A Dense Map Building Approach from Spherical RGBD Images

Tawsif Gokhool[1], Maxime Meilland[2], Patrick Rives[1] and Eduardo Fernández-Moral[3]
[1]*INRIA, Sophia Antipolis, France*
[2]*CNRS-I3S, University of Nice Sophia Antipolis, Nice, France*
[3]*Mapir Group, Universidad de Málaga, Málaga, Spain*
{*tawsif.gokhool, patrick.rives*}*@inria.fr, meilland@i3s.unice.fr, eduardofernandez@uma.es*

Keywords: Visual Odometry, RGBD Dense Tracking, Spherical/Omnidirectional Vision, Optimisation, Pose Graph.

Abstract: Visual mapping is a required capability for practical autonomous mobile robots where there exists a growing industry with applications ranging from the service to industrial sectors. Prior to map building, Visual Odometry(VO) is an essential step required in the process of pose graph construction. In this work, we first propose to tackle the pose estimation problem by using both photometric and geometric information in a direct RGBD image registration method. Secondly, the mapping problem is tackled with a pose graph representation, whereby, given a database of augmented visual spheres, a travelled trajectory with redundant information is pruned out to a skeletal pose graph. Both methods are evaluated with data acquired with a recently proposed omnidirectional RGBD sensor for indoor environments.

1 INTRODUCTION

This paper is focussed in the context of Visual Odometry (VO), defined as the process of estimating the relative motion of a mobile agent using vision sensors. This incremental technique computes the pose of a vehicle based on the movements induced by onboard cameras. Over the years, VO has been useful to compensate other similar techniques such as wheel odometry which is highly affected by dead reckoning in uneven terrains. On the other hand, global positioning system (GPS) has shown its limitation in aerial and underwater applications (Fraundorfer and Scaramuzza, 2012).

Odometry techniques in general require accurate relative motion estimation to reduce trajectory drift. Over the last decade, VO coupled with SLAM approaches have evolved in two main categories; feature based and dense techniques. Feature based methods rely on a preceeding identification and extraction phase. Registration allows images which are further apart , but are affected by outliers . Dense technique, which uses the entire information content has become increasingly popular recently as registration is performed using a direct image alignment (Meilland and Comport, 2013). The latter is generally more accurate but is restricted by smaller interframe displacements.

An important attribute of a reliable mapping system is scalability. Accumulation of frames over extended trajectories causes the issue of memory space over time. In order to reduce data redundancy, maintaining a sparse pose graph in a skeletal representation is often desirable.Pose Graph pruning is therefore a central issue as discussed in (Konolige and Agrawal, 2008).

This paper extends the results presented in (Meilland et al., 2011) in several ways. We introduce a cost function combining both geometric and photometric constraints and use it in a direct image registration method for pose estimation. Additionally, the problem of pose graph optimization is addressed, whereby, given a database of augmented visual spheres, an explored trajectory with redundant information is pruned out to a skeletal pose graph.

2 RELATED WORKS

2.1 Visual Odometry

After the well established theory of optical flow for motion estimation using Intensity based images, the community has now turned to the fusion of both intensity and depth information for registration and tracking. This trend is recurrent especially due to the advent of consumer based RGB-D sensors such as Microsoft's Kinect or Asus's Xtion Pro. A break-

through of this technique was devised by (Harville et al., 1999), who were among the first to formalize registration as a Brightness Change Constraint Equation (BCCE) and a Depth Change Constraint Equation (DCCE). They argued that tracking is best achieved with intelligent fusion of the two constraints mentioned above in order to reduce the effect of drift, occlusions or illumination changes.

Their work was further extended in (Rahimi et al., 2001), where a Maximum Likelihood (ML) based differential tracker was developed and the problem of drift and loop closure were also addressed. To improve the tracking performance, the measurement model incorporated the fusion of multiple frames. A similar formulation was proposed in (Wang et al., 2006), but encapsulated in a Bayesian framework. A 2D-3D pose estimation along with an intensity cost function helped to improve feature correspondence as well as drift reduction.

