# BiCamSLAM: Two times mono is more than stereo

Joan Solà, André Monin and Michel Devy
LAAS-CNRS
Toulouse, France
{jsola,monin,michel}@laas.fr

***Abstract*— This paper is an invitation to use mono-vision techniques on stereo-vision equipped robots. By using monocular algorithms on both cameras, the advantages of mono-vision (bearing-only, with infinity range but no 3D instant information) and stereo-vision (3D information only up to a limited range) naturally add up to provide interesting possibilities, that are here developed and demonstrated using an EKF-based monocular SLAM algorithm. Mainly we obtain: a) fast 3D mapping with long term, absolute angular references; b) great landmark updating flexibility; and c) the possibility of stereo rig extrinsic self-calibration, providing a much more robust and accurate sensor. Experimental results show the pertinence of the proposed ideas, which should be easily exportable (and we encourage to do so) to other, more performing, vision-based SLAM algorithms.**

## I. Introduction

On account of all that has been –and is continuing to be– published, Simultaneous Localization and Mapping (SLAM) is almost a full-right discipline inside robotics. With the objective of interactively perceiving and modeling the robot's environment and keeping localized while moving, SLAM must put into play robust and scalable real-time algorithms that fall into three main categories: estimation (in the prediction-correction sense), perception (with its signal processing), and decision (what we could call *strategy*). The problem is now well understood: the whole decade of the nineties was devoted to solve it in 2D with range and bearing sensors, and big progress was achieved in the estimation side –to the point that some claim that the subject is approaching saturation, which is a kind of optimality obtained by evolutionary mechanisms– and recent research has focused on the perception side. The best example of this trend is vision-based SLAM, and specially mono-vision SLAM, where bearings-only measurements reduce observability, thus delaying good 3D estimates. Davison [1] showed real-time feasibility of mono-vision SLAM with affordable hardware, using the original Extended Kalman Filter (EKF) SLAM algorithm. We showed in [2] that undelayed landmark initialization (*i.e.* mapping the landmarks from their first, partial observation) was needed when considering remote landmarks and/or singular trajectories of the camera. In this direction, an interesting work [3] has recently appeared that uses the constant-time FastSLAM2.0 algorithm [4].

Stereo-vision SLAM has also received considerable attention. The ability to directly obtain 3D measurements allows us to use the best available SLAM algorithms and obtain very good results with little effort in the conceptual parts. Good works on stereo SLAM usually put the accent on advanced image processing, that may require highly specialized programming (we think about the real-time construction and querying of big data bases and the hardware implementation of robust feature trackers of [5] for instance). The drawback of stereo-based systems is a limited range of 3D observability (the dense-fog effect: remote objects cannot be considered), and that they strongly depend on precise calibrations to be able to extend it.

This work is a conceptual work. Although specific solutions are developed from previous work of the authors [2] and experiments are shown, its main contribution is to bring the powers of mono- and stereo-vision SLAM together with the following benefits: *1)* important objects for reactive navigation, which are close to the robot, are rapidly mapped with stereo-like triangulation; *2)* good orientation localization is achieved with bearings-only measurements of remote landmarks (thus eliminating the dense-fog effect); and *3)* updates can be performed on any landmark that is only visible from one camera. Additionally, *4)* precise previous calibration of the stereo rig extrinsic parameters of the cameras is no longer necessary; because *5)* dynamic self-calibration of these stereo rig extrinsic parameters can be incorporated, thus making such an intrinsically delicate sensor more robust and accurate. These two latter assertions are demonstrated with a simplified self-calibration procedure based on the same EKF used for SLAM. The key for all these benefits is using mono-vision algorithms in both cameras instead of a stereo one: we get enhanced observability with a much greater flexibility.

This paper is organized as follows: Section II revises and actualizes the necessary material for mono-vision SLAM and presents the main ideas that will be exploited later. Section III explains how to set up Bi-Camera SLAM and a simple, EKF-based self-calibration solution. Section IV shows the experimental results and finally section V gives conclusions and future directions.

