Estimating ego-motion using a panoramic sensor: Comparison between a bio-inspired and a camera-calibrated method

S. Gourichon⋆, J.A. Meyer⋆

⋆AnimatLab – LIP6
8, rue du Capitaine Scott
75015 Paris, France
stephane.gourichon@lip6.fr, jean-arcady.meyer@lip6.fr

S. H. Ieng†, L.Smadja†, R. Benosman†

†Laboratoire des Instruments et Systèmes
Université Pierre et Marie Curie
4 place Jussieu
75252 Paris cedex 05
rbo@lis.jussieu.fr

Abstract

We introduce and compare two algorithms related to ego-motion, applicable to a robot using a panoramic visual sensor in an unknown environment. The first method, computationally cheap, extends a family of bio-inspired navigation systems by providing an orientation estimation that previously had to be obtained from external references. The second method solves the general structure-from-motion problem and is here adapted to the panoramic sensor case. Both methods produce good results in a test scene, though the calibrated method has a greater precision. A step-by-step comparison of the methods is included, with mobile robotics applications in mind.

1 Introduction

In Gourichon et al. (2002), a biomimetic homing navigation method was applied to a robot, based on the work of Cartwright and Collett (1983, 1987) on searching behaviour in bees. Because this method assumed that the orientations of two snapshots – i.e., two one-dimensional projections of thresholded luminances on a circular retina respectively taken in a given goal place and in any other place close to the former – were known to the robot, only constrained moves were allowed, committing the robot to always face the same direction. To alleviate such constraint, a visual compass has been designed, thus making it possible to monitor the robot's changes of direction along any path, thanks to simple measures of parallax and to simple extensions of Cartwright and Collett's logic of snapshot comparisons.

Turns out besides that monitoring a robot's orientation is part of the more general issue of estimating its ego-motion (Chang and Hebert, 2000; Faugeras and Maybank, 1990; Tomasi and Kanade, 1992; Spetsakis and Aloimonos, 1990), and that sophisticated 3D-vision algorithms may be used to monitor both the translations and rotations of the robot. Incidentally, the same algorithms may also serve for the 3D reconstructions of unknown environments (Ikeuchi et al., 2001; Narayanan and Kanade, 1998; Faugeras et al., 1998).

To assess the respective merits of these two independent lines of approach, an experiment has been set up to demonstrate their effectiveness at determining the orientation of a robot equipped with a catadioptric sensor. Two methods have been accordingly implemented. The first one calls upon simple bio-inspired parallax comparisons, while the second is more mathematically inclined and extensively relies on non Euclidian geometries and minimizations techniques.

This paper describes the basic principles underlying these methods, and the results that were obtained

in simple experimental conditions. The chances that these methods will prove to be efficient in more challenging environments are also discussed, and indications for future improvements are provided.

2 A visual compass from a biologically inspired context

The first method in this comparison is a simple and computationally cheap visual compass designed for an animat able to perceive the azimuths of some landmarks relative to its body. After an observation stage, done once and for all for a given location, there is an exploitation stage when, from any location in the vicinity, the visual compass can estimate the absolute orientation of the animat's body using only perceived relative angle measurements.

Only azimuths are used, no elevation, which makes sensor calibration unnecessary. A simple segmentation algorithm and a matching procedure by dynamic programming are used.

The essence of the algorithm is to take advantage of the local linearity of the variation of any azimuth with respect to the animat motion. This allows, at the observation stage, to capture relevant information without knowing the complete motion of the animat. At the exploitation stage, this also allows to very simply retrieve the animat's orientation from any later view taken in the vicinity of the observation place.

2.1 Early image processing

It is necessary to extract azimuths from the sensor's image. The image processing used here is voluntarily simple. It starts from a color panoramic image (see figure 7 right). The first step is to unroll a ring portion of the color input image according to rectangular-polar transformation to obtain a rectangular colored strip indexed by azimuth. Then a simple segmentation procedure is used, that selects portions of the panorama with highly saturated (colorful) pixels, and separates regions of nearly constant hue at azimuths where the average hue variation is beyond a threshold. The result is a set of regions sorted by azimuths, with possible holes caused by portions of the initial panorama without bright colors. A real-world panorama typically results in less than 100 regions. In this experiment, the regions correspond to the colored cardboard.

The segmentation algorithm, though very rough, is the same we already used successfully in real-world robotic short range guidance experiments in Gourichon et al. (2002).

2.2 Matching algorithm

The visual compass requires the ability to establish correspondences between objects appearing in different views. Here we called upon a dynamic programming algorithm that we already used for short range guidance. It is explained in more details in Gourichon et al. (2002). It globally maximizes a matching score from a local score function that takes color information (particularly hue) into account. This step remains reasonably cheap ($O(n^2)$ with $n < 100$) because the segmentation algorithm provided a small set of regions.

