Photometric Visual Gyroscope for Full-View Spherical Camera

Antoine N. André
Université de Picardie Jules Verne, MIS laboratory, 33 Rue Saint-Leu, 80039 Amiens Cedex, France
antoine.andre@u-picardie.fr

Guillaume Caron*
CNRS-AIST Joint Robotics Laboratory (JRL), IRL, 1-1-1 Umezono, Tsukuba 305-8560, Japan
guillaume.caron@cnrs.fr

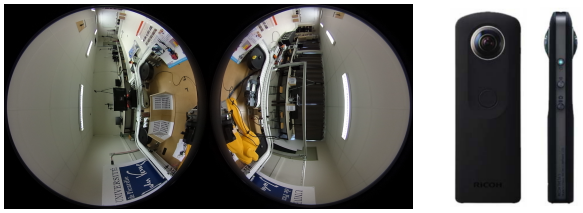
Abstract

This article presents a visual gyroscope based on the spherical representation of dual-fisheye cameras. By using the fully available view of the scene thanks to a dual-fisheye camera and projecting the photometric information on a sphere, a highly precise orientation estimation along with a great convergence domain can be achieved. This method is validated with the help of images taken from the PanoraMIS dataset to evaluate quantitatively the performances.

1. Introduction

A Visual Gyroscope (VG) estimates the 3D orientation of a camera from a captured image with respect to landmarks in structured environments [9] or to a reference image [5]. This classical estimation problem is the heart of many computer vision applications, for example upright correction of spherical panoramas [9], stabilization of 360 degrees (deg) videos [8] and robot motion estimation [3].

VG algorithms consider two kinds of information, direct or indirect. An Indirect VG (IVG) leverages image features, either handcrafted as patches around image points [5, 8] or learnt to estimate the optical flow [7]. Instead, a Direct VG (DVG) almost considers directly pixel brightness of the whole image as input of a 3D rotation optimization method [11]. DVGs almost consider directly pixel brightness because the latter usually undergo a transform to another space before estimating the 3D rotation, *e.g.* a spherical Fourier transform to solve for the rotation with phase correlation [11] or a Mixture of Photometric Potentials (MPP) transform to optimize for the rotation minimizing the Sum-of-Squared-Differences between two MPPs [3].



(a) Dual-fisheye image (b) Ricoh Theta S

Figure 1. (a) Dual-fisheye image taken from the *PanoraMIS* dataset and used as full-view spherical data for the PVG. (b) Dual-fisheye camera used for the data acquisition.

Despite the absence of features, DVGs could recently reach very large estimation domains such as up to 360 deg in static environment [3]. But the rotation estimation error is still of the order of 1 deg (for real-time or close to real-time setups) whereas Neural Networks could recently reach 0.3 deg [7], though within a much tighter estimation domain of $[-5; 5]$ deg around each axis. This error level is still problematic as it gets to be accumulated over time within 360 deg video stabilization [8] or visual odometry [12].

We believe the rather low accuracy of state-of-the-art real-time DVGs is due to the transformation of the pixelic information. Hence, this article proposes a new DVG to improve the rotation estimation accuracy by directly considering pixel brightness of captured images lifted on a sphere, named PVG (Photometric Visual Gyroscope). PVG implements a new hybrid multi-resolution/sub-sampling scheme in two *levels* allowing to decrease significantly the best known estimation errors [7] together with a much larger convergence domain, as assessed by the evaluation made on many images available in the *PanoraMIS* dataset [1].

The rest of the paper presents the PVG algorithm (Sec. 2) and the evaluation of its linearity and accuracy results compared to the state-of-the-art (Sec. 3) before conclusion.

*G. Caron is also with Université de Picardie Jules Verne, MIS, France. This work is carried out as part of the Interreg VA FCE ADAPT project "Assistive Devices for empowering disAbled People through robotic Technologies" (adapt-project.com). The Interreg FCE Programme is a European Territorial Cooperation programme that aims to fund high quality cooperation projects in the Channel border region between France and England. The Programme is funded by the European Regional Development Fund.

2. Spherical Photometric Visual Gyroscope

This section presents first the method to project dual-fisheye images on a sphere (Cartesian representation). Then, the PVG's optimization algorithm is detailed.

