors and polystyrene blocs) all concentrated on one side of the environment of the robot (visible in either side of the robot on Figure 34) having saturated blue and white pixels. These boxes compared to the overall unsaturated grey-ish colour of the remaining of the environment create a marked asymmetric feature in the environment that make the system believe it is either turning more or less than it actually does, depending on the feature being on one side of the robot or the other (see Section 3.1.3 for a discussion). To limit this effect, we restrict the distance measurement between images to the field of view corresponding to the front and back of the robot. This is shown on Figure 40 where one can see a clear improvement for a width of the field of view in the range $[60°; 90°]$ but a performance gradually going down for narrower fields of view. As expected, the effect of the lateral strong features diminishes with the width of the field of view to the extent of not being visible anymore for very narrow fields of view. However, when this happens the performance is notably worse because very little of the environment remains available for the distance measurements.

3.2.5 Other possible improvements

We have chosen to use a linear mapping between pixels of the omni-directional and panoramic images (see Section 3.1.1). Different types of mapping can be used. For example, giving more importance to the pixels at the periphery of the omni-directional images could be a good idea because these pixels will typically correspond

to features that are far away from the robot, thus more stable. However, these pixels also correspond to tall objects that are near the robot, which does happen often given that robots are often not taller than objects in their environment. Moreover, the results we obtained indoors in our lab where not as good in that case. This is due to the fact that the background is cluttered, unlike the area in which the robot is moving, making the distance function more noisy and thus the minimum not as well marked. For the same reason, we keep as much as possible of the "doughnut" in the omni-directional images.

We have mentioned that the unwrapping uses the nearest neighbour pixel. Using bi-linear interpolation is possible and certainly produces smoother images. However, this involves more processing and we have shown that this does not improve the performance of the system [Labrosse, 2004]. Similarly, images could be blurred, which would result in a smoother distance function and a wider minimum, which could be desirable for the minimisation. However, this is again more processing and it has been shown in the past that blurring images tends to move features in the images (e.g. [Marr, 1982]) which would introduce a bias in the rotation estimation. Moreover, our minimisation procedure has shown that it never fails to find the global minimum (especially if the heading is computed often enough, Section 3.3, but also note Figure 41).

In all the experiments reported here, we use the Euclidean distance to measure the similarity between images. We have mentioned above the possibility of using the Manhattan distance instead but previously published results do not show any difference in performance between the two [Mitchell and Labrosse, 2004]. We have also shown that the combination of the Euclidean distance with the RGB colour space is not optimum. More work needs to be done on that front.

3.3 The algorithm

In the previous sections, different parameters of the system have been evaluated independently and it can be seen that some produce better results than others. It remains to determine what the best value for these parameters are and to then devise an algorithm that will provide good heading measurement. This is done now.

We have seen that increasing the angular resolution of the panoramic images does dramatically improve the performance of the system. However, this also introduces an important drawback: heavier computation. Indeed, increasing the resolution implies increasing the size of the grabbed images and thus more processing at the unwrapping and distance measurement stages. Given that estimating headings is only one of the tasks the robot has to perform when navigating, we will not use higher angular resolutions than one pixel per degree in the panoramic images. The panoramic images need only be 360 pixels wide and the grabbed images

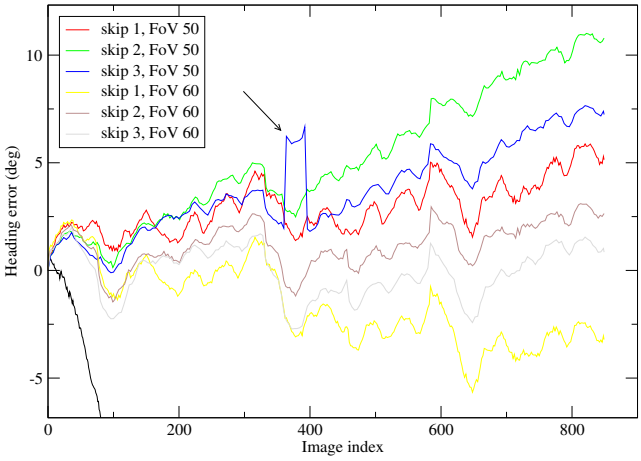


Figure 41: CIRCLE1: error between VICON and visual headings for different constant spatial samplings with interpolation of the distance function and various widths of the front and back field of view. Arrowed is a problem with the minimisation, see text.

200×200 pixels.

Interpolation of the distance function also improved the performance of the system. Moreover, the simple parabolic interpolation is not computationally expensive, although we will see below that it is sometimes not enough. We will thus use a revised version of the interpolation (see below).

Reducing the field of view to the regions corresponding to the front and back of the robot improves significantly the performance and has the advantage of also reducing the amount of computation to perform. We have seen that the performance is stable with variations of the width of the field of view. Figure 41 shows the error in headings for various skips and widths of the front and back field of view. Although the different results are not very different (all with an error below 10% fo the total rotation), a field of view of 60° with a skip of 3 provides the best result. Such a value for the field of view is a good compromise between loosing information (if details in the environment are only present on the sides of the robot) and reducing computation and parallax errors. Moreover, reducing the field of view also narrows the peak in the distance function, which improves the speed of the minimisation but can also make it fail (a case is arrowed on Figure 41). This will be discussed further below.

A parameter that is more difficult to specify is the space sampling since previous experiments have shown that the performance of the system depends largely on it. Moreover, we have seen with the datasets presented so far that the optimum value varies within datasets (see for example Figure 20 on page 13) and between datasets. This is not surprising since, as we have mentioned before, there is a trade-off between sampling often, thus introducing less error each time on the

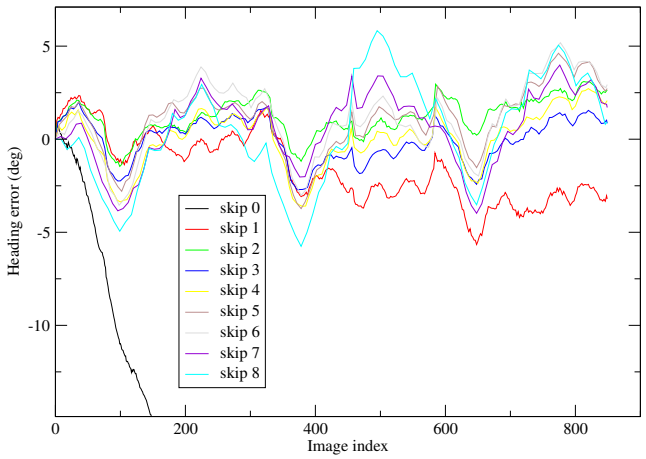


Figure 42: CIRCLE1: error between VICON and visual headings for different constant spatial samplings with interpolation of the distance function and a field of view of 60° front and back

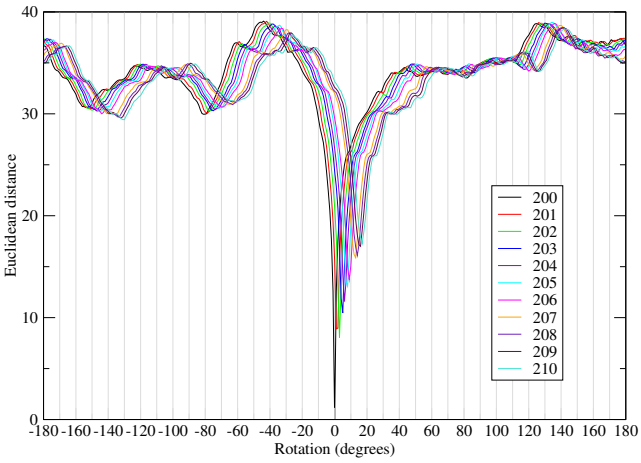


Figure 43: CIRCLE1: the Euclidean distance between image 200 and images 200 to 210 as a function of the column-wise shift of the second image of the pairs