2.3 Using the parallax in two stages

As summarized above, the biologically-inspired visual compass operates in two stages: an observation stage, and an exploitation stage.

2.3.1 Notations

Let us introduce the notations illustrated in figure 1. Given a view taken at place X, we name $\theta_{a \to b}^{X}$ the signed angle between the azimuth of feature a and the azimuth of feature b, where a feature may be either a landmark number such as i, or a reference direction such as $north$ or $body$.

We call parallax of a landmark during an animat's motion the variation of azimuth caused by this motion. Thus, for landmark i, for a motion from M to N, the parallax is $D_i^{M \to N} = \mu(\theta_{north \to i}^{N} - \theta_{north \to i}^{M})$ where $\mu(x)$ is the number closest to zero in the set (x modulo 2π).

2.3.2 Observation stage

During the observation stage the animat observes the parallax caused by ego-motion. The animat has to know its orientation during this stage, but the translation can be unknown. It takes three views (named M,N,P) that are not aligned and have in common at least three landmarks (named a,b,c). The ideal situation is one distant landmarks and two close landmarks, as we will illustrate later. The worst situation is when the observation region and the three landmarks lie on a circle.
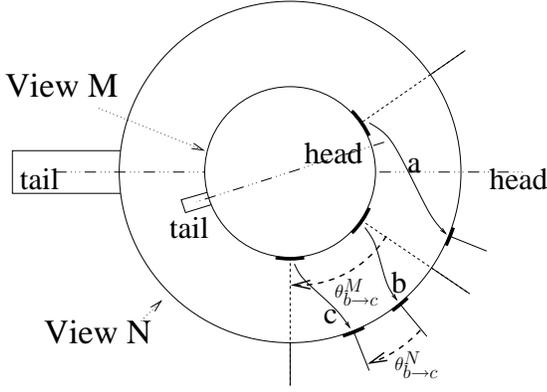
Figure 1: Illustration for notation $\theta_{a\rightarrow b}^N$. Curved arrows represent three landmarks a,b,c paired between views M and N by the matching process.

From the parallax measured we will calculate what we'll call a "bearing". It is a set of three real numbers $\lambda_a, \lambda_b, \lambda_c$ that are attached to the triplet and will be used to estimate a direction.

For $(i, j, k) \in \{(a, b, c)(b, c, a)(c, a, b)\}$, let

$$\lambda_i = D_j^{M\rightarrow N} D_k^{M\rightarrow P} - D_j^{M\rightarrow P} D_k^{M\rightarrow N}$$

If the sum of the three λ's is close to zero, the triplet is unsuitable and simply ignored. The reason for this formula and details are given in Gourichon (2003). Usually, more than 3 landmarks are available in a standard situation. Then, all suitable triplets are enumerated and used to build corresponding bearings.

2.3.3 Exploitation stage

Each time a new view X of unknown orientation is taken, that has common landmarks with the views used at the observation stage, an exploitation stage can be conducted. It consists in estimating the animat's orientation, using a previously memorized bearing.

If the orientation of the new view is completely unknown, then it is first artificially rotated so that the landmark that has the highest lambda in the bearing appears to have the same azimuth as in view M. This prevents angular wrapping artifacts of pseudo-parallaxes in equation 1 below.

The denominator attached to the bearing is calculated:

$$d = \lambda_a + \lambda_b + \lambda_c$$

as well as a linear combination that depends on X and that we call the invariant:

$$I(X) = \sum_{i\in\{a,b,c\}} \lambda_i \underbrace{\mu(\theta_{body\rightarrow i}^X - \theta_{body\rightarrow i}^M)}_{\text{pseudo-parallax}} \quad (1)$$

Then the current orientation of the animat's body $\theta_{north\rightarrow body}^X$, that can be estimated from this bearing is simply:

$$\theta_{north\rightarrow body}^X = \theta_{north\rightarrow body}^M - \frac{I(X)}{d} \quad (2)$$

Let us summarize how it works. The pseudo-parallax is a real parallax if views M and X have the same orientation reference, i.e. $\theta_{north\rightarrow body}^X = \theta_{north\rightarrow body}^M$. The three λ's are calculated so that the linear combination of parallaxes is constant, in first order approximation, in the vicinity of the observation places. As a consequence, if the view X is correctly oriented, $I(X) \simeq 0$. This is exact by construction, for $X = M$, $X = N$ and $X = P$.

At the observation step the linear combination $I(X)$ involves pseudo-parallaxes, because it is calculated as if the view X was correctly oriented, while it is not necessarily the case. This generates the same offset at the level of each parallax, which contributes to the linear combination. Since the linear combination is chosen to yield a constant result with real parallaxes, the difference between actual and expected values is the error in orientation (the only unknown value) multiplied by the sum of the linear coefficients d (which is known). Hence equation 2.