2.1. Spherical representation of dual-fisheye images

Dual-fisheye cameras can easily be modeled using a unified central projection model involving two main steps that are *(i)* the projection on a unit sphere and *(ii)* a perspective projection on the image plane [4]. The process used to calibrate and combine the two dual-fisheye cameras have already been described in [1]. Once the calibration of the two cameras have been done, the two parts of the dual-fisheye images are then projected on a single sphere.

In order to do this, an isocahedron (20 equilateral triangles and 12 vertices) is firstly considered, before being subdivided n times. The unit norm points ${}^c\mathbf{X}_{\mathcal{S}_i} = [X_{\mathcal{S}_i}, Y_{\mathcal{S}_i}, Z_{\mathcal{S}_i}]^\top \in \mathbb{R}^3, i \in [1, P]$ resulting from this subdivision are uniformly spaced on the whole sphere surface [10] (partially presented in Fig. 2).

2.2. Alignment method based on pixelic information

Consider $I_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}}) \in [\![0, 255]\!]$, the spherical image brightness[1] in the spherical image $I_{\mathcal{S}}$ acquired at the orientation $\mathbf{r} \in \mathbb{R}^3$ (i omitted for compactness). The input of the PVG is the stacking of all spherical brightness as vector $\mathbf{I}_{\mathcal{S}}(\mathbf{r}) \in [\![0, 255]\!]^{P\times 1}$:

$$\mathbf{I}_{\mathcal{S}}(\mathbf{r}) = [\ldots, I_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}}), \ldots]^\top. \tag{1}$$

Vector $\mathbf{r}$ represents the camera 3D rotation with axis $\mathbf{w} = [w_X, w_Y, w_Z]^\top \in \mathbb{R}^3 : ||\mathbf{w}|| = 1$ and angle $\theta \in \mathbb{R}$.

The brightness vector $\mathbf{I}^*_{\mathcal{S}}$ is built from the desired image $I^*_{\mathcal{S}}$, as $\mathbf{I}_{\mathcal{S}}(\mathbf{r})$ is built with (1) from $I_{\mathcal{S}}$. PVG is designed to minimize the Sum of Squared Differences (SSD):

$$\mathscr{C}(\mathbf{p}) = \frac{1}{2}||\mathbf{I}_{\mathcal{S}}(\mathbf{r}) - \mathbf{I}^*_{\mathcal{S}}||^2, \tag{2}$$

within a Levenberg-Marquardt optimization method computing iteratively the rotation increment $\delta\mathbf{r} \in \mathbb{R}^3$ from the spherical image Jacobian $\mathbf{L}_{\mathbf{I}_{\mathcal{S}}} \in \mathbb{R}^{P\times 3}$:

$$\delta\mathbf{r}^{(it)} = -\left(\mathbf{H} + \mu\, \mathrm{diag}\,(\mathbf{H})\right)^{-1} {\mathbf{L}_{\mathbf{I}_{\mathcal{S}}}}^\top (\mathbf{I}_{\mathcal{S}}(\mathbf{r}) - \mathbf{I}^*_{\mathcal{S}}), \tag{3}$$

with $\mathbf{H} = {\mathbf{L}_{\mathbf{I}_{\mathcal{S}}}}^\top \mathbf{L}_{\mathbf{I}_{\mathcal{S}}}$, at each iteration it until convergence. In (3), $\mu \in \mathbb{R}^*_+$ is the damping parameter. $\mathbf{L}_{\mathbf{I}_{\mathcal{S}}}$ links the variations of spherical image brightness to the 3 camera pose DoF, thanks to the *Optical Flow Constraint Equation* (OFCE), valid in Lambertian scenes [6]. $\mathbf{L}_{\mathbf{I}_{\mathcal{S}}}$ is the stacking of interaction matrices $\mathbf{L}_{I_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}})}$ evaluated for all ${}^c\mathbf{X}_{\mathcal{S}}$, each composing spherical image gradients

[1] Any other quantization could be considered.

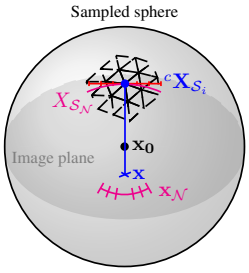


Figure 2. Spherical representation of the dual-fisheye camera: partial view of the subdivided isocahedron, a pixel projected on the sphere (blue), its neighbors along the x direction (red), back-projection of the neighbors to the image plane (pink).