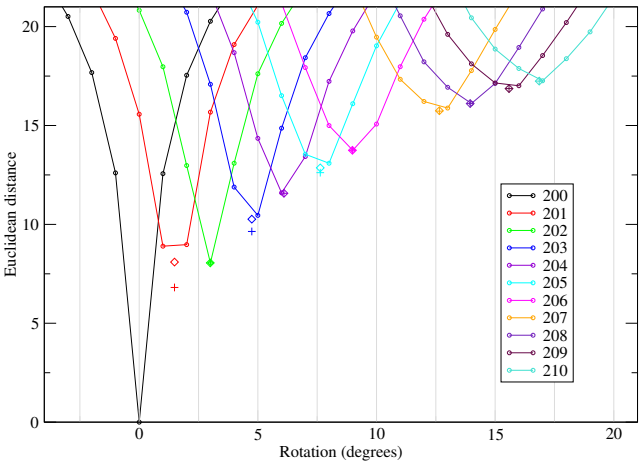


Figure 44: CIRCLE1: close-up of Figure 43. The '◇' shows the interpolation using parabolic interpolation (Equation (3) page 3) while the '+' shows the mixture of linear and parabolic interpolation (see text).

grounds that less parallax error introduces less measurement error, and accumulating less often un-avoidable errors, implying a low sampling rate. Fortunately, as Figure 42 shows, the performance is not as dependent on the sampling rate when using a narrower field of view (compare with Figure 38). However, it is clear that choosing a good value is necessary since, although the error does not significantly drift in any case, it is increasingly variable with lower sampling rates for this dataset. It could be tempting to simply use one frame out of three grabbed (or adjust the frame rate to an equivalent rate) since this is what gives the best performance. However, this is not satisfying as the ideal rate obviously depends on the speed of the robot and the variability of the environment (a mixture of closeness of the obstacles and variability). We thus need to match the spatial sampling rate to the grabbed images and their evolution.

The only information we have about the evolution of the images is the distance between successive images. Figure 43 shows the distance between a reference image and successive following images as a function of the column-wise shift of the second image of the pairs. It is clear that the top parts of the curves are all similar (in fact differ mostly by a horizontal shift due to the rotation of the robot between the different frames). The bottom part of the curves changes smoothly in two ways: a horizontal shift shows the rotation of the robot[11] while an increase of the minimum shows increasing parallax error between frames. This is even more visible on Figure 44. However, the distance minimum is not a monotonous function of image index (or space sampling), even when the result of the parabolic interpolation is used ('◇' on Figure 44). Close inspection of the distance function between, e.g., frames 200 and 201 show that actually the parabolic interpolation does not provide a good estimate of the height of the minimum (although the angle provided by it is good). This is visible on Figure 45 that shows, for images 200 and 201 the distance using one and two pixels per degree (the later has been scaled so that it matches the former at a rotation of 1°). However, note that for other pairs, the parabolic interpolation provides a good estimate of both the rotation and the value of the minimum. For images 200 and 201, the curve for two pixels per degree has a minimum close to the real minimum of the function (because of the near-symmetry of the function). This minimum would be better approximated by the intersection of the extrapolation of the two values before and after the minimum (dashed lines on Figure 45). This is even truer for pairs of images very close

[11] The larger horizontal gap between images 206 and 207 corresponds to one of the time glitches.

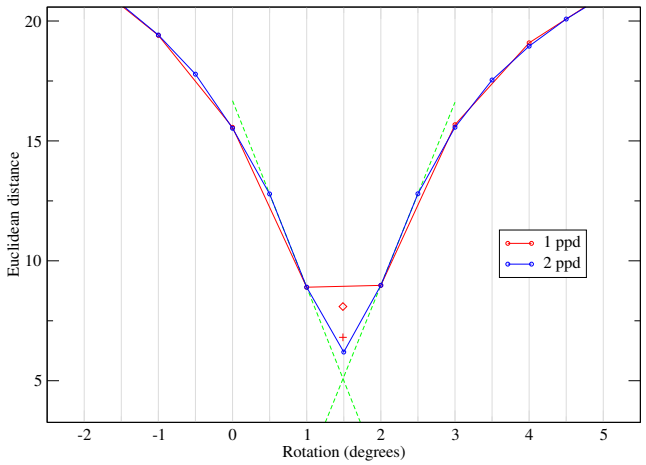
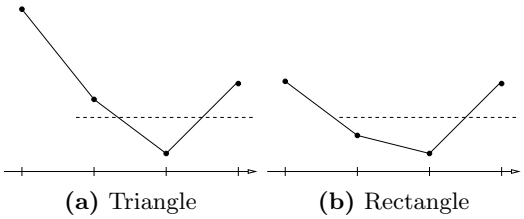


Figure 45: CIRCLE1: close-up of the distance between images 200 and 201 using angular resolutions of 1 and 2 pixels per degree. The '◇' shows the interpolation using parabolic interpolation (Equation (3) page 3) while the '+' shows the mixture of linear and parabolic interpolation (see text).



(a) Triangle **(b)** Rectangle

Figure 46: Two possible configurations of the minimum of the distance between images as a function of the column-wise shift of the second image

to each other (results not shown here but visible when one interpolates between the case of the pair 200-200 and 200-201 on Figure 44).

We thus distinguish two cases. The first is when the configuration of the bottom points of the distance function form a triangle. This is when the lower of the two points on either side of the minimum is above the mid-height between the minimum and the point on the other side of the minimum (Figure 46(a)). In this case, a parabolic interpolation is used. The second case is when the two lower points are roughly at the same height, i.e. the four lower points are in an almost rectangular configuration (Figure 46(b)). In this case, a parabolic interpolation is sometimes better (see for example the distance between images 200 and 209 on Figure 44). On the other hand, other cases are better modelled with the linear extrapolation (Figure 45). The difference between these two cases is the amplitude of the distance function (difference between maximum and minimum values) relative to that of the distance between the reference image and itself. Based on our observations of the evolution of that amplitude

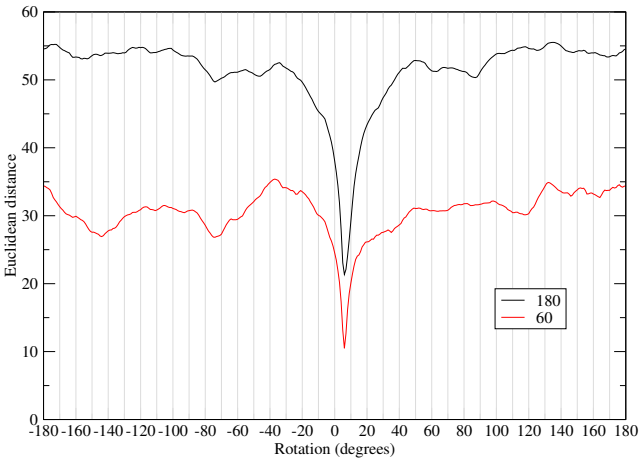


Figure 47: CIRCLE1: distance between images 200 and 204 for two widths of the front and back field of view

for many cases, we decided to use a ratio of 0.5; when the amplitude of the distance function between images i and j is below half of the amplitude of the distance between image i and itself, a parabolic interpolation is used. Otherwise, a proportion of the linear extrapolation and of the parabolic interpolation is used (in the same proportion of that between the amplitudes). This is to avoid a sudden change of the sub-pixel height of the distance function. The result of this procedure is shown with the symbol '+' on Figures 44 and 45. One can see that this method gives monotonicity of the height of the minimum, and thus of the sub-pixel amplitude.