To make things better, more that a single invariant may be available, because more than one bearing may have been recorded at the observation step, from as many triplets of landmarks. To contribute to the overall direction estimate, each bearing yields a unit vector pointing to the estimated body direction. These unit vectors are summed up and their resultant provides the aggregated direction estimate. The length of the resultant vector is used as a self-confidence estimator: it is maximal when all bearings can be used and all indicate the same direction, while it is reduced if the different estimated directions contradict each other. Additional details are given in Gourichon (2003).

2.3.4 Example

For clarity of the example and figure, the equations in this example are simplified and use absolute az-

imuths θ instead of parallaxes, and would not always work properly *as is* because of angular wrapping effects.

Figure 2 shows an example animat motion and an observed invariant. Since $\theta_{north\rightarrow i} = \theta_{north\rightarrow body} + \theta_{body\rightarrow i}$, this linear equation it is very easy to solve for $\theta_{north\rightarrow body}$.

If the animat moves or rotates, or a combination of both, and takes a new view X, $\theta^X_{north\rightarrow body}$ is as simple to retrieve as:

$$-\frac{43.25 \cdot \theta^X_{body\rightarrow a} + 92 \cdot \theta^X_{body\rightarrow b} + 55.25 \cdot \theta^X_{body\rightarrow c}}{43.25 + 92 + 55.25}$$

It is interesting to notice that the λs are calculated based on a particular motion, which is unknown at any time. Only a constant orientation during this motion is assumed. In a sense, this method separates orientation from motion, or compensates motion to retrieve orientation.

One of the prerequisite for this method is that this separation is possible. For any particular triplet, there is a region where it is not true (figure 3).
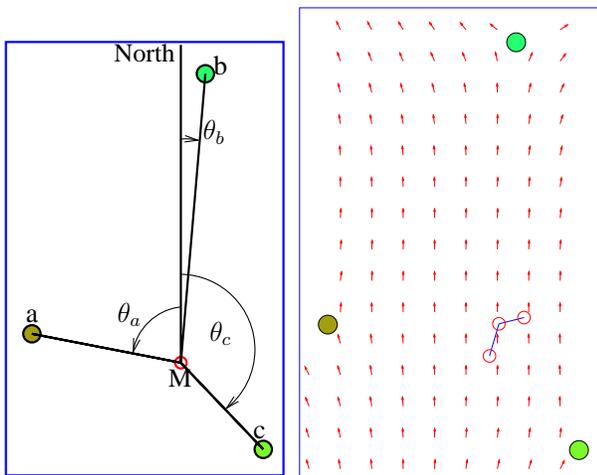

Figure 2: Example illustrating the principle of the bio-inspired visual compass. Shaded circles are landmarks. Moving along the three connected open circles (right) the animat observes that $43.25 \cdot \theta_{north\rightarrow a} + 92 \cdot \theta_{north\rightarrow b} + 55.25 \cdot \theta_{north\rightarrow c}$ is approximately constant. Since this is a linear equation it is very easy to solve for $\theta_{north\rightarrow body}$. Actually, this equation would suffer from angular wrapping effects, which is why parallax are used instead of azimuths, and the view is first pre-oriented in favor of the most stable landmark (the one with the highest lambda, see text). The vector field on the right shows the estimated north direction at points of a grid.
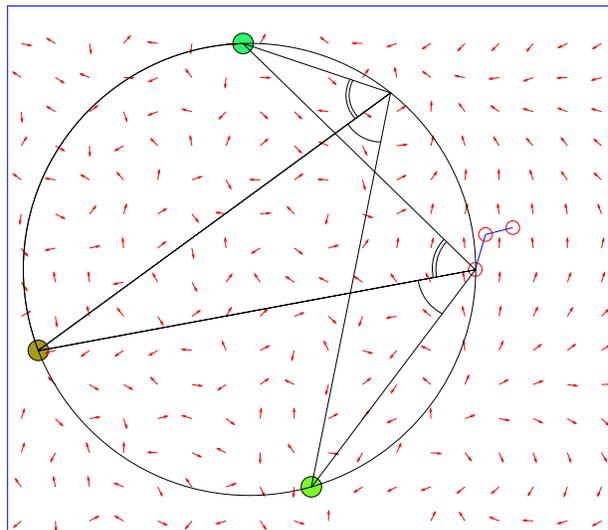

Figure 3: Example illustrating a degenerate case and a limitation of the bio-inspired method. The circle traversing any three landmarks has the geometrical property that the perceived relative angles are the same from any position (represented here at two positions). From the animat's point of view it means that orientation (turning around itself) cannot be separated from motion (moving on the circle). As a result, if the observation region is too close to this circle, the corresponding triplet will produce unreliable results. In this example the observed paramaters are $\lambda_a = -31$, $\lambda_b = 14$, $\lambda_c = 16$ and $d = -1$. d is much smaller than any lambda, which is a measurable sign of a bad triplet in this region. In a real situation, the compass would have ignored it in favor of other triplets.