$\nabla I_{\mathcal{S}^c\mathbf{X}_{\mathcal{S}}} \in [\![-255, 255]\!]^3$ to the geometric interaction matrix $\mathbf{L}_{{}^c\mathbf{X}_{\mathcal{S}}} = [{}^c\mathbf{X}_{\mathcal{S}}]_\times$ (skew-symmetric matrix of ${}^c\mathbf{X}_{\mathcal{S}}$ [3]):

$$\mathbf{L}_{I_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}})} = -\nabla I^\top_{\mathcal{S}^c\mathbf{X}_{\mathcal{S}}} \mathbf{L}_{{}^c\mathbf{X}_{\mathcal{S}}}. \tag{4}$$

2.3. Spherical image gradient computation

In order to approximate as closely as possible the gradient directly on the sphere, the three derivatives of $\nabla I_{\mathcal{S}^c\mathbf{X}_{\mathcal{S}}}$ (i.e. $\frac{\partial I_{\mathcal{S}}}{\partial X_{\mathcal{S}}}$, $\frac{\partial I_{\mathcal{S}}}{\partial Y_{\mathcal{S}}}$ and $\frac{\partial I_{\mathcal{S}}}{\partial Z_{\mathcal{S}}}$) are computed using finite differences. The sampling used for this finite difference gradient computation is defined from the image plane to sphere projection $pr_\xi^{-1}()$ (see [2] for the detailed projection function and intrinsic parameters) of the size of one pixel at the principal point of coordinates u_0 and v_0, as follow:

$$\Delta_{X_{\mathcal{S}}} = \Delta_{Y_{\mathcal{S}}} = \Delta_{Z_{\mathcal{S}}} = \left\| pr_\xi^{-1} \begin{pmatrix} u_0 + 1 \\ v_0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\| \tag{5}$$

Then, for each spherical pixel, three neighborhoods are defined, one for each axis. Each of these N-neighbors (with $N = 6$ typically) are calculated accordingly to the following equation that presents the case of the neighbors computation along the sphere abscissa axis:

$$X_{\mathcal{SN}} = \left\{ \frac{\left(X_{\mathcal{S}} + k\Delta_{X_{\mathcal{S}}}, Y_{\mathcal{S}}, Z_{\mathcal{S}}\right)^{\mathrm{T}}}{\|\left(X_{\mathcal{S}} + k\Delta_{X_{\mathcal{S}}}, Y_{\mathcal{S}}, Z_{\mathcal{S}}\right)^{\mathrm{T}}\|} \right\} \quad -\frac{N}{2} \leq k \leq \frac{N}{2}, k \neq 0 \tag{6}$$

The same procedure is then applied similarly for the neighbors along the two others directions $Y_{\mathcal{SN}}$ and $Z_{\mathcal{SN}}$. With this neighbors computation, a fine cartesian spherical gradient evaluation of the images can be done, completing the PVG optimization expression presented previously.

2.4. Hybrid visual alignment method

As PVG uses pixels lifted to the equivalent sphere, the image resolution should match the isocahedron subdivision level. For n subdivisions levels, $P = \frac{1}{2}(20 \cdot 4^n) + 2$ points form the sphere.

Level of subdiv. n	Nb. of features P	Scale factor s	Resulting size	New Nb. of pixels
Full image	—	1	1280×720	921600
5	10242	9.48	135×76	10260

Table 1. Re-sampling table of images to match the number of spherical pixels depending on the icosahedron subdivision level.

Input images usually don't match the number of spherical points by a large factor (see Tab. 1). However, by considering the full image resolution, better accuracy can be achieved, thanks to an oversampling of image gradients, whereas the convergence domain is smaller (as assessed by experiments, Sec. 3). Inversely, re-sampling the image accordingly to the (lower) number of spherical pixels (see Tab. 1, 2nd row) leads to a larger convergence domain, though for a poorer accuracy.

To combine both the adequate spherical sampling and the high accuracy granted by the super sampling, PVG implements a hybrid method: first, PVG runs with re-sampled imaged, then it uses the original images (i.e. oversampled).