Recent works of the same domain includes that of (Newcombe et al., 2011) whereby a preliminary frame to frame registration is fed to a surface reconstruction module which improves the perceived model over time together with the estimated pose. The RGB-D slam framework of (Henry et al., 2012) used a variant of the Iterative Closest Point algorithm (ICP) - direct alignment of 3D point clouds (Besl and McKay, 1992), (Zhang, 1992), together with photometry. Their work also included a surfel representation of the environment to have a more compact representation of the information available. A robust ICP flavour was experimented in (Lui et al., 2012) using an inverse depth parameterisation. Other related works merging photometric and geometric information for Visual Odometry (VO) can be found in (Tykkälä et al., 2011), (Tykkälä et al., 2013), (Kerl et al., 2013).

2.2 Keyframe Selection

When exploring vast scale environments, many frames sharing redundant information clutter the memory space considerably. The idea to select keyframes based on a predefined criteria happens to be very useful in the conception of a sparse skeletal pose graph. Furthermore, performing frame to frame registration introduces drift in the trajectory due to uncertainty in the estimated pose as pointed out in (Kerl et al., 2013).

Therefore, in order to overcome this issue, frame to keyframe odometry is rather desirable. Common techniques applied constitute of introducing keyframes when two such frames share very few features between them as defined by the view clustering criteria of (Konolige and Bowman, 2009), or a threshold on the number of features shared between a central frame and its corresponding adjacent frames (Royer et al., 2007). Strasdat et al. (Strasdat et al., 2010), introduce a new frame whenever a certain distance threshold between camera poses is exceeded. Wang et al. (Wang et al., 2006) modeled a temporal criteria to take into account the interpose frame difference as well as feature overlap among them. On the other hand Meilland et al. (Meilland et al., 2011) used a selection criteria based on the Median Absolute Deviation (MAD) in intensity error between a reference and a current frame to reinitialize on a predefined threshold.

Recently, information theory (Kretzschmar et al., 2010) was introduced to prune out nodes with a minimal expected information gain. On a similar note, (Kim and Eustice, 2013) set up a salient keyframe selection criteria based on the ratio between the covariance of the measurement and that of the innovation to encode the entropy between two corresponding nodes. However, this criteria is modeled based on the probabilistic framework of iSAM where the covariances are easily extracted. On the other hand, we found that the criteria based on a differential entropy approach introduced by (Kerl et al., 2013) is more suitable for our system of geo-referenced spheres which will be discussed in section 4.3.

3 SPHERICAL SYSTEM REPRESENTATION

A new sensor for a large field of view RGBD image acquisition has been used in this paper. This device integrates a set of Asus Xtion Pro Live (Asus XPL) sensors and allows to build a spherical representation as sketched in figure 1.[1]

The chosen configuration offers the advantage of creating full 360° RGBD images of the scene isometrically, i.e. the same solid angle is assigned to each pixel. This permits to apply directly some operations, like point cloud reconstruction, photo consistency alignment or image subsampling. To build the images, the sphere $\mathbb{S}^2$ is sampled according to the resolution of our device, so that the equator (θ direction) contains 1920 samples in the range $[0, 2\pi]$, while the meridian (ϕ direction) is sampled with the same spacing, containing 960 samples in the range $[-\pi/2, \pi/2]$.

For spherical warping, a virtual sphere with the

[1]The conception of this sensor system is out of scope of this paper and details are omitted due to ongoing patent consideration.

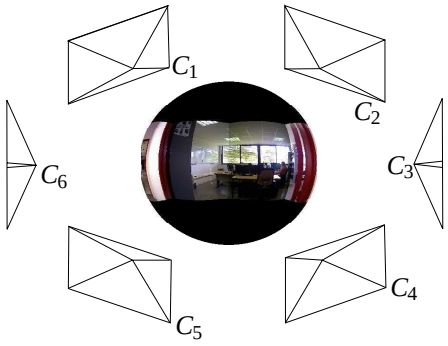

Figure 1: Multi RGBD acquisition system.