Note that this procedure is not very expensive given that most of the needed values are computed anyway during the minimisation of the function and that the configuration of the points allows us to simplify to the extreme all the interpolation and extrapolation computations.

We can now turn back to determining the spatial sampling rate. For this, we measure the amplitude of the distance as a function of the column-wise shifts for all pairs of images of the dataset that correspond to skipping three frames, the best sampling rate for this dataset[12]. To compute the amplitude, we need to obtain both the minimum and the maximum of the distance function. The minimum can be obtained by exhaustive search. However, this is expensive. Instead, we perform a local minimisation that, if the rotation between tested images is not too important, will be the global minimum. However, when the width of the field of view is made narrower, the distance function becomes less regular, presenting a narrower peak and many local minima, e.g. Figure 47. This implies that a better minimisation procedure must be used. We perform a

[12]The pairs of images containing the timing glitches have not been used in these because they introduce uncharacteristic elements in these measurements.

Table 4: CIRCLE1: statistics on the relative amplitude of the distance for the best value of the skip (3) and the following with a field of view of 60°

Skip	Average	Std. Dev.	Median	Min.	Max.
3	0.6413	0.0486	0.6430	0.4881	0.7617
4	0.6053	0.0541	0.6055	0.3448	0.7284

combination of a local search and a global search: the local minimum is first sought for using a simple descent algorithm, which in most cases is successful and does not need more than $r + 1$ evaluations of the distance, where r is the rotation between the two frames. From this local minimum, a number of positions on both sides of it are evaluated in turn. If any of them is lower than the previous minimum, then the local minimisation is started again from that position. In all experiments reported in the following sections, eight values spanning a rotation of 20° centred on the local minimum were evaluated. If any of these values is indeed lower, then the previous local minimisation only had to evaluate a small number of values given the shape of the function, if the initial starting position was not too wrong. The algorithm described below ensures that this is the case and in particular solves failures such as the one arrowed on Figure 41. This minimisation thus usually does not need a large number of evaluations of the distance function (generally no more than the amount of degrees of the rotation between successive images).

The maximum value of the distance is taken as the value corresponding to a rotation of 180° from the minimum. This does not necessarily correspond to the absolute maximum of the function, especially when the field of view does not cover the whole image, but does correspond to the theoretical least correlation between the images. Moreover, as seen on Figure 43, this value is stable when the number of skipped images increases.

Table 4 gives some statics on the relative amplitude of the distance between all pairs of images corresponding to values of the skip of 3 and 4 and with front and back field of view of 60°. The relative amplitude is obtained by normalising it to that of the amplitude between the first (reference) image of each pair and itself. The amplitudes are normalised because they depend on the values of the pixels of the images (that themselves depend on factors such as illumination and colour of the environment, sensitivity of the camera, etc.). The statistics are gathered for the best value of the skip and the following because if the amplitude is used as a threshold, then any value between that corresponding to the best skip and the following for a particular reference image is acceptable. Despite an apparent large range of amplitude values between the minimum and maximum values, the distribution is fairly compact as shown by the standard deviation. Moreover, because the average and median values are close, the distribution is well balanced.

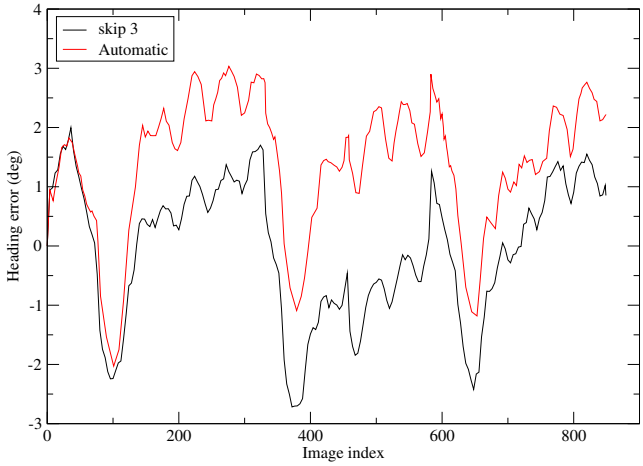


Figure 48: CIRCLE1: error between VICON and visual headings for spatial sampling based on the best skip and on the threshold on the relative amplitude

The idea of the algorithm is as follows. From a reference image for which a heading is known, a frame is grabbed and the relative amplitude of the distance between the reference image and the grabbed image is computed. If the value is above the threshold, then the grabbed image is discarded because judged too close (in Cartesian and image space) to the reference image. On the contrary, if it is below the threshold, then it becomes the new reference, after having computed the change in orientation between the two images and updated the current heading. The process is then repeated with the new reference image and heading. In practise, because an amplitude below the threshold means that the image is already too distant, the previous image is used to update the heading and kept as the next reference. In slowly changing appearances, this leads to infrequent heading measurements. Because one might want frequent heading updates and because the minimisation procedure works better from a good estimate of the position of the minimum, a current heading is computed from the reference image and heading using every grabbed image but the reference heading is only updated when the reference image changes. This way, an estimate of the heading is available at all times.

Because of the way the threshold is used, we want to use the median amplitude corresponding to the best skip plus one (i.e. corresponding to a skip of 4 for the CIRCLE1 dataset); we will thus use the value 0.6055 for all subsequent experiments, see Table 4. Figure 48 shows the error between VICON and visual headings for the best fixed skip and the skip determined using the threshold on the relative amplitude of the distance for the CIRCLE1 dataset. Although the algorithm does not perform quite as well as the fixed sampling, the difference is small: 1° at the end of a 1097.285° rotation (as measured by the VICON system).

Table 5: The parameters of the visual compass algorithm

Angular resolution of the appearance	1° per pixel
Size of the appearances	360 × 45 pixels
Distance interpolation	mixture of parabolic interpolation and linear extrapolation
Front and back field of view	60°
Distance minimisation	local + regular scanning
Space sampling	threshold on the relative amplitude: 0.6055

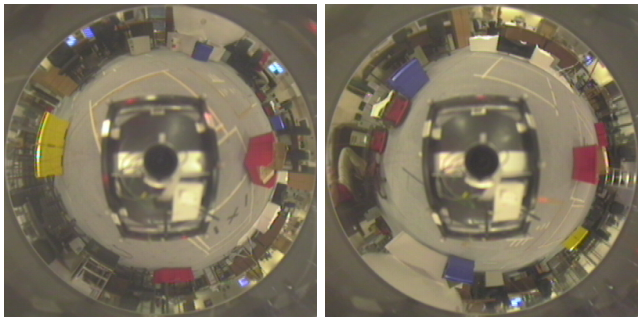


Figure 49: Two omni-directional images of CIRCLE2

3.4 Experiments

In this section, we apply the algorithm with the parameters estimated and evaluated in Section 3.3 to different datasets to evaluate the performance of the visual compass. The parameters are recapitulated in Table 5.

3.4.1 The datasets

The datasets cover an indoors environment (some of these datasets have been described above) and two outdoors environments. The indoors datasets have all been grabbed in the same environment but with different trajectories and/or different objects visible in the images:

- ROTATELEFT: a pure rotation (on the spot), a few objects surrounding the robot (see Section 3.2.2);
- STRAIGHT1: a pure translation, a few objects close to the robot (see Section 3.2.3);
- STRAIGHT2: a pure translation, no object in the vicinity of the robot (but at the extremities of the trajectory);
- CIRCLE1: a circular trajectory, a few objects outside the trajectory, some close to it (see Section 3.2.4);
- CIRCLE2: a circular trajectory, many colourful objects outside and inside the trajectory. In particular, a bright red box at the centre of the trajectory appears almost static in the images (on the right of the omni-directional images of Figure 49).