3 A camera calibration-based visual compass

This method estimates both rotational and translational components of motion from two views only. In a first stage, it calibrates the internal parameters of the camera, thus making it possible to get elevations from the camera image. In a second stage, a matching procedure based on robust methods is applied. It is rotation and scale invariant to cope with the non linear resolution of the sensor. Finally a structure-from-motion direct linear algorithm retrieves the motion parameters using a robust method, prodiving complete ego-motion (i.e. translation and rotation).

3.1 The necessity of calibration

The use of catadioptric sensors to retrieve the complete motion between two frames implies necessarily entails knowing the 3D orientation of the line of sight associated with each image pixel. In other words, the 3D direction in space which corresponds to the direction in which it is observing the scene must be known (J. Fabrizzio, 2002; Geyer and Daniilidis, 2002, 2001). A catadioptric sensor being an association of a camera and a mirror, it is possible to elaborate a model that associates to each view point (camera, mirror) a 3D coordinates system. It is then important to retrieve the unknowns of the model, which are the camera internal parameters, the distance between the camera and the mirror and the orientation of the mirror. The mirror being generally manufactured with a very high precision, its parameters are assumed to be known. Moreover, by construction, most panoramic catadioptric sensors fulfill what is called the *single viewpoint constraint*, which means that all incident rays reflected by the mirror and intersecting in the camera focal point also intersect in a single point called the mirror viewpoint by Baker and Nayar (1998). Retrieving the unknown parameters makes it possible to check the validity of the hypothesis of the single view point constraint, and allows a better computation of the 3D vector of sight associated with each pixel (see figure 4). This is an advantage as the process of determining the motion from two consecutive images is very sensitive to noise and can provide distorted results if the calibration does not provide accurate parameters (see next section).

3.2 Matching catadioptric images

To be able to recognize common real-world features appearing in two images, it is necessary to perform a match.

Most of the existing matching methods rely on the same principle. They first extract features from two "rough" images, generally high curvature points. Then, they select around each point a neighborhood of points and compute a correlation between all the neighborhoods. The best score implies the best fit (Faugeras, 1993).
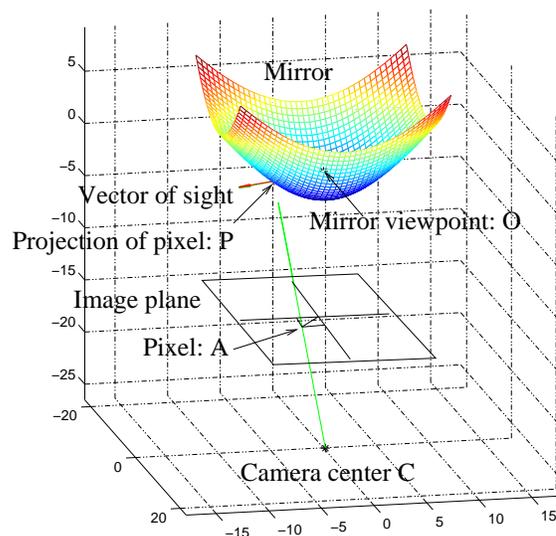

Figure 4: The aim of the calibration if to find the precise relationship between pixel coordinates in the image and the 3D direction of the vector of sight. This representation relies on the estimated model provided by calibration.

3.2.1 Feature extraction

The feature extraction normally used is a corner extraction, performed using a classical Harris detector (Harris and Stevens, 1988), directly applied to images, i.e. without paying attention to their catadioptric origin.

For this experiment, black spots were used instead (see figure 7). The feature points were the center of the spots.

3.2.2 Window definition

Matching interest points in stereoscopic views requires a sharp neighborhood extraction. Classical approaches use predefined size patches, often squares, centered on interest point. Unfortunately, in the present case, similar approaches would not give reliable results because the shape of the mirror does not guarantee an homogeneous resolution (it is maximum at the center and decreases when getting near the borders). This is why a dedicated procedure, derived from Svoboda and Pajdla (2001),which takes into account the spatial resolution variation by means of dynamically resized patches is used instead.

The shape of the neighborhood windows are not squares but diamonds defined by their vertices (see figure 5). Thanks to calibration, there is a two-way correspondence between each pixel in the image and
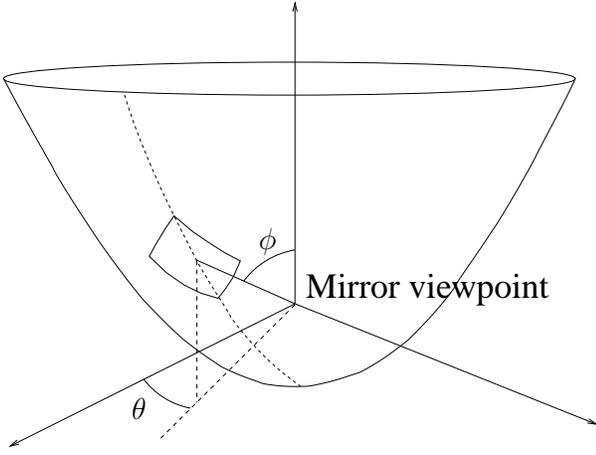
Figure 5: Window definition. The diamond is defined from the mirror surface, its vertices are obtained by setting azimuth (θ) and elevation (ϕ) ranges.