## II. Vision-based SLAM

We consider as vision-based SLAM those SLAM approaches where the external world is measured only by means of vision sensors: the cameras. This is true regardless of whether we use other kind of sensors for the robot motion: in mono-vision SLAM, these other sensors will provide the scale factor which is missing in bearings-only measures.

In a large sense, SLAM may be viewed as a generic set of procedures to recover, online and incrementally, the geometrical properties of the surrounding world. Thus only geomet-

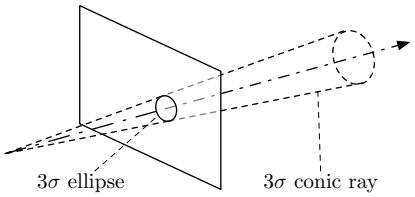


Fig. 1. The conic ray back-projects the elliptic representation of the gaussian 2D measure. It extends to infinity.

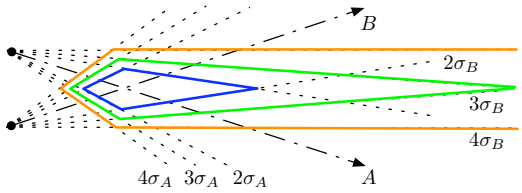


Fig. 2. Different regions of intersection for $4\sigma$ (*orange*), $3\sigma$ (*green*) and $2\sigma$ (*blue*) ray widths when the outer $4\sigma$ bounds are parallel. The angle between rays axes $A$ and $B$ is $\alpha = 4\,(\sigma_A + \sigma_B)$.

rical information from the images is explicitly exploited. This resumes to the position of certain features in 2D images (here 2D points), which we will uniquely associate to particular landmarks in the 3D space. Photometrical information is only used to perform this association via feature matching.

### A. 3D observability from vision

When a new feature is detected in the image, the back-projection of its noisy-measured position defines a conic-shaped *pdf* for the landmark position, called *ray*, that extends to infinity (Fig. 1).

Just a couple of ideas (not to be strictly interpreted) to help to understand the observability concept that we use. Let us consider two features extracted from two images and matched because they correspond to the same landmark: their back-projections are two conic rays $A$ and $B$ that extend to infinity. Their angular widths can be defined as a multiple of the standard deviations $\sigma_A$ and $\sigma_B$ of the angular errors (Fig. 2), which depend on the accuracy of the cameras' poses (extrinsic parameters), on the cameras' angular resolution (intrinsic parameters) and on the accuracy of the methods used to extract and match points. We say that the landmark's depth is observed if the region of intersection of these rays is *a)* closed and *b)* sufficiently small. If we consider, for example, the two external $4\sigma$ bounds of the rays to be parallel, then we can insure that the $3\sigma$ intersection region (which covers 98% probability) is closed and that the $2\sigma$ one (covering 74%) is small. The depth's sigma-to-mean ratio is in such case better than $0.25$. The angle $\alpha$ between the two rays axes is then $\alpha = 4(\sigma_A + \sigma_B) = constant$.

We can plot (Fig. 3) the locus of those points where the two angular observations differ exactly in this angle $\alpha$. Inside the obtained circular region, depth is observable; outside it is not. Given overall angular uncertainties $\sigma_A$ and $\sigma_B$, this region's radius is directly proportional to the distance between the two cameras. In 3D, revolution of this region around the axis joining both cameras produces something

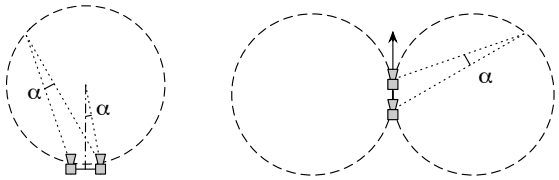


Fig. 3. Simplified depth observability regions in a stereo rig (*left*) and a camera traveling forward (*right*). The angle $\alpha$ is the one that assures observability via difference of points of view.

like a torus-shaped region. In a stereo configuration or for a lateral motion of a moving camera like in aerial images (Fig. 3 *left*), this region is in front of the sensor. Beyond the region's border (remark that, unlike range scanners, a camera is capable of sensing objects at infinity) stereo is at no profit: if we want to consider distant landmarks we have to use mono-vision techniques.