Headings provided by the visual compass are evaluated against headings provided by the VICON system.

The ground truth for the outdoors datasets is obtained using a magnetic compass, since the VICON system is not available outdoors. The first two datasets are only used to evaluate the performance of the magnetic compass:

- PAVEMENTSTRAIGHT: a pure translation on a straight pavement between a road and a grassy area, no close object apart from the author being static behind the robot at its starting position;
- GRASSSTRAIGHT: a pure translation on the grass by the pavement used in experiment PAVEMENTSTRAIGHT;
- GRASS1: a random trajectory in a grassy area surrounded by bushes;
- GRASS2: a random trajectory on a grassy area surrounded by bushes and under trees.
- CARPARK: a random trajectory on a flat but rough surface in a car park almost empty of cars;

The datasets acquired on the grassy area all contain numerous frames showing moving cars and people.

3.4.2 Results

For each dataset, the heading as a function of image index is computed using the algorithm described in Section 3.3 using the parameters recapitulated in Table 5 ("Atm" with the threshold used, 0.6055 being displayed as 0.6) and compared with the best result (lower error at the end) obtainable with a fixed space sampling ("Fx" with the number of skipped frames). The results are evaluated using various measures. The maximum of the absolute value of the error and the mean and standard deviation of the error are used to assess the variation of the error on each dataset. The slope of the robust linear regression of the error as a function of time and distance travelled is used to evaluate the trend of the error. The distance travelled is not always applicable (in ROTATELEFT) or available (the VICON system not being available outdoors and not having a GPS, outdoors experiments don't contain this information). All datasets have been acquired with the robot driving at similar speeds (although not exactly the same speed due, for example, to difficulties to drive in grass). This means that the slopes as a function of time are roughly

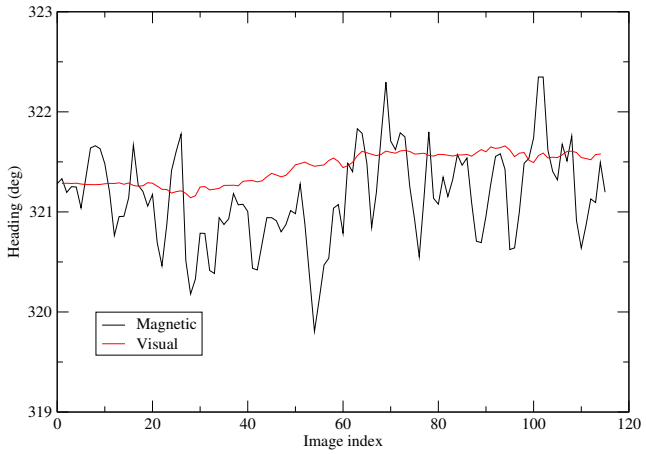


Figure 50: PAVEMENTSTRAIGHT: magnetic and visual heading with all the images

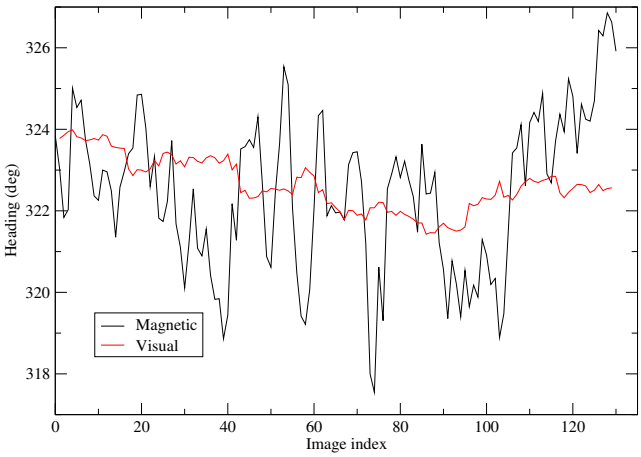


Figure 51: GRASSSTRAIGHT: magnetic and visual heading with all the images

comparable. In all cases, the number of reference frame is given with the total number of frames in parenthesis). All the results are given in Tables 6 and 7.

Table 6 shows no surprise in the results. In all cases, the error remains low (less than 7°) and in the more general case of combined translation and rotation the error's drift also remains low (about 0.2°/m in the difficult case of CIRCLE2 where a bright object remains static in the images — a red box at the centre of the trajectory). The manual method (fixed spatial sampling rate) performs systematically slightly better than the automatic method. However, comparable results are obtained for all the datasets with parameters determined from one of these sets, the datasets being different both in the trajectory and the environment. The automatic method for STRAIGHT1 performs significantly worse than the manual method and compared to the other datasets. This is because the environment does not vary much when seen with the reduced field of view and thus large numbers of frames are skipped, introducing an important error due to parallax. A similar example is GRASS2 (below).

Figures 50 and 51 show the heading of the robot as returned by the magnetic compass and computed by the visual compass using all the images. The visual heading is much smoother than the magnetic heading. Moreover, the magnetic compass provides a heading that is less smooth on the bumpy area of the grass than on the smooth pavement. Tis is because the magnetic compass is sensitive to yaw and tilt. This is obviously good for the stability of programs that might use such information to control the robot.

To capture these datasets, the robot was given constant forward speed and null rotational speed command and was positioned on the straight pavement by the grassy area used in these experiments or on the grassy area by the pavement. It was thus easy to see that the actual trajectory was in fact not straight, although no actual precise measurements were taken. This was even truer for the run on grass, due to the slippery and bumpy surface. For both datasets, the visual heading seems to drift. However, close inspection of the graphs show that the overall tendency of the magnetic heading behaves in the same way. This shows that the divergence is in fact due to the small amounts of rotation the robot undertook. This experiment shows that errors of about 5.5° can be due to the noisy behaviour of the magnetic compass.

It is interesting to note that frames 100 to 106 of GRASSSTRAIGHT contain a moving car that was large in the images and brighter than the rest of the pixels on the images, Figure 52. This has no effect on the visual compass because when the car was at its largest in the view, it was out of the narrowed field of view used for the heading computation.

The experiments with the outdoors datasets show interesting problems as well as a generally good performance. The statistics shown in Table 7 for the datasets GRASS1 and GRASS2 (starred) were gathered only on a subset because of these problems, to get statistics showing the "normal" performance of the system without the problems shown here.

Figure 53 shows the error between the magnetic and visual headings for the dataset GRASS1. As can be seen, up to about frame 700, the error remains low (less than 20°) but jumps to a high value in a few frames (Table 7 shows the statistics taken up to frame 700). This is due to a combination of several effects: a not very fast rotation, a bump in the ground shifting the camera so that the visual effect of the rotation was attenuated and a bland and featureless environment, especially in the field of view kept for the computation, Figure 54.