3. Performance of the measure

The method has been implemented leveraging the dual-fisheye camera model and tools of the libPeR_base `C++` library[2] to study PVG's performance. In order to do so, the publicly available dataset *PanoraMIS*[3] has been used [1]. The presented experiments use the third sequence (Sequence 3) that contains images of multiple 360 deg pure rotation around a single axis with a rotation increment of 2.5 deg between successive images. This acquisition is repeated five times at different locations (named CP_i for $i \in [0, 4]$), thus giving a set of 720 images (144 images per full rotation). PVG's results are also compared with the prior visual gyroscope leveraging the MPP of spherical images (Sec. 1). In all the experiments $n = 5$ subdivision levels for the sphere, and the damping parameter (3) is set as $\mu = 0.01$.

3.1. Accuracy

The first experiment evaluates PVG's accuracy (closeness of agreement between a test result and the accepted reference value). For this, the orientation between each successive pair of images $\mathbf{I_k}$ and $\mathbf{I_{k+1}}$ is estimated. All successive rotation ground truths are 2.5 deg plotted as the black line in Figure 3 that shows a high accuracy with an average estimation error of 0.0981 deg (std: 0.0442 deg), when considering the full image. Tab. 2 shows the contribution in accuracy of PVG over the MPP-based gyroscope [3].

However, when considering images with the same number of pixels as the number of points sampling the sphere, the accuracy decreases to reach an average error of 0.79 deg. This result confirms the interest in considering full resolution images to achieve the best accuracy.

[2] https://github.com/PerceptionRobotique/libPeR_base
[3] https://mis.u-picardie.fr/~panoramis

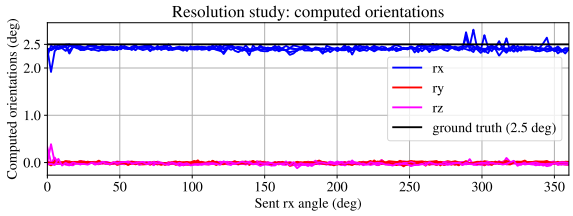


Figure 3. Accuracy test carried over the whole images data-set, with a ground truth of 2.5 deg between successive images.

3.2. Linearity and convergence domain

To avoid being in a particular case when studying the convergence domain of the method, the experiments are using the whole set of images, giving a more accurate experimental evaluation of the convergence width. As the dataset contains full rotations around a single axis at five different locations, the experiment is split in five sub-studies, one for each considered location of the camera. For each location, the image at mid-rotation (180 deg) is set the reference image and, then, PVG aligns its spherical representation to each image in the $[-180, 180]$ deg range.

To highlight the interest of the hybrid alignment (Sec. 2.4), the experiment is split in two parts, in order to study the convergence domains when either, first, considering re-sampled images or, second, full images. Fig. 4 presents the aggregation of the five sub-experiments to provide a global study of the convergence domain width. On the other hand, Fig. 5 presents the results of linearity when full resolution images are considered

As a summary, if Fig. 4 highlights a loss in accuracy (with an error level of 0.79 deg) compared to using the full images (especially when considering small angles that are not detected), it is nothing but the counterpart of a ± 45 deg convergence domain, much larger than the ± 12.5 deg convergence domain when using full images (Fig. 5).

This convergence domain can be summarized with the representation of the cost function (2). Fig. 6a shows it together with the domain where the gradient descent-like method of 3 is results in an accurate orientation estimation.

Used method	PVG	MPP
Accuracy (deg)	0.0981	3.15
Convergence domain (deg)	± 45	± 180
Computing time (ms)	883.49	57160.00

Table 2. PVG (hybrid) versus MPP-based gyroscope [3], both applied to the pure rotation images of the *PanoraMIS* dataset.

(a) Resulting mean convergence and linearity

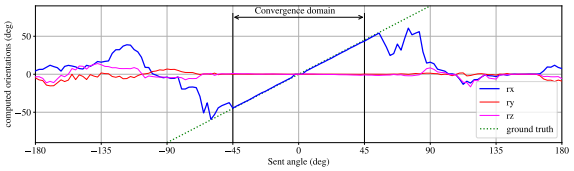


Figure 4. Convergence and linearity study in case of images re-sampled before estimating their orientations once aggregated to provide a global conclusion.

These results show the main interest in considering the hybrid PVG to benefit a greater convergence domain and a high accuracy. Though the 90 deg convergence domain width of the hybrid PVG is 4 times narrower than the domain of the MPP-based gyroscope, the hybrid PVG increases 37 times the estimated rotation accuracy.