In mono-vision, this paper wants to specifically consider singular motions, *ie.* a single camera moving forward (Fig. 3 *right*). In the motion axis, depth recovery is simply impossible. Close to and around this axis, which in robotics is typically the zone we are interested in, observability is only possible if the region's radius becomes very large. This implies the necessity of very large displacements of the camera during the initialization process, something that can only by accomplished with undelayed initializations.

Combining both mono- and stereo-vision we get an instant observability of close frontal objects while still utilizing the information of distant ones: the first beneficiary is the robot localization as we will dispose of long term absolute angular references. It is known that it is precisely the accumulation of angular errors in odometry which makes simple SLAM algorithms (such as EKF-SLAM) become inconsistent [6] and fail. Thus, this long term observability will improve EKF-SLAM performance.

### B. Undelayed mono-vision SLAM

The core algorithm of this work is FIS-SLAM, a bearings-only EKF-based SLAM algorithm presented in [2], which is briefly described as follows. FIS-SLAM focuses on an undelayed way to initialize landmarks, which at the first observation are only partially observed. For this, the conic ray is first truncated at minimum and maximum considered depths, and then approximated by a geometric series of Gaussian members. All these members are initialized in the EKF-SLAM map as if they were different landmarks, and are assigned a uniform initial probability or *weight* $\Lambda_i$. As new observations are incorporated, these weights will evolve with time following the likelihood update form $\Lambda_i^+ = \Lambda_i\,\lambda_i$, where $\lambda_i$ is the likelihood of member $i$ given the current measurement, $\Lambda_i$ is the member's weight, and $(\cdot)^+$ denotes "the updated value of $(\cdot)$". The ray's members are progressively deleted as new observations make their weights drop below a certain threshold, and the remaining ones are used to correct the SLAM map by means of a special EKF-based update scheme named Federated Information Sharing (FIS). Inspired by the Principle of Measurement

Reproduction [7], FIS performs a federated sharing of the information provided by the observation among all ray's members before applying an EKF update on each one of them. As all members are initialized from the beginning, partially observed remote landmarks may be used as long term angular references. The landmark is considered fully 3D observed when only one member is left. Many other facts on FIS-SLAM are not relevant for the understanding of the present work.

More recent works [3], [8] use a newer and more mathematically defendable solution for undelayed initialization based on an inverse parametrization of the landmark's depth.

### C. Feature detection and matching

For feature detection, a heuristic strategy is used to select a region of interest in the image.[1] The strongest Harris point [9] in the region is selected for landmark initialization. Its associated ray is calculated and initialized in the map. A small rectangular region or *patch* around the point is stored as the landmark's appearance descriptor, and the current pose of the robot is memorized.

For matching, we follow the *active search* approach (also referred to as *top-down*) [1], [3]: At subsequent observations, the joint estimates of landmark and robot positions are projected into the image (giving what is called the expectation) and are used to draw the $3\sigma$ ellipses in the image (or sets of ellipses in the case of rays) inside which the feature will be searched. A predefined number of landmarks with the biggest ellipse surfaces are selected as those being the most interesting to be measured. These surfaces are compared by means of the expectation covariance's determinant (in the case of rays we just take the biggest determinant of its members). Then the stored patch is warped (zoomed and rotated) the amounts defined by the change in the robot position. A search for the best correlation of this modified patch inside the elliptic region and a final parabolic interpolation with its cardinal neighbors provide a sub-pixellic measurement. This approach combines the simplicity of patch descriptors and correlation-based scans with the robustness of invariant matching: instead of invariant descriptors like [10], [11], we appropriately *vary* them before each scan using the information available in the map. False matches are also drastically minimized as they will normally fall outside the predicted ellipses.

## III. Bi-camera SLAM

Bi-camera vision *is not* stereo vision. It is just two times mono-vision that takes advantage of the enhanced observability that instant ray triangulation provides, like stereo-vision does. We avoid image rectification, and allow us to use raw distorted images taken from distinct cameras, in any number or configuration, with the only condition of having overlapping fields of view. In this paper, however, we use a classical stereo rig of two nominally equal cameras, with individually calibrated intrinsic parameters (including radial distortion), and nominal, uncalibrated extrinsic parameters.