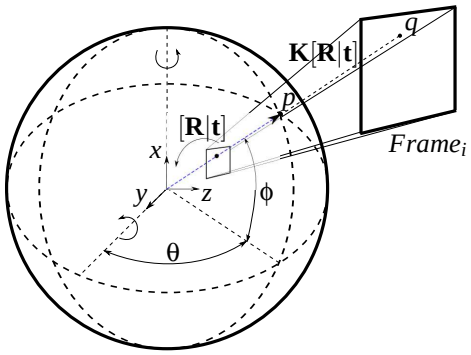

Figure 2: Spherical image construction.

above sampling and radius $\rho = 1$ is used to project the sample points into image coordinates (u,v), (see figure 2). For that, the extrinsic calibration of each sensor is taken into account. Thus, a point p in $\mathbb{S}^2$ is parameterized in $\mathbb{R}^3$ with

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \rho\ cos(\phi)\ sin(\theta) \\ \rho\ sin(\phi) \\ \rho\ cos(\phi)\ cos(\theta) \end{bmatrix} \quad (1)$$

where ρ in this case equals 1 (unit sphere). This point $q = (u,v)$ on image coordinates is found by applying perspective projection to p, through the homogeneous matrix $\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ where $\mathbf{K} \in \mathbb{R}^{3\times3}$ is the camera projection model and $[\mathbf{R}|\mathbf{t}] \in \mathbb{SE}(3)$ is the relative position of the camera (extrinsic calibration). Nearest neighbor interpolation is used to assign the intensity and depth values to the respective spherical images.

The point cloud from the spherical representation (figure 3) is obtained by applying equation (1), substituting ρ by the measured depth and the values of θ and ϕ by its corresponding image location.

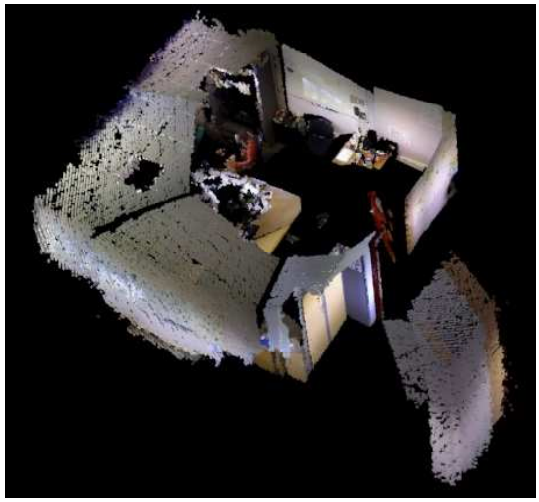

Figure 3: Point cloud obtained from the spherical RGBD acquisition system.

4 SPHERICAL RGBD REGISTRATION

To accurately recover the position of the RGBD spheres with respect to one-another, a 6 dof pose estimation method is proposed based on accurate dense localization. Considering an RGBD sphere $S(I,\rho)$ as defined in Section 3, the objective is now to compute a pose between a reference sphere and the next one. The localisation problem is then to estimate the relative pose $\widehat{\mathbf{T}}$ between the two consecutive spheres.
Inter-frame incremental displacement is defined as an element of the Lie groups applied on the smooth differential manifold of $\mathbb{SE}(3)$ (Blanco, 2010), also known as the group of direct affine isometries. Motion is parametrized as a twist (a velocity screw motion around an axis in space), denoted as $\mathbf{x} = \{[\omega,\upsilon]|\upsilon \in \mathbb{R}^3, \hat{\omega} \in so(3)\} \in \mathfrak{se}(3)$: $\omega = \begin{bmatrix}\omega_x & \omega_y & \omega_z\end{bmatrix}$, $\upsilon = \begin{bmatrix}\upsilon_x & \upsilon_y & \upsilon_z\end{bmatrix}$, with $so(3) = \{\hat{\omega} \in \mathbb{R}^{3\times3} | \hat{\omega} = -\hat{\omega}^T\}$, where ω and υ are the angular and linear velocities respectively . The reconstruction of a group action $\widehat{\mathbf{T}} \in \mathbb{SE}(3)$ from the twist consists of applying the exponential map using Rodriguez formula (Ma et al., 2004). Thereon, $\widehat{\mathbf{T}}$ is denoted as the transformation (pose) recovered between the current frame $I_c \in \mathbb{R}^{m\times n}$ observed at time t and the reference frame $I^* \in \mathbb{R}^{m\times n}$.