Figure 55 shows the error between magnetic and visual headings for the dataset GRASS2. At first sight, the performance with this dataset seems much worse than with the others. This is due to a combination

Table 6: Quantitative evaluation of the performance for the different indoors datasets

Dataset	Method	Nb (total)	Max (°)	Mean (°)	Std. dev. (°)	Slope (°/s)	Slope (°/m)
RotateLeft	Fx 11	72 (853)	1.79800000	-0.24539069	0.86898691	0.00963005	n/a
	Atm 0.6	25 (853)	2.12040000	-0.95199560	0.56033512	-0.00496272	n/a
Straight1	Fx 1	118 (234)	1.73656901	-1.00399279	0.42532528	0.00432601	0.07255171
	Atm 0.6	15 (234)	6.47863117	1.91005913	2.04736241	0.10216857	1.69372456
Straight2	Fx 1	120 (238)	3.23443161	1.05001711	1.06265287	0.05922039	0.99584576
	Atm 0.6	13 (238)	3.99161069	0.97141982	1.53962810	0.06773449	1.13329002
Circle1	Fx 3	214 (851)	2.71730019	-0.05506429	1.17721532	-0.00032850	-0.00285344
	Atm 0.6	204 (851)	3.03350294	1.58213383	0.99226045	0.00231989	0.01966643
Circle2	Fx 8	94 (837)	5.00147745	-1.58073909	2.05148765	-0.01566614	-0.13296038
	Atm 0.6	184 (837)	6.87676737	-2.92980200	2.04042571	-0.02510505	-0.21267754

Table 7: Quantitative evaluation of the performance for the different outdoors datasets (subsets for starred sets)

Dataset	Method	Nb (total)	Max (°)	Mean (°)	Std. dev. (°)	Slope (°/s)
Grass1*	Fx 0	701 (701)	16.61526370	4.33354863	5.64275273	-0.01032254
	Atm 0.6	398 (701)	17.51966342	4.46453569	7.05506766	-0.06312815
Grass2*	Fx 2	349 (1043)	65.45388912	-15.45585612	20.15729050	-0.15617237
	Atm 0.6	259 (1043)	77.26036245	-30.74021243	21.69265574	-0.22989795
CarPark	Fx 3	216 (860)	36.65223503	2.58878988	13.66782165	0.06901577
	Atm 0.6	196 (860)	31.98683776	4.36235851	13.49039756	0.07723853

of problems. The first one happens between frames 51 and 54, shown on Figure 56. As in some previously shown cases, a bright car is passing by the robot. This time however, the car is in the field of view. Moreover, the other half of the field of view (front) is almost featureless, resulting in a wrong estimation of the change in heading. The graph obtained when skipping three frames at each measurement shows that if these frames are indeed avoided, then the problem does not happen. Figure 57 shows the distance as a function of rotation between frame 51 and frames 51 to 53, when the car was passing in the field of view. The curves are qualitatively different from the "normal" curve, e.g. the one for image 51 and itself. In particular, the global minimum of the distance does not correspond to the rotation of the robot. They do however correspond to pushing the car out of the field of view. Figure 58 shows the panoramic images corresponding to frames 51 to 53, frames 52 and 53 having been un-rotated according to the global minimum of the distance to frame 51. It is clear that the car is indeed "pushed" out of the field of view. The distance function from frame 51 to frames 52 and 53 does present two marked minima that correspond to pushing the car either way of the field of view, the higher one being for when the car "moves" most.

Note that the minimisation procedure as described in Section 3.3 will not reach the global minimum of the distance between images 51 and 53, Figure 57 but rather the local minimum close to 0°, which does correspond to the true rotation between the two frames. However, because the minimum is much higher than what it should be (because overall the second image is very different from the first because of the car), the relative amplitude goes below the threshold and the change in heading is thus calculated with the previous image (52), for which the "good" minimum is merged with the one corresponding to pushing the car out of the field of view, hence the introduced error.

Table 7 gives the statistics measured from frame 70. Despite this, the error is still much larger than with the other datasets and is drifting significantly more. This happens from around frame 400, at which point the robot arrives under trees, the grass largely disappearing, thus only having brown/grey/dark green colours in its view that also became darker, Figure 59. All this contributes to having less contrast and thus less features to use.

Finally, the CarPark dataset shows good results, Figure 60. However, important variations are visible. These are due to the fact that the environment was visually largely asymmetric (mostly due to inhomogeneous lighting and shadows) with a side much brighter than the other, Figure 61. As mentioned before, this tends to "pull" the heading in one way or the other, depending on the rotation.

3.5 Discussion

We discuss here several aspects shown during the experiments reported above.

The proposed system generally copes well with dynamic environments. In all cases but one, the moving cars and people did not affect the performance of the system. This is partly due to the narrowed field of

Figure 58: GRASS2: the panoramic images corresponding to frame 51 (top) and frames 52 and 53 (middle and bottom) after un-rotation using the global minimum of the distance function on Figure 57. Regions outside the field of view are darker.

view, in effect eliminating most moving objects from the view, but also to the relative small size in the images of these objects, because of the projection on the mirror. However, reducing the field of view constrains the environment even further as this requires features everywhere (the isotropy of the environment pushed to the extreme). Moreover, if large moving objects happen to be in the field of view, then the system introduces an error related to the projection of the motion of the object. The relationship is more or less strong depending on the amount of stable features present in the remain parts of the field of view and on the relative brightness of the moving object. We have indeed shown in Section 3.1.2 that the brighter the colour is the more pull or push it creates. This shows that a better combination of colour space and distance metric must be found. These however are problems inherent to vision; human beings suffer from similar problems.

Another effect of moving objects is that successive images appear very different from each other, forcing the system to use more frames, which is generally not desirable, particularly in such situation. This could possibly be solved by running two instances of the algorithm at the same time with different values of the threshold on the distance amplitude. This would at least allow the detection of something going wrong and possibly the recombination of the two results to obtain a more robust estimation of the heading. Another solution could be to use an idea similar to the one behind the mapping algorithm in [Neal and Labrosse, 2004]: images seldom seen are at first used (in the topological map) and then discarded because not judged "strong" enough (too different from neighbouring images and thus not reinforced by them). When an image gets discarded, the heading could be recomputed from the kept images immediately surrounding it. This needs to be tested and will be when the visual compass is integrated with the mapping algorithm (see below).

The width of the front and back field of view could be adapted according to the images. For example, lack of features can be easily detected with simple methods such as colour variance in the images. Moving objects in the field of view also make the distance as a function of the rotation qualitatively different from the "normal" distance function, fact that could be detected. When such a situation is detected, the field of view could then be increased and/or the space sampling rate modified to tackle these problems.

Despite these problems, we have shown that the algorithm described performs well, providing an estimation of the heading with an error generally well below the maximum error insects [Cartwright and Collett, 1983, Åkesson and Wehner, 2002] and algorithms derived from the snapshot model can cope with (in [Ruchti, 2000] a maximum error of 45° is given). Perhaps more importantly, results show that the error drifts only very slowly[13], which makes it suitable for tasks such as homing or even long-range navigation where the path is specified in terms of visual targets. Moreover, it has been shown that the performance of the system can be improved even further by using higher resolution images. Although this does imply heavier processing, the computations performed in the system being only low level ones, they could be easily implemented in hardware and would thus be faster, and not using processing power of the computer, allowing the use of higher resolution images.

Inertial navigation systems are nowadays often used in robotic applications to incrementally compute the pose of the robot while it moves. Studies on the performance of such systems to estimate the heading reveal that drift of error can be as high as 1.35°/s [Barshan and Durrant-Whyte, 1995], an order of magnitude higher than our system. A better performance can be obtained by inertial sensors when their output is used in an (extended) Kalman filter. In that case, drifts of the order of 3°/min can be obtained but can still produce errors in the heading estimation as high as 12° [Barshan and Durrant-Whyte, 1995] or more recently around 9°

[13] However, the drift probably does not remain low when the robot undergoes a systematically biased trajectory in an asymmetrical environment.