3.3. Robustness to translation

With the characteristics of PVG known in the presence of pure rotation, a final experiment studies the robustness to translation. Fig. 6b presents the results of hybrid PVG to align the reference image of the first location CP_0 with images at any location within the demonstrated convergence domain. PVG's accuracy decreases as the distance to CP_0 increases, with a mean error of 2.57 deg overall. This error increase, compared to considering images of pure rotations only, shows the sensitivity of PVG to translation, however still more accurate than the MPP-based gyroscope in the presence of pure rotation (Tab. 2).

4. Conclusion

This article presented a Photometric Visual Gyroscope based on dual-fisheye images. By projecting the captured images and their photometric information on a sphere, the method presents a high level of accuracy of up to 0.0981 deg in a ± 45 deg convergence domain. To achieve these results, this article presented a method to account for the image resolution, before its projection on the sphere, enhancing its convergence domain. Future works target the validation of the method by extending the gyroscope to the servoing of a robot arm embedding a dual-fisheye camera.

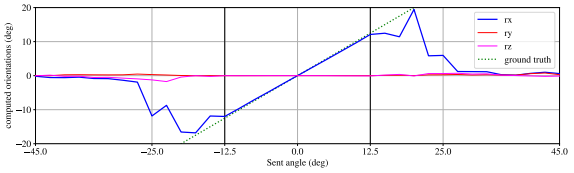


Figure 5. Linearity test carried with full resolution images.

(a) Aggregation of all cost evaluations

(b) Robustness to translation in orientation estimation

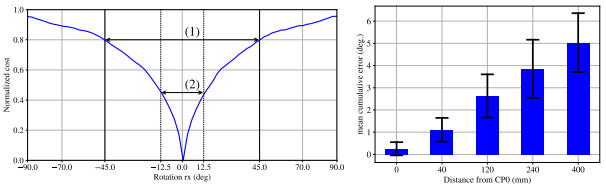


Figure 6. (a) Cost function evaluation between a reference image and images in the $[-90, 90]$ deg range: convergence domain with (1) re-sampled images and with (2) full resolution images. (b) Mean and std orientation estimation error, function of 5 locations.

References

[1] H.-E. Benseddik, F. Morbidi, and G. Caron. Panoramis: An ultra-wide field of view image dataset for vision-based robot-motion estimation. *The Int. J. of Robotics Research*, 39(9):1037–1051, 2020.

[2] G. Caron, E. Marchand, and E. Mouaddib. Photometric visual servoing for omnidirectional cameras. *Autonomous Robots*, 35(2-3):177–193, Oct. 2013.

[3] G. Caron and F. Morbidi. Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials. In *IEEE Int. Conf. on Robotics and Automation*, pages 820–827, Brisbane, Australia, May 2018.

[4] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *European Conf. on Computer Vision*, pages 445–461, 2000.

[5] H. Hadj-Abdelkader, O. Tahri, and H.-E. Benseddik. Closed form solution for rotation estimation using photometric spherical moments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 627–634, 2018.

[6] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185 – 203, 1981.

[7] D. Kim, S. Pathak, A. Moro, A. Yamashita, and H. Asama. Self-supervised optical flow derotation network for rotation estimation of a spherical camera. *Advanced Robotics*, 35(2):118–128, 2021.

[8] J. Kopf. 360° video stabilization. *ACM Trans. Graph.*, 35(6), nov 2016.

[9] H. Li, J. Zhao, J.-C. Bazin, and Y.-H. Liu. Quasi-globally optimal and near/true real-time vanishing point estimation in manhattan world. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 44(3):1503–1518, 2022.

[10] S. Li and Y. Hai. A full-view spherical image format. In *Int. Conf. on Pattern Recognition*, pages 2337–2340, 2010.

[11] A. Makadia and K. Daniilidis. Rotation recovery from spherical images without correspondences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(7):1170–1175, 2006.

[12] S. Pathak, A. Moro, H. Fujii, A. Yamashita, and H. Asama. Distortion-resistant spherical visual odometry for UAV-based bridge inspection. In *Int. Workshop on Advanced Image Technology*, volume 11049, pages 316 – 321, 2019.