4.1 Cost Function Formulation

In (Meilland et al., 2011), the pose estimation was tackled as an image registration problem whose ob-

jective was to perform a direct image alignment by minimizing a cost based on the image intensity only:

$$\mathfrak{F}_I = \frac{1}{2}\sum_i^k \eta_{HUB}\left\| I\left(w(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x});\mathcal{P}_i^*)\right) - I^*\left(w(\mathbf{I};\mathcal{P}_i^*)\right)\right\|^2, \tag{2}$$

where $w(.)$ is the warping function that projects a 3-D point $\mathcal{P}_i$, given the pose T, onto the sphere and η_{HUB} is a robust weighting function on the error given by Huber's M-estimator (Huber, 1981). The latter plays an important role in reducing the effect of outliers by measuring the similarity between two corresponding pixels. Hence the weight computed accommodates partially the uncertainty associated to each pairing between a reference and a current frame.

With the aim of robustifying the above cost function, a geometric point to plane constraint (Chen and Medioni, 1992) is added to the equation (2) as follows:

$$\mathfrak{F}_S = \frac{1}{2}\|e_I\|^2_{D_I} + \frac{\lambda^2}{2}\|e_\rho\|^2_{D_\mathcal{D}}, \tag{3}$$

which can be written in its explicit form:

$$\begin{aligned}\mathfrak{F}_S = \quad & \frac{1}{2}\sum_i^k \eta_{HUB}\left\| I\left(w(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x});\mathcal{P}_i^*)\right) - I^*\left(w(\mathbf{I};\mathcal{P}_i^*)\right)\right\|^2 \\ + \quad & \frac{\lambda^2}{2}\sum_i^k \eta_{HUB}\left\| n^T\left(\mathcal{P}_i - \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x})\mathcal{P}_i^*\right)\right\|^2,\end{aligned} \tag{4}$$

such that $\mathcal{P} \in (X,Y,Z) \longrightarrow (\theta,\phi,\rho)$ and λ is a tuning parameter to effectively balance the two cost functions. n^T is the normal map computed from the cross product of adjacent points on the grid structured depth map.

Since the unknown $\mathbf{x}$ is common in both parts of equation (4), the error function is stacked in a single vector computed simultaneously as shown:

$$\mathbf{e(x)}_S = \begin{bmatrix} \eta_{HUB}\left(I\left(w(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x});\mathcal{P}^*)\right) - I^*(\mathcal{P}^*)\right) \\ \lambda\eta_{HUB}\left(n^T\left(\mathcal{P} - \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x})\mathcal{P}^*\right)\right)\end{bmatrix} \tag{5}$$

The Jacobian matrix $\mathbf{J}_S$ is the total Jacobian relative to the augmented cost function defined above and is given as:

$$\mathbf{J}_S = \begin{bmatrix}\mathbf{J}_{I^*}\mathbf{J}_w\mathbf{J}_T \\ \lambda n^T\mathbf{J}_D\end{bmatrix}, \tag{6}$$

Where, respectively, J_{I^*} is the jacobian w.r.t. the intensity, and $\mathbf{J}_w$ is the jacobian w.r.t. the warping function, $\mathbf{J}_T$ is the jacobian w.r.t. the pose and $\mathbf{J}_D$ is the jacobian w.r.t. the depth.

Similarly, the weighting function for each part of cost function is stacked in a block diagonal matrix where $\mathbf{D}_I, \mathbf{D}_\mathcal{D} \in \mathbb{R}^{mn\times mn}$ are the confidence level in illumination and depth respectively for each corresponding feature pair:

$$\mathbf{D}_S = \begin{bmatrix}\mathbf{D}_I & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\mathcal{D}\end{bmatrix} \tag{7}$$

Linearization of the above cost function leads to a classic closed form solution given by an Iterative Least Mean Squares (ILMS) and the incremental motion $\mathbf{x}$ is given by the following expression:

$$\mathbf{x} = -(\mathbf{J}_S^T\mathbf{D}_S\mathbf{J}_S)^{-1}\mathbf{J}_S^T\mathbf{D}_S\mathbf{e}(\mathbf{x})_S \tag{8}$$

Using an iterative optimization scheme, the estimate is updated at each step by an homogeneous transformation:

$$\widehat{\mathbf{T}} \longleftarrow \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \tag{9}$$

where $\widehat{\mathbf{T}} = [\mathbf{R} \quad \mathbf{t}]$ is the current pose estimate with respect to the reference available from the previous iteration.