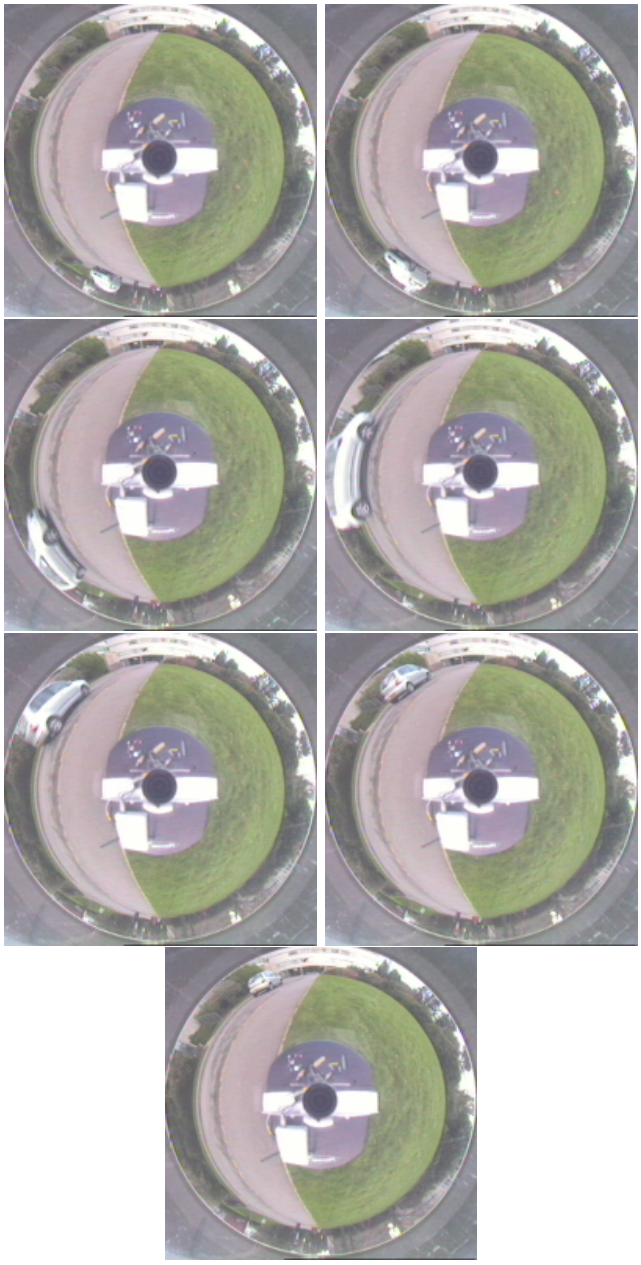


Figure 52: GRASSSTRAIGHT: frames 100 to 106

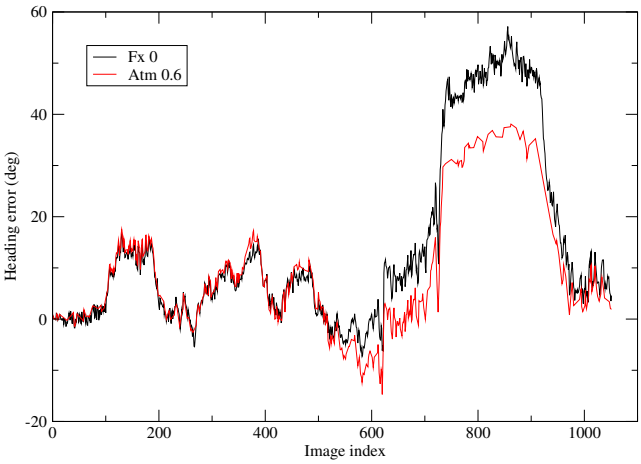


Figure 53: GRASS1: error between the magnetic and visual headings

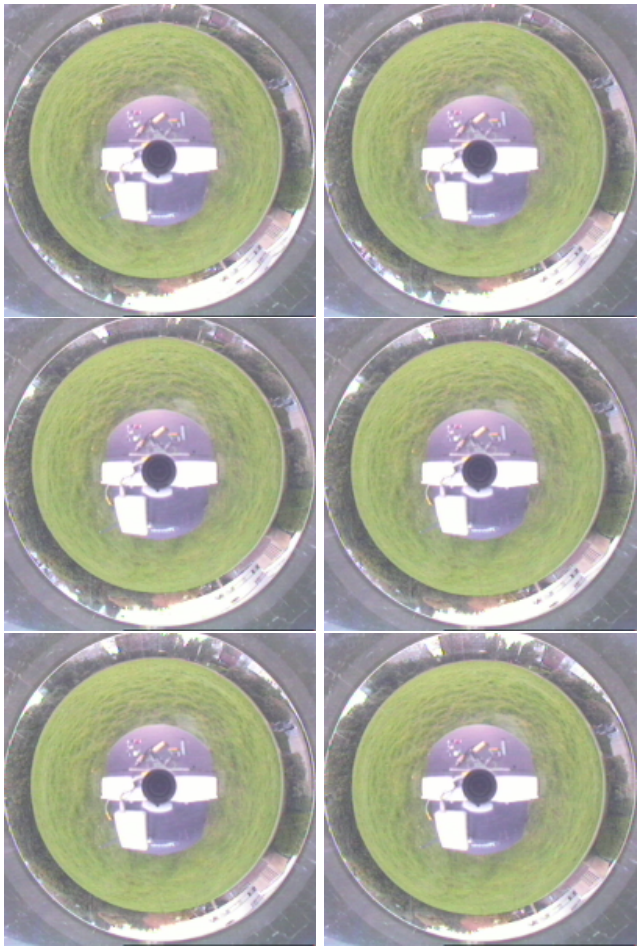


Figure 54: GRASS1: frames 725 to 730

[Hogg *et al.*, 2002]. This shows that our visual compass performs better than inertial sensors and could be improved even further by integrating it with or without other sensors in a Kalman filter (or more generally an extended information filter).

Finally, we have seen that our system provides much smoother headings than the one provided by the magnetic compass used in these experiments. Moreover, our system is not sensitive to yaw and tilt, to some extent. However, if the terrain becomes rough enough images could change dramatically, which would deteriorate significantly the performance of our system. We are currently envisaging the use of passively or actively controlled stabilisation platforms for the camera.

We intend on integrating the present system with a homing algorithm and a mapping algorithm using similar techniques, early versions of which having been presented in the past [Mitchell and Labrosse, 2004, Neal and Labrosse, 2004], in particular to help disambiguate wrong matching. These will allow fre-

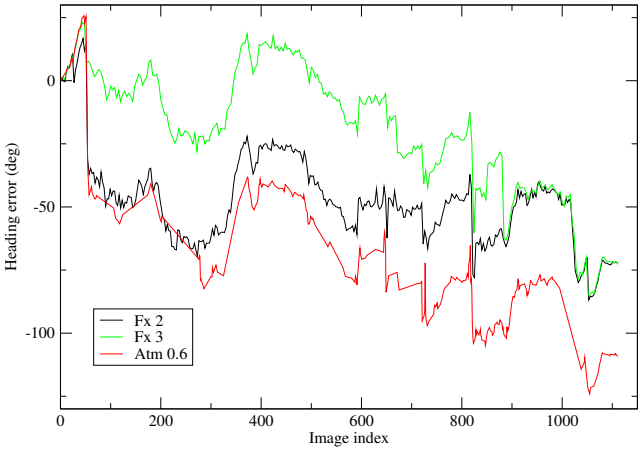


Figure 55: GRASS2: error between the magnetic and visual headings

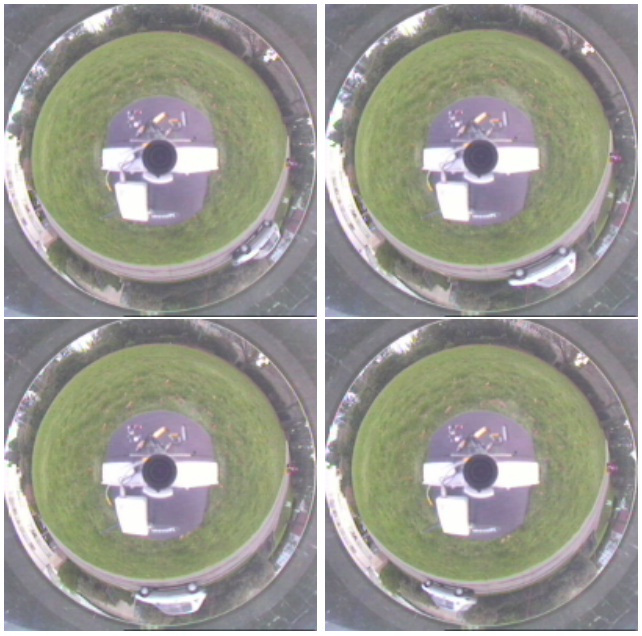


Figure 56: GRASS2: frames 51 to 54

quent re-localisation and thus re-alignment of the heading, similar to procedures insects seem to adopt [Srinivasan *et al.*, 1997].