[1] A simple one: divide the image with a grid. Select a grid element with no landmarks inside and use it as the initialization region.

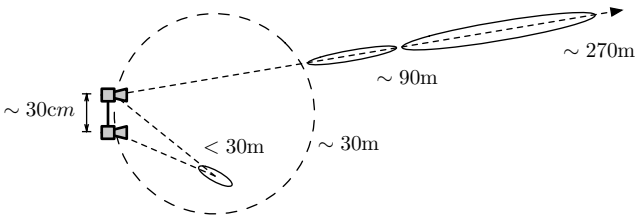


Fig. 4. BiCam initialization. Use stereo capabilities when possible. Use mono otherwise. When combined with self-calibration, get rays ranging hundreds of meters with very few members.

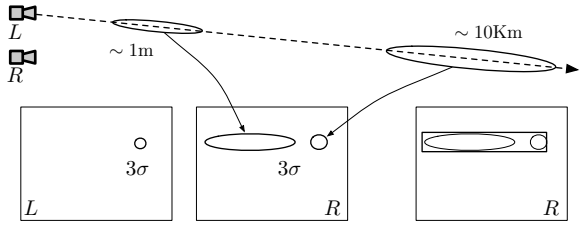


Fig. 5. The 2-member ray in BiCam initialization. The left- and right- $3\sigma$ projections, and the scan region in the right image.

### A. Landmarks initialization

As a general idea, one can simply initialize landmarks following mono-vision techniques from the first camera, and then observe them from the second one: we will determine their 3D positions with more or less accuracy depending on if they are located inside or outside the stereo observability region. Understanding the angular properties that generated these observability regions (section II-A), we can *a-priori* evaluate, from both images, whether each landmark is fully 3D-observable or not (see next paragraph). *1)* If it is fully observable, it is clear that initializing the whole ray and then deleting all but the right members is not so clever. Better, we compute its distance by triangulation, and initialize a "ray" of one single member at this distance using one of the views. Then we update it with the second view to refine its position. *2)* If it is not fully observable, a ray is initialized with its closest member already outside the region. As the farther member's distances follow a geometric series, we easily reach ranges of several hundred meters with very few members (Fig. 4). The ray is immediately updated with the observation from the other camera.

A detailed description of the observability evaluation method is illustrated in Fig. 5. Assume a new feature is detected in the left image. We define (without initializing it) a 2-members ray in the left camera's frame: one member $\{\bar{p}_1; P_1\}$ is at the minimum considered distance and the other $\{\bar{p}_\infty; P_\infty\}$ at the maximum, virtually at infinity. This ray is projected onto the right image: the nearby member becomes an elongated ellipse; the remote one, that projects exactly at the vanishing point of the ray, is a rounded, smaller ellipse. Let $H_p$ and $H_c$ be the Jacobian matrices of the right camera observation function $h(p, c)$ with respect to the point position $p$ and the right camera pose $c$. Let $R$ be the covariances matrix of the measurements and $C$ that of the camera pose uncertainty. Define $H_1 \doteq [H_p|_{\bar{p}_1,\bar{c}} \; H_c|_{\bar{p}_1,\bar{c}}]$,

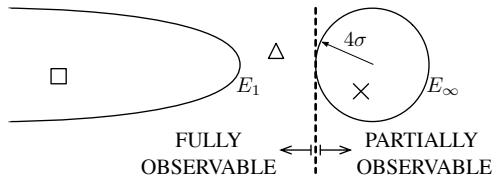


Fig. 6. Deciding on 3D observability. A $4\sigma$ criterion is *a-priori* reasoned in the 2D image plane. The measure marked "cross" corresponds to a landmark outside the stereo observability region. Landmarks measured "square" and "triangle" are inside.