This method, requires an initial guess of the pose and it happens that an initial solution does not give an optimum of $\mathbf{x}$ depending on how far the initialization has been made with respect to the reference. To be able to recover a pose close to the solution, equations (5) to (9) are evaluated iteratively until a tolerated threshold in $\|\mathbf{x}\| < \varepsilon$ is reached. This leads us to the term Iterative Re-weighted Least Squares (IRLS) where the Gauss-Newton algorithm is well adapted for these genres of non linear unconstrained optimization problem.

4.2 Information Selection

In order to estimate displacement between two frames, a set of correspondences between them has to be found to constrain the motion model (R,t) efficiently. This is a vital step in visual odometry as bad feature matches lead to pronounced deviation from the real motion. In literature, two mainstreams are identified: the first one based on feature extraction while the second one uses dense (correspondence-free) methods (Fraundorfer and Scaramuzza, 2012).

Both approaches exhibit their advantages and inconveniences. The former, based on point feature detection needs to undergo an identification phase where primitives such as blobs, corners, lines or edges are usual candidates. Good features are characterized in terms of several properties such as stability, computational efficiency, distinctiveness or invariance to geometric and photometric changes.

On the other hand, dense methods make use of the entire photometric and geometric information content for tracking. Along that streamline, (Dellaert and Collins, 1999) argued that instead of using all the information content which is computation intensive, selection of a subset of good pixels that yield enough information about the 6 degrees of freedom (DOF) state vector could considerably reduce computational cost as well as data redundancy without compromising on the accuracy of the estimation.

Recently, a similar approach was independently devised in (Meilland et al., 2010) which relies on finding the most informative pixel subset based on the decomposition of the Jacobian matrix obtained from equation (2). We hereby outline the backbone of the algorithm which has been implemented in the system and is also illustrated in figure 4:

- Decomposition of the Jacobian **J** gives in itself a saliency map pertaining to each DOF of **x**
- Columnwise, **J** is sorted out for the best pixel which is indexed in decreasing order of magnitude
- Each DOF is then looked-up for the best ranked pixel i in ascending order and the i^{th} line is lifted to a new table as shown in figure 4
- In case a particular pixel gives the best indexing in more than one DOF and given that it has already been selected, we proceed to the second best pixel and so on.
- The selection process is performed iteratively until all the pixels have been successfully sorted out
- The final result is a table of best ranked pixels with respect to their response to a particular DOF

Therefore, instead of using all the pixels from the intensity and depth map, a wise selection of the top 10-20% of the pixels are used for registration.

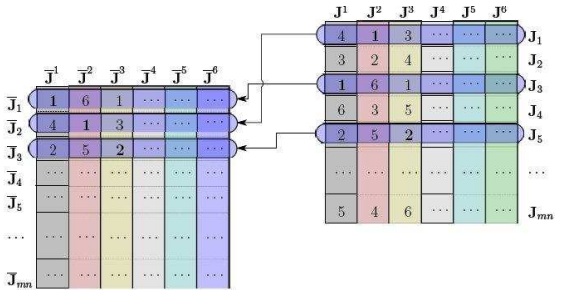

Figure 4: Saliency map.

4.3 Salient Frames Criterion

We define the differential entropy of a random variable **x** with dimensions n such that $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$ as:

$$H(\mathbf{x}) = \frac{n}{2}\big(1 + \ln(2\pi)\big) + \frac{1}{2}\ln(|\Sigma|), \qquad (10)$$

where, Σ is the covariance matrix of the estimate **x** which is obtained by the inverse of the Fisher Information matrix computed from the normal equations (8):

$$\Sigma = \big(\mathbf{J}_S^T \mathbf{D}_S \mathbf{J}_S\big)^{-1} \qquad (11)$$

The entropy ratio between a motion estimate $x_{k:k+j}$ from a reference frame k to a current frame $k+j$ is obtained by the following deduction:

$$\alpha = \frac{H(\mathbf{x}_{k:k+j})}{H(\mathbf{x}_{k:k+1})}, \qquad (12)$$

where the denominator is just the entropy relative to the consecutive of the k^{th} frame in question. The greater the gap between the reference and the current frame, the smaller the value of α. Hence a preset on the value of α is used to decide whenever a keyframe needs to be inserted or not.