Finally, it is interesting to note that the measure used to determine when to change the reference image in the proposed algorithm is similar to the one used in a previous publication on mapping; the Network Affinity Threshold (NAT) in [Neal and Labrosse, 2004] is indeed based on the distance between previously stored nodes (representing images) and the current image.

4 Conclusion

We have introduced here a theory to incrementally estimate the heading of a robot using sub-symbolic matching of successive panoramic images grabbed from the robot. We have seen that many factors can contribute

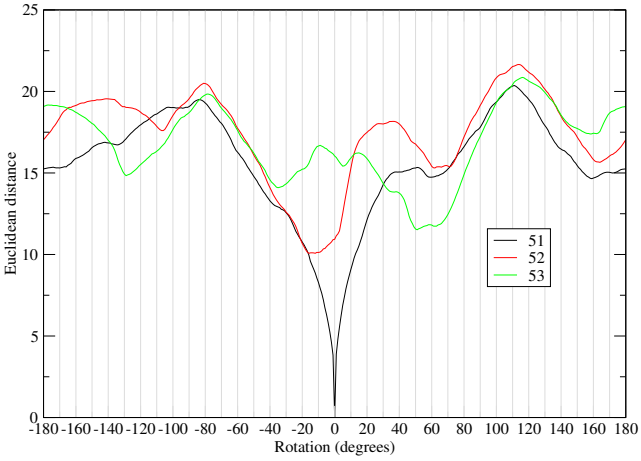


Figure 57: GRASS2: distance between image 51 and images 51 to 53 corresponding to the car passing in the field of view of the robot, Figure 56

Figure 59: GRASS2: frames 400 and 500

positively and/or negatively to the performance of the method and we have performed a careful study with real data of some of these factors. Others remain to be studied, such as the combination of colour space and distance metric.

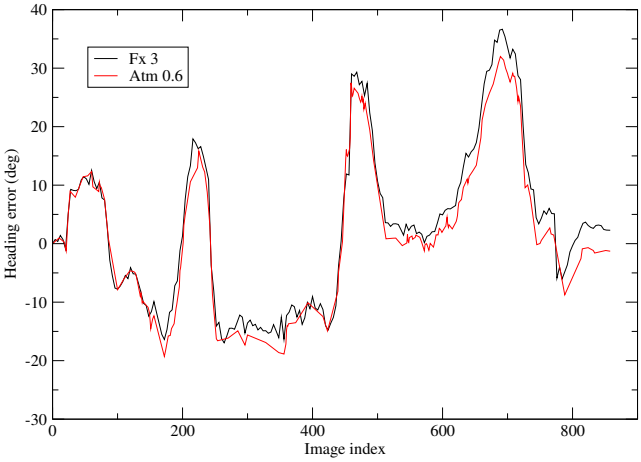


Figure 60: CARPARK: error between the magnetic and visual headings

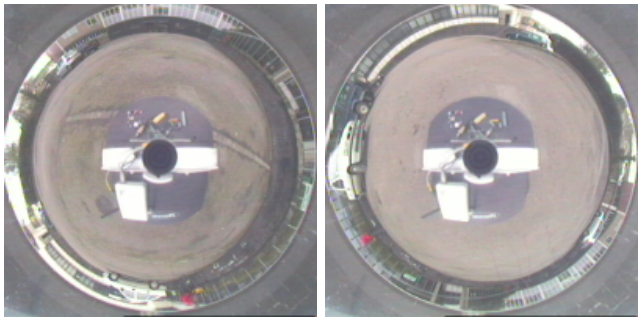


Figure 61: CarPark: frames 26 and 198

An algorithm implementing a solution to the problem as been presented and results using real data acquired in un-modified indoors and outdoors dynamic environments have been presented. Problems inherent more to the visual aspect of the method rather than to the algorithm itself have been discussed. In particular, we have shown that apart from increasing the angular resolution of the panoramic images (only difficult by the extra computation involved), improving the combination of distance metric and colour space is where the method can be improved most. Despite these problems, the performance of the system has been shown to be good and in particular drifting only slowly, which is an important characteristic, especially since the algorithm proceeds by integrating local changes in heading. Finally, the performance is more than sufficient to provide the heading information to methods deriving from the snapshot model and for mapping and navigation methods we are currently working on.

Acknowledgement

The equipment used for this work was partly funded by HEFCW Science Research Investment Fund (SRIF) grants from 1999 (for the VICON system) and 2002 (for the research lab and the robotic equipment).

The author gratefully acknowledges the help from Dr Mark Neal provided during many animated discussions.

References

[Åkesson and Wehner, 2002] Susanne Åkesson and Rüdiger Wehner. Visual navigation in desert ants *Cataglyphis fortis*: are snapshots coupled to a celestial system of reference? *Journal of Experimental Biology*, 205:1971–1978, 2002.

[Barshan and Durrant-Whyte, 1995] Billur Barshan and Hugh F. Durrant-Whyte. Inertial navigation systems for mobile robots. *IEEE Transactions on Robotics and Automation*, 11(3):328–342, 1995.

[Bichsel and Pentland, 1994] Martin Bichsel and Alex P. Pentland. Human face recognition and the face image set's topology. *CVGIP: Image Understanding*, 59(2):254–261, 1994.

[Bisset *et al.*, 2003] David Lindsey Bisset, Michael David Aldred, and Stephen John Wiseman. Light detection apparatus. United States Patent US 6,590,222 B1, 2003. Also UK Patent GB 2 344 884 A, 2000.

[Cartwright and Collett, 1983] B. A. Cartwright and T. S. Collett. Landmark learning in bees: experiments and models. *Journal of Comparative Physiology*, 151:521–543, 1983.

[Cartwright and Collett, 1987] B. A. Cartwright and T. S. Collett. Landmark maps for honeybees. *Biological Cybernetics*, 57(1/2):85–93, 1987.

[Cozman and Krotkov, 1995] Fabio Cozman and Eric Krotkov. Robot localization using a computer vision sextant. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, pages 106–111, 1995.

[Cozman *et al.*, 2000] Fabio Cozman, Eric Krotkov, and Carlos Guestrin. Outdoor visual position estimation for planetary rovers. *Autonomous Robots*, 9(2):135–150, 2000.

[Franz *et al.*, 1997] Matthias O. Franz, Bernhard Schölkopf, and Heinrich H. Bülthoff. Homing by parameterized scene matching. In *Advances in Artificial Life: Proceedings of the European Conference on Artificial Life*, pages 236–245, 1997.