Figure 6: Motion estimation step. The sensor was moved from referential R_1 to referential R_2 between image 1 and image 2. If A_1 and A_2 are two correctly matched points there is a relation between O_1, O_2, P_1, and P_2. The coherence of this relation among many points allows at the same time to estimate the motion and reject mismatched points.

its projection on the mirror. Each feature point is projected on the mirror. The quadrilateral patch definition is an original method that defines the associated window on the mirror surface by fixing angular ranges around the point's azimuth θ and elevation ϕ (see figure 5). Then, the vertices of the diamond are projected again from the mirror to pixels in the image plane. By setting the neighborhood window on the mirror instead of the image, we are able to cope with the spatial resolution requirement. Dimensions of the window increase with the distance to the center, under the assumption of fixed angular parameters.

3.2.3 Matching and outliers removal

Optimal matching requires an appropriate sampling of the extracted neighborhoods which may have different size. Each window is accordingly resized using a classical bilinear interpolation.
Since the windows are projected on the mirror at a known azimuth, disparity angle for each couple of points can easily be computed. Thus, patches are correctly oriented before the matching process.

The correlation score computation should ensure one, and only one, best match for each point; therefore we use a centered and normalized cross correlation. Points are paired if they mutually give the best similarity scores.
This is still not enough to get reliable points. In order to remove outliers, a threshold extracted from
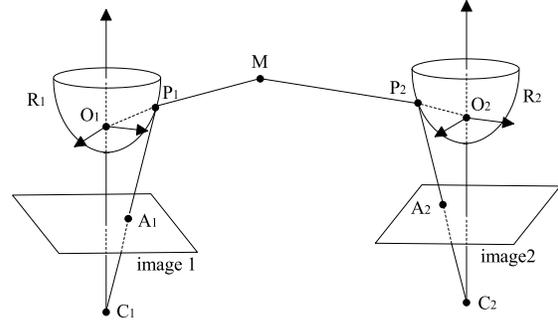
matched points is applied on the measured angular disparity set. Pair of matched points are removed if their relative angles are too far from the computed average value ($|\theta - \overline{\theta}| > 2.5\sigma$). Additional details are given in Ieng et al. (2003).

3.3 Extrinsic parameters estimation

The next step is to actually find the motion between our two matched images. This step is based on previous work from Faugeras and Maybank (1990); Chang and Hebert (2000) and Svoboda et al. (1998).

Let us consider a couple of matched points, A_1 in image 1 and another A_2 in image 2. If all the previous steps worked correctly, the two matched points correspond to the same location M in the surrounding scene (figure 6). If it is true, then O_1, O_2, P_1, and P_2 lie on the same plane:

$$\overrightarrow{\mathbf{O_1P_1}}_{|R1}.(\overrightarrow{\mathbf{O_1O_2}}_{|R1} \wedge \overrightarrow{\mathbf{O_2P_2}}_{|R1}) = 0 \quad (3)$$

Changing the form of the equation as followed leads to a well known property in stereovision, which is interesting because P_1 and P_2 are actually known:

$$\mathbf{P_1^t}.E.\mathbf{P_2} = 0 \text{ assuming that } \begin{cases} \mathbf{P_1} = \overrightarrow{\mathbf{O_1P_1}}_{|R1} \\ E = \overrightarrow{\mathbf{O_1O_2}}_{|R1} \wedge R_2^1 \\ \mathbf{P_2} = \overrightarrow{\mathbf{O_2P_2}}_{|R2} \end{cases}$$

The only unknown is E, the essential matrix defined by Longuet-Higgins (1981). This matrix repre-

sents the coordinate transformation between referentials $R1$ and $R2$, that is the translation and rotation that was performed between the views we are working on, i.e. rotation R_2^1 and translation T_2^1 components.

Equation (3) is simplified to a more explicit form: $\mathbf{U}.\mathbf{e} = 0$, where $\mathbf{U} = (p_{1x}p_{2x},\ p_{1x}p_{2y},\ p_{1x}p_{2z},\ p_{1y}p_{2x},\ p_{1y}p_{2y},\ p_{1y}p_{2z},\ p_{1z}p_{2x},\ p_{1z}p_{2y},\ p_{1z},\ p_{2z})$ and e is a column vector containing all components of E.

This is a dot product with eight terms. In concrete terms, if all matching was perfect and measurements without noise, seven of those equations would be enough to know the motion. Thus, this method needs at the bare minimum, 7 matched points in the environment.