$H_\infty \doteq [H_p|_{\bar{p}_\infty,\bar{c}}\ H_c|_{\bar{p}_\infty,\bar{c}}]$, $Q_1 \doteq \mathrm{diag}(P_1, C)$ and $Q_\infty \doteq \mathrm{diag}(P_\infty, C)$. The projected $n\sigma$ ellipses are centered at $\bar{e}_1 = h(\bar{p}_1, \bar{c})$ and $\bar{e}_\infty = h(\bar{p}_\infty, \bar{c})$ and are described by the expectations' covariances matrices

$$E_1 = H_1\, Q_1\, H_1^\top + R \tag{1}$$

$$E_\infty = H_\infty\, Q_\infty\, H_\infty^\top + R. \tag{2}$$

The region including both $3\sigma$ ellipses is scanned for a feature match. The found pixel $y$ is sent to the following $4\sigma$ test (Fig.6), equivalent to that in section II-A: *the measured landmark is fully 3D observable if and only if the measured feature falls strictly at the left-hand side of the $E_\infty$ ellipse's leftmost $4\sigma$ border*. If we write the measured pixel as $y = [y_h, y_v]^\top$, and the remote expectation as

$$\bar{e}_\infty = \begin{bmatrix} \bar{e}_{\infty,h} \\ \bar{e}_{\infty,v} \end{bmatrix} \qquad E_\infty = \begin{bmatrix} \sigma^2_{\infty,h} & \sigma^2_{\infty,hv} \\ \sigma^2_{\infty,hv} & \sigma^2_{\infty,v} \end{bmatrix},$$

where $(\cdot)_h$ denotes horizontal coordinates, then this criterion resumes simply to

$$y_h < (\bar{e}_{\infty,h} - 4\,\sigma_{\infty,h}) \iff \text{3D OBSERVABLE.} \tag{3}$$

The landmark is then initialized either as a single point or as a ray as indicated above. Notice that this method inherently accounts for arbitrary extrinsic parameters accuracies: the size of $E_\infty$ will vary accordingly, and hence the 3D-observable region bounds too. This naturally allows us to self-calibrate these extrinsic parameters.

### B. Stereo rig self-calibration

Stereo rigs are mechanically delicate, specially for big base lines. We believe that stereo assemblies are only practical if they are very small or if their main extrinsic parameters are continuously self-calibrated. Outdoors operation will often impose this second case.

Not all six extrinsic parameters (three for translation, three for orientation) need to be calibrated. In fact, the notion of *self-calibration* inherently requires the system to possess its own gauge. In our case, the metric dimensions or *scale factor* of the whole world-robot system can only be obtained either from the stereo rig base line (and notice that then it is absurd to self-calibrate the gauge!) or from the odometry sensors, which often are much less accurate than any rude measurement we could make of this base line. Additionally, vision measurements are much more sensible to the cameras' orientations than to any of the other two translation parameters (cameras are actually angular sensors). This means that vision measurements will contain little information about these translation parameters. In consequence, self-calibration should only concern orientation, and more precisely, the orientation of the right camera with respect to the left one. The relative error of the overall scale factor will mainly be the relative error we did when measuring the rig's base line.

We have used a very simple self-calibration solution which has given promising results: we just add three angles (or any other orientation representation we are familiar with) to the EKF-SLAM state vector (not forgetting the Jacobians of all involved functions with respect to them) and let EKF make the rest. The evolution function of the extrinsic parameters is simply $c^+ = c + \gamma$, where $\gamma$ is a white, Gaussian, low energy process noise that accounts for eventual de-calibrations (due to vibrations or the like). For short-duration experiments we set $\gamma = 0$. This solution lacks some robustness but is included here as an illustration of the BiCam capability of working with on-line extrinsic calibration. This fact (this lack of robustness) is further discussed in sections IV and V.

### C. Updates

Thanks to the mono-vision formulation, updates can be performed at any mono-observation of landmarks. This includes any nearby or remote landmark that is only visible from one camera.

As indicated in II-C, the determinant of the expectation's covariances matrix is a measure of the information we will gain when measuring a particular landmark. This is so because the uncertainty in the measurement space can be associated to the surface of the corresponding ellipse, which is proportional to the square root of this determinant. Therefore, we suggest as a first step to organize all candidates to be updated in descending order of expectation determinants, without caring if they are points or rays, or in the left- or right- image, and update at each frame a predefined number of them (usually around 10). A second group of updates should be performed on remote landmarks (points or rays) to minimize the angular drift. Updates are processed sequentially, with all Jacobians being re-calculated at each individual update to decrease linearization errors.