5 RESULTS

5.1 Simulated Data

Our first fold of experiments have been performed on a synthetic dataset available along with the provided ground truth. Our algorithms have been thoroughly tested on this dataset of around 600 images in order

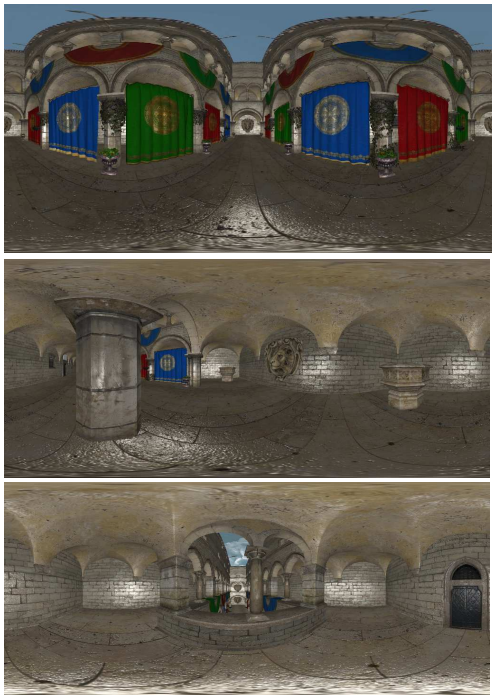

Figure 5: Example of spherical views obtained from the *Synthetic* dataset.

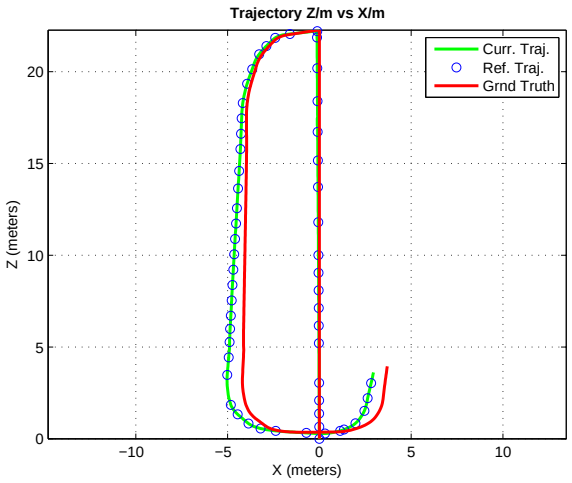

Figure 6: Trajectory comparison between ground truth and that of the Intensity with Geometric cost function with the keyframe criteria based on the differential entropy approach

to validate the convergence of the various cost functions discussed in the paper before moving on to real data. Figure 6 shows the trajectory plot obtained from the augmented cost function (4). As illustrated, we observe a small but noticeable offset between ground truth and the computed trajectory using the proposed approach. This is accountable for the fact that VO is prone to drift phenomena.

5.2 Real Data

Our experimentations are performed on a dataset of around 170 intensity and depth images within an office environment constituted of several rooms and corridors. Figure 7 illustrates the different types of places observed along the trajectory.

Figure 7: Example of spherical views from the *Office* dataset.

Figure 8(a) shows the trajectory obtained using the cost functions of equations (2) and (4) respectively. We observe that the algorithm using vision-only performs poorly in low-textured regions such as corridors or in the presence of reflections from window panes. Such circumstances lead to erroneous estimated poses coming from poor convergence of the algorithm. On the other hand, the *photometry + geometry* cost function takes care of these discrepancies by relying more on the depth information available. This is justified by the overall faster convergence of the algorithm as profiled in figure 8(b). The high spikes of the figure capped at 200 iterations are due to the non-convergence of the intensity cost function while the new approach still manage to converge at lower iterations. Finally, figures 8(c) and 8(d), depicts the error norm of each frame at convergence.