[Franz *et al.*, 1998] Matthias O. Franz, Bernhard Schölkopf, Hanspeter A. Mallot, and Heinrich H. Bülthoff. Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics*, 79:191–202, 1998.

[Frier *et al.*, 1996] Helen J. Frier, Emma Edwards, Claire Smith, Susi Neale, and Thomas S. Collett. Magnetic compass cues and visual pattern learning in honeybees. *Journal of Experimental Biology*, 199(6):1353–1361, 1996.

[Gaspar *et al.*, 2000] José Gaspar, Niall Winters, and José Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transactions on Robotics and Automation*, 16(6):890–898, 2000.

[Goedemé *et al.*, 2005] Toon Goedemé, Tinne Tuytelaars, Luc Van Gool, Dirk Vanhooydonck, Eric Demeester, and Marnix Nuttin. Is structure needed for omnidirectional visual homing? In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 303–308, 2005.

[Gonzales-Barbosa and Lacroix, 2002] José-Joel Gonzales-Barbosa and Simon Lacroix. Rover localization in natural environments by indexing panoramic images. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1365–1370, Washington, USA, 2002.

[Gourichon, 2004] Stéphane Gourichon. *Utilisation d'un compas visuel pour la navigation d'un robot mobile*. PhD thesis, Université Paris VI, Paris, France, 2004.

[Graham *et al.*, 2003] Paul Graham, Karine Fauria, and Thomas S. Collett. The influence of beacon-aiming on the routes of wood ants. *Journal of Experimental Biology*, 206:535–541, 2003.

[Graham *et al.*, 2004] Paul Graham, Virginie Durier, and Thomas S. Collett. The binding and recall of snapshot memories in wood ants (*Formica rufa* L.). *Journal of Experimental Biology*, 207:393–398, 2004.

[Hogg *et al.*, 2002] Robert W. Hogg, Arturo L. Rankin, Stergios I. Roumeliotis, Michael C. McHenry, Daniel M. Helmick, Charles F. Bergh, and Larry Matthies. Algorithms and sensors for small robot path following. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3850–3857, Washington DC, USA, 2002.

[Jogan and Leonardis, 2000] Matjaž Jogan and Aleš Leonardis. Robust localization using panoramic view-based recognition. In *Proceedings of the International Conference on Pattern Recognition*, volume 4, pages 136–139, Barcelona, Spain, 2000.

[Judd and Collett, 1998] S. P. D. Judd and T. S. Collett. Multiple stored views and landmark guidance in ants. *Nature*, 392:710–714, 1998.

[Labrosse, 2004] Frédéric Labrosse. Visual compass. In *Proceedings of Towards Autonomous Robotic Systems*, pages 85–92, University of Essex, Colchester, UK, 2004.

[Lu *et al.*, 1998] Haw-minn Lu, Yeshaiahu Fainman, and Robert Hecht-Nielsen. Image manifolds. In *Proceedings of SPIE; Applications of Artificial Neural Networks in Image Processing III*, volume 3307, pages 52–63, San Jose, CA, USA, 1998.

[Marr, 1982] David Marr. *Vision*. W. H. Freeman and Company, 1982.

[Mitchell and Labrosse, 2004] Tristan Mitchell and Frédéric Labrosse. Visual homing: a purely appearance-based approach. In *Proceedings of Towards Autonomous Robotic Systems*, pages 101–108, University of Essex, Colchester, UK, 2004.

[Möller *et al.*, 1999] Ralph Möller, Marinus Maris, and Dimitrios Lambrinos. A neural model of landmark navigation in insects. *Neurocomputing*, 26–27:801–808, 1999.

[Möller, 2001] Ralf Möller. Do insects use templates or parameters for landmark navigation? *Journal of Theoretical Biology*, 210(1):33–45, 2001.

[Naval Jr. *et al.*, 1997] Prospero C. Naval Jr., Masayuki Mukunoki, Michihiko Minoh, and Katsuo Ikeda. Estimating camera position and orientation from geographical map and mountain image. In *Proceedings of the 38th Research Meeting of the Pattern Sensing Group, Society of Instrument and Control Engineers*, pages 9–16, Tokyo, Japan, 1997.

[Nayar *et al.*, 1996] Shree K. Nayar, Sameer A. Nene, and Hiroshi Murase. Subspace methods for robot vision. *IEEE Transactions on Robotics and Automation*, 12(5):750–758, 1996.

[Neal and Labrosse, 2004] Mark Neal and Frédéric Labrosse. Rotation-invariant appearance based maps for robot navigation using an artificial immune network algorithm. In *Proceedings of the Congress on Evolutionary Computation*, volume 1, pages 863–870, Portland, Oregon, USA, 2004.

[Nelson and Aloimonos, 1988] Randal C. Nelson and John Aloimonos. Finding motion parameters from spherical flow fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273, 1988.

[Röfer, 1995] Thomas Röfer. Image based homing using a self-organizing feature map. In *Proceeding of the International Conference on Artificial Neural Networks*, volume 1, pages 475–480, 1995.

[Röfer, 1997] Thomas Röfer. Controlling a wheelchair with image-based homing. In *Proceedings of the AISB Symposium on Spatial Reasoning in Mobile Robots and Animals*, Manchester University, UK, 1997.

[Ruchti, 2000] Sepp Ruchti. Landmark and compass reference in landmark navigation. Master's thesis, Artificial Intelligence Lab des Institutes für Informatik der Universität Zürich, 2000.

[Shaw and Barnes, 2003] Andy Shaw and Dave Barnes. Landmark recognition for localisation and navigation of aerial vehicles. In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 1, pages 42–47, 2003.

[Srinivasan *et al.*, 1997] Mandyam V. Srinivasan, Shao Wu Zhang, and N. J. Bidwell. Visually mediated odometry in honeybees. *Journal of Experimental Biology*, 200:2513–2522, 1997.

[Tenenbaum, 1998] Joshua B. Tenenbaum. Mapping a manifold of perceptual observations. *Advances in Neural Information Processing Systems*, (10), 1998.

[Thompson *et al.*, 1993] William B. Thompson, Thomas C. Henderson, Thomas L. Colvin, Lisa B. Dick, and Carolyn M. Valiquette. Vision-based localization. In *Proceedings of DARPA Image Understanding Workshop*, pages 491–498, 1993.

[Vardy and Oppacher, 2003] Andrew Vardy and Franz Oppacher. Low-level visual homing. In *Advances in Artificial Life: Proceedings of the European Conference on Artificial Life*, 2003.

[Vassallo *et al.*, 2002] Raquel Frizera Vassallo, José Santos-Victor, and Hans Jörg Schneebeli. Using motor representations for topological mapping and navigation. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 478–483, Lausanne, Switzerland, 2002.

[Wehner *et al.*, 1996] Rüdiger Wehner, Barbara Michel, and Per Antonsen. Visual navigation in insects: coupling egocentric and geocentric information. *Journal of Experimental Biology*, 199:129–140, 1996.

[Woodland and Labrosse, 2005] Alan Woodland and Frédéric Labrosse. On the separation of luminance from colour in images. In *Proceedings of the International Conference on Vision, Video, and Graphics*, pages 29–36, University of Edinburgh, UK, 2005.

[Zeil *et al.*, 1996] Jochen Zeil, Almut Kelber, and Rüdinger Voss. Structure and function of learning flights in bees and wasps. *Journal of Experimental Biology*, 199:245–252, 1996.

[Zeil *et al.*, 2003] Jochen Zeil, Martin I. Hofmann, and Javaan S. Chahl. Catchment areas of panoramic snapshots in outdoor scenes. *Journal of the Optical Society of America A*, 20(3):450–469, 2003.