## IV. Experiments

Some indoor experiments are presented here to illustrate the proposed ideas. A robot with a stereo head looking forward is run for some 15m in straight line inside the robotics lab at LAAS (Fig. 7). Over 500 image-pairs are taken at approximatively 5Hz frequency. The robot approaches the objects to be mapped, a situation that is common in mobile robotics but that presents observability difficulties for mono-vision SLAM because the trajectory is singular. The stereo rig consists of two intrinsically calibrated cameras with $55^\circ$ FOV at $512 \times 384$ pixels resolution. They are arranged nominally in parallel, separated $330\mathrm{mm}$ and slightly heading $5^\circ$ down. The left camera is taken as reference, and the orientation of the right one is initialized with an uncertainty of $1^\circ$ standard deviation. A simple

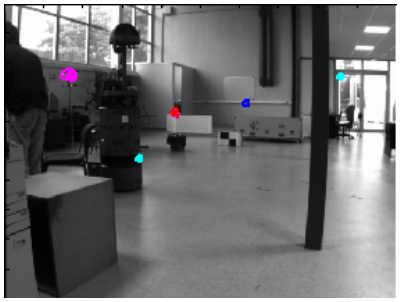

Fig. 7. The LAAS robotics lab. The robot will approach the scene in a straight forward trajectory.

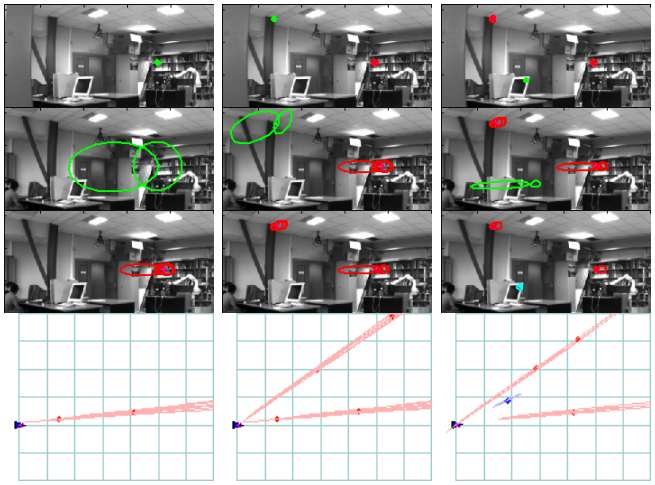

Fig. 8. Initialization sequence in the presence of extrinsic self-calibration. The initialization sequence for the three first frames, one per column, is shown. Start at column 1: a 2-member ray (*green*) is defined from the left view (*top row*). It is projected onto the right image (*second row*). The two $3\sigma$ ellipses (*green*) define a region which is scanned for a feature match. If this match is not on the left of the right-hand $4\sigma$ ellipse (33% bigger than drawn), the landmark is not 3D observable and is initialized as a ray (*red, third row*). The resulting map is shown (*bottom row*, the grid at 2m spacing). Subsequent observations (*columns 2 and 3*) decrease calibration uncertainty and hence the ellipses sizes too. After 3 frames a newly detected landmark at a similar range is already 3D observable, thanks to the enhanced extrinsic precision, and can be initialized as a single Gaussian (*blue*).

2D odometry model is used for motion predictions where the added uncertainty position and orientation variances are proportional to the measured displacement. These experiments want to particularly show the self-calibration and the initialization mechanisms where the landmarks can be mapped with either a single Gaussian or a ray depending on the 3D observability. Results on the accuracy of the resulting map are also reported. Illustrating videos can be found on the author's web page at `http://www.laas.fr/~jsola/objects/videos/icra07/video-N.mov` where `N` is a video number.