Actually, there are mismatched points and their position is not exactly known. So the problem in transformed into minimizing $\|\mathbf{U}.\mathbf{e}\|$. As estimating E means minimizing errors, we solved our optimization problem by weighting contribution of each interest point.

The reader can refer to the work of Zhang (1998) for a complete review of the robust existing techniques.

4 Experimental setup

In order to be able to compare both methods in the real world, we used the same testbed. Although both methods are meant to work on unprepared environment, we validated them here using a constrained, prepared one. This makes it easier to cope with experimental problems and have reproducible results.

The experimental setup is shown in figure 7. Around a flat and rectangular surface were displayed several colored patterns. A series of 2D positions forming a grid was determined. At each 2D position, we grabbed an image, forming a database of several view points. Each method was applied on the images of the database, with the goal of determining the orientation between selected frames.

5 Experimental results

5.1 Biologically-inspired method

After all the views have been acquired, three views were selected in the center of the arena for the observation step, which built 120 bearings.
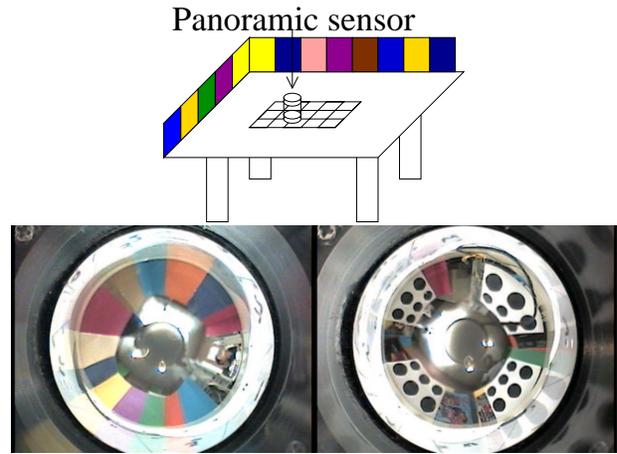

Figure 7: Experimental setup. For the bio-inspired visual compass, colored cardboard was attached around a table to provide environmental features that were easy to segment. For the calibrated method, black spots were displayed to ensure that points were available at different elevations.

Then, for each available view (including the ones used for observation) we applied the bio-inspired algorithm to estimate the orientation of the sensor when the views were taken. Results are shown on the left of figure 8. In this example, the mean error is 1.39 degrees and the standard deviation 3.54 degrees.

5.2 Sensor calibration-based method

After all the views have been acquired, one view was selected as reference (shown in red in figure 8).

Then, for each availavle view we applied the calibrated method, estimating its orientation with respect to the reference view.

In this example, the mean error is 0.155 degrees, and standard deviation is 0.156 degrees.

6 Discussion

In this run, the results of the camera-calibrated approach are considerably more precise (standard deviation 0,16 degrees versus 3.54) in this specific case. Naturally, additional experiments would be necessary to draw more generic conclusions. Furthermore, none of the two methods was tested in the case of an unprepared environment in which they are intended to be used in the future.

A number of differences are summarized in table 1.

	Bio-inspired visual compass	Calibrated method
translation	not calculated here	calculated
rotation	one angle	three angles
features located by	azimuths	azimuths and elevations
determination of interest points	simple 1D segmentation	center of displayed spots or corners
minimum features needed	3 regions (landmarks)	7 points
matching criterion	color (mainly hue)	2D neighborhood
matching procedure	1D dynamic programming	multi-resolution resampling and correlation
aggregation	simple vector sum	sophisticated minimization techniques

Table 1: Feature comparison between the two methods. Most differences illustrate the differences in the contexts in which those methods were developed.
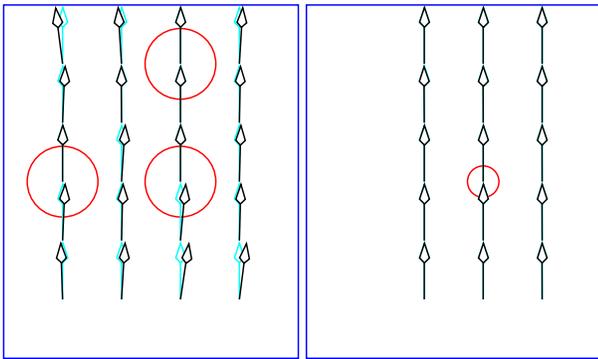

Figure 8: Experimental result. Left: biologically inspired method. The circles represent the three locations that were used for the observation stage. Each black arrow shows the orientation computed at its origin. The shaded arrows show the true reference direction. Right: Sensor calibration-based method. The error is too small to be noticed visually. Only three rows of images were taken.

The simplification affordable by assuming planar motion alone was experimented by Faugeras et al. (1998) in the case of classical projective cameras. In our case with a catadioptric sensor, further simplification can be done because the axis of rotation is the same as the camera-mirror axis direction. The choice of the bio-inspired visual compass to use only azimuths makes it independent of mirror shape and camera parameters, which makes calibration unnecessary.