We show the dynamic observability decision criterion with extrinsic self-calibration. The first three frames of the sequence are detailed in the three columns of Fig. 8. Observe how, on the first frame, extrinsic self-calibration is poor and results in big decision ellipses, giving place to initializations of nearby landmarks in the form of rays. Observations from the right camera refine the extrinsic precision and subsequent

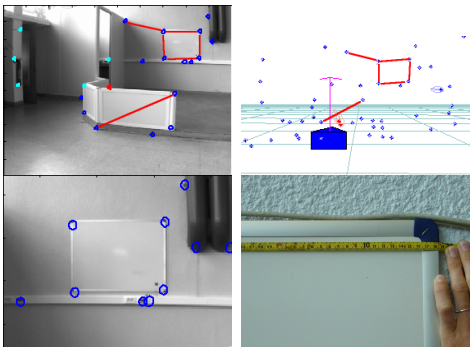

Fig. 9. Metric mapping. The magnitudes of some segments in the real lab are compared to those in the map (red lines). Thirteen points at the further end wall are tested for co-planarity.

TABLE I

MAP TO GROUND TRUTH COMPARISON.

| segment | location | real (cm) | mapped | std. dev. |
|---|---|---|---|---|
| A | board | 119 | 119.6 | 0.81 |
| B | board | 86 | 84.3 | 0.87 |
| C | board | 115 | 114.8 | 1.11 |
| D | board | 88 | 89.0 | 0.72 |
| E | wall | 134 | 132.5 | 0.91 |
| F | fence | 125 | 124.5 | 1.21 |

decision ellipses become smaller. On the third frame, the stereo rig is already quite accurate and is able to fully observe the 3D position of new landmarks. Previous rays are continuously observed from both cameras and will rapidly converge to single Gaussians as self-calibration enhances accuracy.

To contrast the resulting map against reality, two tests are performed: metric consistency and planarity (Fig. 9). *1)* The four corners of the white board are taken with other nine points on the end wall to test co-planarity: the mapped points are found to be coplanar within 4.9cm of standard deviation. *2)* The lengths of the real and mapped segments marked in red in Fig. 9 are summarized in table I. We observe consistent estimates with errors in the order of one centimeter for landmarks that are still about 4m away from the robot.

A typical evolution of the three self-calibrated Euler angles is illustrated in Fig. 10. We observe the following behavior: *1)* Pitch angle (cameras tilt, $5^\circ$ nominal value) is observable from the first matched landmark. It rapidly converges to an angle of $4.87^\circ$ and remains very stable during the whole experiment. *2)* Roll angle is observable after at least two landmarks are mapped. It may take some frames for this condition to arrive but then it also converges relatively fast and quite stably. *3)* Yaw angle is very weakly observable because it is coupled with the distance to the landmarks: both yaw angle and landmark depth variations produce a similar effect in the right image, *i.e.* the feature moves following the landmark's epipolar line. However, yaw does start converging from the initial observations, but after some frames it does it insecurely and slowly: see how from frame 15 onwards yaw uncertainty is already bigger than roll one, which started converging later. As it can be appreciated in Fig. 10 *right* yaw

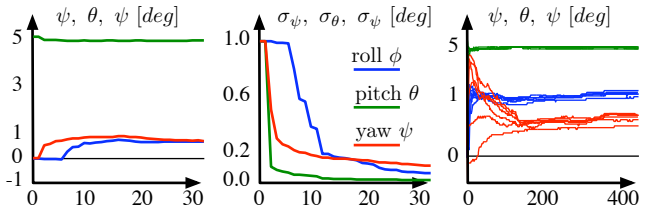


Fig. 10. Evolution of the self-calibrated orientation angles. *Left*: Euler angles during the first 30 frames. *Center*: their standard deviations. *Right*: 6 superimposed calibrations showing calibration repeatability and stability along 400 frames, with reasonable convergence at about frame 150 (300% zoomed in, green curve out of scale).

TABLE II

SELF-CALIBRATION ERRORS WITH RESPECT TO OFF-LINE CALIBRATION.

| Angle | off-line | self-cal. | error | $\sigma$ (stat.) | $\sigma$ (estim.) |
|---|---|---|---|---|---|
| roll $\phi$ | 0.61° | 0.60° | −0.01° | 0.038° | 0.021° |
| pitch $\theta$ | 4.74° | 4.87° | 0.13° | 0.006° | 0.006° |
| yaw $\psi$ | 0.51° | 0.33° | −0.18° | 0.108° | 0.018° |

only shows reasonable convergence after 150 frames. Before, yaw is not very stable neither very repeatable among different experiments and its estimates are clearly inconsistent: its true standard deviation, which can be appreciated in Fig. 10 *right* to be about 1°, is much larger than its estimated value, which from Fig. 10 *center* is about 0.1° at frame 30.