One might wonder why the original bio-inspired ego-motion method focuses only on rotation. The reason is that the biomimetic navigation models needs correctly oriented snapshots (hence this work) to produce a direction to go to reach the goal (where the reference snapshot was taken). Franz et al. (1998b) proved that the direction to go can be considered as an approximate ego-motion vector, which error in direction is less than $90°$. This is of course insufficient for many applications of ego-motion, but actually sufficient for robot navigation.

The visual compass needs at the bare minimum 3 identifiable features, all visible from the 3 observation views. By contrast, the calibrated method needs at the bare minimum 7 points from the reference view. In both methods, the more visible features, the better the precision of the estimation. This difference is directly related to the number of parameters estimated.

The early image processing presented here for the bio-inspired visual compass are a bit too simplistic, particularly the segmentation algorithm which is too dependent on colored objects. It proved sufficient for real robot navigation, but according to recent experiments it is the first thing that needs improvement to do real robot orientation. The calibrated method normally uses more robust, though heavier, corner extraction techniques (although black spots were used for this experiment).

There is also room for improvement in the aggregation step of the bio-inspired method. We will test a modified aggregation algorithm, that weights bearing contributions according to their denominator. The calibrated method uses robust minimization algorithms which, again, are costly. An adaptation of those iterative techniques to robotics would probably reduce its cost by spreading calculation in time:

making a quick rough estimate and using as an input on next frame acquisition, which will at each new frame refine the position found.

Despite all the contrast between techniques, in complexity and computational cost, both methods produce comparable results in this experiment. The error of the calibrated method is one order of magnitude smaller than the bio-inspired method, which is an expected result, showing the benefit of the the more costly methods. Not all robotics applications need such a great precision, though.

Of course all methods fail when their prerequisites are not satisfied. Since the bio-inspired method relies on local linearity of the variation of azimuths with respect to the animat motion, it fails when this assumption becomes too approximative. This happens in particular if the animat approaches one of the landmarks too closely: the associated azimuth varies more than during the observation because of shorter distance, causing I constancy to fail. In a real robotic application, an enhanced aggregation technique would ignore the too close landmark because it only results in outliers during the aggregation, but this would succeed only if enough other bearings are available. Another possibility, instead or in combination with the previous one, would be to call upon the complementarity of the visual compass property (observe once, use many times) with other source of information, for example odometry (used locally only to avoid drifts) to make other observations and extend the range of the visual compass as much as needed, including the information in a topological map, like Franz et al. (1998a) did with a different orientation strategy.

While we already used the same segmentation technique in Gourichon et al. (2002), we noticed that the bio-inspired visual compass is sensitive to wrong segmentations and wrong matchings, while the insect-inspired navigation strategies are usually very robust and forgiving in this matter. We think that there is an interesting parallel with the results of Svoboda and Sturm (1996) showing that the noise in the calibration parameters affects rotation estimation more than translation estimation.

Due to its simplifying assumption of a sole rotation around a vertical axis, the bio-inspired method makes itself vulnerable to situations where the motion is not even. It will fail if a rotation around a different axis occurs, much the same way many navigation models fail when they wrongly assume that they know their correct orientation, because the linear relation that is called upon will be disrupted and irrelevant. By contrast, the calibrated method should benefit from the more thorough work it does, by being able to deal with any rotation. In both contexts, gravity may help if properly used (inertial navigation systems and/or low-cost mechanical arrangement maintaining the robot "head" upwards).

Synthetically, all this agrees with the common wisdom that robotics can take advantage of computationally much lighter methods than the ones used for 3D environment reconstruction. If an animat can deduce with a very cheap method an approximate direction to go, sufficient to reach the goal, there is no need for a computationally expensive one. On the contrary, the functional view, common in the 3D reconstruction paradigm, where pictures taken off-line are then fed into a mathematical algorithm that produces a reconstruction, cannot afford approximative results. There is room for hybrid methods, for example where a machine aiming at a 3D reconstruction would decide to move in a certain direction to overcome a perceptual ambiguity or refine a measurement.

In future work, we will test the visual compass on a real robot.

For the bio-inspired compass we will probably need to improve either the segmentation algorithm, the matching algorithm or add a method for selecting triplets that produce the most consistent results, to cope with an environment less easy than the test environment used here.

For the calibrated compass, it could be interesting to mathematically express the assumption about rotation that is done in the other method, and assess if the resulting simplification allows a better precision in difficult situations where too few points are available for example.

We are considering the use of the bio-inspired visual compass as an external direction estimator for the biomimetic models of head-direction cells currently developed by Angelo Arleo at LPPA, Collège de France.

After the visual compass is validated for itself in a robotic context, it should be tested as a complement to a navigation system that needs an orientation estimation, for example the work of Filliat (2001).