In order to analyze calibration after convergence we made 10 runs of 200 frames and collected the estimated calibration angles and standard deviations at the end of each sequence. We computed the statistical standard deviations (with respect to the 10 runs) of these estimated angles. We compared these values against the angles provided by the Matlab camera calibration toolbox. Apart from the mentioned initial stability issues, the results in Table II show a surprisingly good calibration, with similar statistical and estimated standard deviations, except for yaw which shows a clear inconsistency, *i.e.* an overestimate of its standard deviation. This inconsistency is further discussed in the conclusions.

## V. CONCLUSION AND FUTURE WORK

We showed in this paper that using mono-vision SLAM techniques in stereo-vision or multi-camera equipped robots provides several advantages. These advantages have been highlighted and explored with a particular bearings-only SLAM algorithm, although they should come up naturally in any other implementation.

The self-calibration solution proposed here suffers from poor observability and inconsistency problems. Theoretically speaking, lack of observability should not be a problem as an image pair of five 3D-points in general configuration renders the whole system observable, but things are in practice much more delicate. Regarding inconsistency, the fact of the different ray members being projected from one camera to the other seems to be the responsible of the observed fall in uncertainty of the yaw angle, because upon observation of a multi-hypothesized ray from the right camera, the FIS update [2] may produce overestimate values in the direction where the ray's members expectations are more disperse, which is precisely the direction that couples the cameras convergence angle (the yaw angle) with the distance to the landmarks. To insure a consistent, real-time, continuous calibration operation, we believe the inverse depth parametrization in [8], [3] should give much more satisfying results. Nevertheless, our procedure helped to prove with real experiments that, given a dynamic extrinsic calibration with its time-varying uncertainty, the 3D observability can be easily determined at every moment from very simple reasoning on the image plane. Of course one can use the whole BiCam proposals with an offline-calibrated stereo rig.

The ultimate objective of this approach is to be able to perform visual-based SLAM in dynamic environments, detecting and tracking the moving objects in the surroundings of the robot (this has already been solved with a range-and-bearing scanner in 2D [12]). The trajectories of these moving objects are not observable by means of bearings-only measurements unless the robot's trajectory is forced appropriately (in an information maximization sense) [13], something that we consider unpractical in multiple real-life situations in constrained environments.

## REFERENCES

[1] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. International Conference on Computer Vision*, Nice, October 2003.

[2] J. Solà, A. Monin, M. Devy, and T. Lemaire, "Undelayed initialization in bearing only SLAM," in *IEEE International Conference on Intelligent Robots and Systems*, Edmonton, Canada, august 2005.

[3] E. Eade and T. Drummond, "Scalable monocular SLAM," *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 469–476, 2006.

[4] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI)*. Acapulco, Mexico: IJCAI, 2003.

[5] T. D. Barfoot, "Online visual motion estimation using FastSLAM with SIFT features," in *IEEE International Conference on Intelligent Robots and Systems*, august 2005.

[6] J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of the EKF-based SLAM," in *5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisboa, PT, July 2004.

[7] V. A. Tupysev, "A generalized approach to the problem of distributed Kalman filtering," in *AIAA Guidance, Navigation and Control Conference*, Boston, 1998.

[8] J. Montiel, J. Civera, and A. Davison, "Unified inverse depth parametrization for monocular SLAM," in *Proceedings of Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[9] C. Harris and M. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, Manchester (UK), 1988.

[10] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[11] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Seattle, USA, 1994, pp. 593–600.

[12] C.-C. Wang, "Simultaneous localization, mapping and moving object tracking," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, April 2004.

[13] J. Le Cadre and C. Jauffret, "Discrete-time observability and estimability analysis for bearings-only target motion analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 1, pp. 178–201, January 1997.