We noticed that the bio-inspired visual compass could be extended directly to also estimate a translation, at the cost of knowing the motion performed

during observation.

7 Conclusion

The issue of estimating the motion between two real-world views is simultaneously interesting for 3D world reconstruction as well as for robotics.

In this article, we presented and compared two new methods that estimate the orientation of a panoramic view with respect to a reference view in an *a priori* unknown environment, one biologically inspired in the continuation of the work on insect-inspired robot navigation strategies, the other in the continuation of research on 3D scene reconstruction.

Due to their different origins, the two methods differ in input used, results produced, complexity and computational cost. The 3D computer vision approach makes an extensive use of sharp models of sensors to retrieve a high precision. From a bio-inspired perspective, the calibrated method needs too much information and produces more results and with greater precision than really needed for a robotics application.

We think that there is room for inspiration from each other field, that would either offer more robustness where needed in the bio-inspired field, of more flexibility in the classical field by taking advantage of closer interaction with the real world.

References

S. Baker and S. K. Nayar. A theory of catadioptric image formation. *ICCV*, pages 35–42, 1998.

B. A. Cartwright and T. S. Collett. Landmark learning in bees. *J. Comp. Physiol*, 151:521–543, 1983.

B. A. Cartwright and T. S. Collett. Landmark maps for honeybees. *Biol. Cybern.*, 57:85–93, 1987.

P. Chang and M. Hebert. Omnidirectional structure from motion. In *Proceedings of IEEE workshop on omnidirectional vision*, 2000.

O. Faugeras. *Three dimensional computer vision : a geometric viewpoint*. MIT press, 1993.

O. Faugeras and S. Maybank. Motion from point matches :multiplicity of solutions. Technical Report RR-1157, INRIA, february 1990.

Olivier Faugeras, Long Quan, and Peter Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. In *Computer Vision - ECCV'98, 5th European Conference on Computer Vision, Freiburg, Germany, June 2-6, 1998, Proceedings, Volume I*, volume 1406 of *Lecture Notes in Computer Science*. Springer, 1998. ISBN 3-540-64569-1.

David Filliat. *Cartographie et estimation globale de la position pour un robot mobile autonome.* PhD thesis, Université Pierre et Marie Curie (Paris VI), 2001.

M. Franz, B. Scholkopf, P. Georg, H. Mallot, and H. Bulthoff. Learning view graphs for robot navigation. *Autonomous Robots*, 5:111–125, 1998a.

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, 1998b.

C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International journal of computer vision*, december 2001.

C. Geyer and K. Daniilidis. Paracatadioptric camera calibration. *IEEE Transaction on pattern analysis and machine intelligence*, 24(4), april 2002.

Stéphane Gourichon. A biologically inspired visual compass. Technical report, Université Pierre et Marie Curie (Paris 6), 2003.

Stéphane Gourichon, Jean-Arcady Meyer, and Patrick Pirim. Using coloured snapshots for short-range guidance in mobile robots. *International Journal of Robotics and Automation*, 17(4):154–162, 2002.

C. Harris and M. Stevens. A combined corner and edge detector. In *Proceedings of the fourth Alvey Vision Conference*, pages 153–158, 1988.

Sio-Hoï Ieng, Ryad Benosman, and Jean Devars. Multi-angular dynamic matching for catadioptric views. In *Proceedings of ICCV2003*, september 2003. to appear.

K. Ikeuchi, M. Sakauchi, T. Takahashi, M. Murao, I. Sato, and F. Kai. Panoramic image analysis for constructing virtual cities. *IPSJ Transaction on*

Computer Vision and Image Media Abstract, 42 (SIG13-06), december 2001.

R. Benosman J. Fabrizzio, J.-P. Tarel. Calibration of panoramic catadioptric sensors made easier. In *Proceedings of the IEEE workshop on omnidirectional vision*, 2002.

H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, september 1981.

P. J. Narayanan and T. Kanade. Virtual worlds using computer vision. In *Proceedings of the IEEE and ATR workshop on computer vision for virtual reality based human communications*, pages 2–13, january 1998.

M. E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *Int. Journal of Computer Vision*, 4(3):171–183, May 1990.

T. Svoboda and T. Pajdla. Matching in catadioptric images with appropriate windows and outliers removal. *CAIP*, pages 733–740, 2001.

T. Svoboda, T. Pajdla, and V. Hlavac. Motion estimation using central panoramic cameras. In *Proceedings of the IEEE International conference on Intelligent Vehicules*, pages 335–340, october 1998.

T. Svoboda and Peter Sturm. What can be done with a badly calibrated camera in ego-motion estimation ? Technical Report CTU-CMP-19, Center for Machine Perception, Czech Technical University, Prague, Nov 1996.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2), 1992.

Z. Zhang. Determining the epipolar geometry and its uncertainty. *International Journal on Computer Vision*, 27(2):161–195, 1998.