Figure 9 focusses on the keyframe criteria discussed in section 4.3 for the same dataset: MAD (*method 1*) and Entropy ratio (*method 2*). Over here, we fix the *photometry + geometry* approach with the keyframe criteria as the only variants. While the MAD acts on the residual warping error after convergence, the entropy ratio α abstracts the uncertainty in the estimated pose along the trajectory. The number of spheres initialized for *method 1* is around 50 while *method 2* revealed 27 re-initializations. However, we believe that greater reduction is achievable with lesser inter frame acquisition so that the pose estimation algorithm is better initialized leading to a faster and more accurate convergence. A heuristic threshold of 15 was chosen for the case of the MAD and 0.96 for that of α.

6 CONCLUSIONS

In this paper, a spherical RGBD registration method has been proposed and we have demonstrated that both photometry and geometry can co-exist within the same framework. The idea is to implement a robust registration process that compensates mutually for poorly conditioned photometric images as well as depth maps. Furthermore, we have compared two different criteria for keyframe selection and have shown an improvement by using a criterion based on differential entropy. Results have been presented with synthetic and real data. For the latter, a novel sensor was used to acquire an indoor dataset at the INRIA Sophia Antipolis offices.

Our future works will be directed towards the exploitation of our new spherical mapping system on board of a mobile robot platform for simultaneous localisation and mapping (SLAM) as well as autonomous navigation.

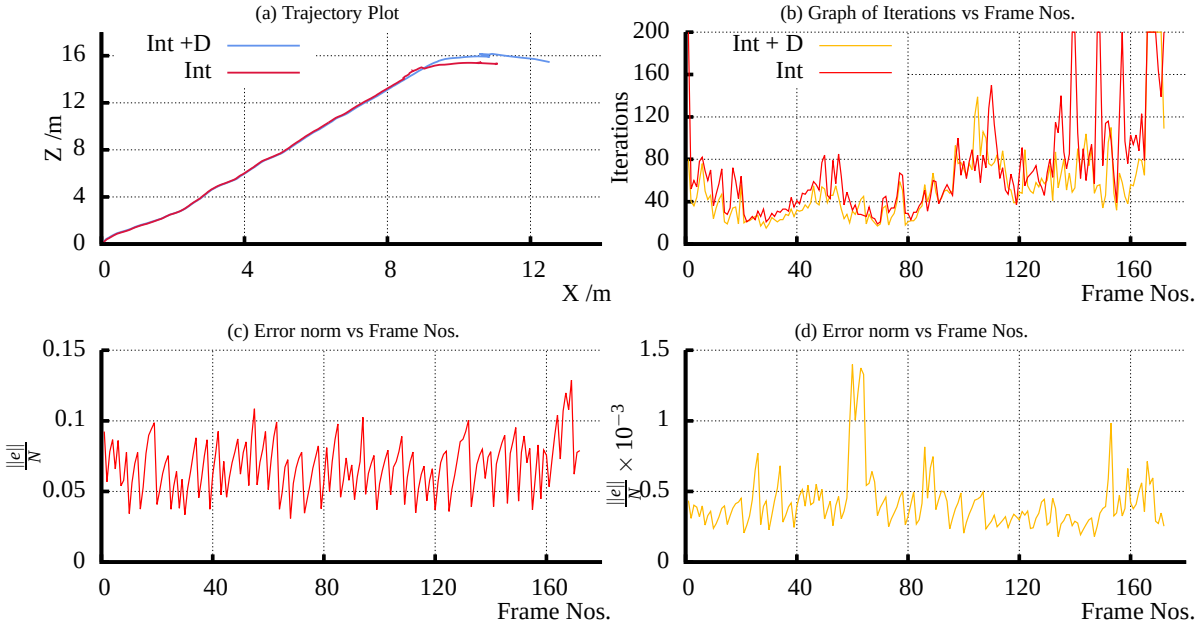

Figure 8: Performance comparison between Intensity only and Intensity and Depth cost functions.

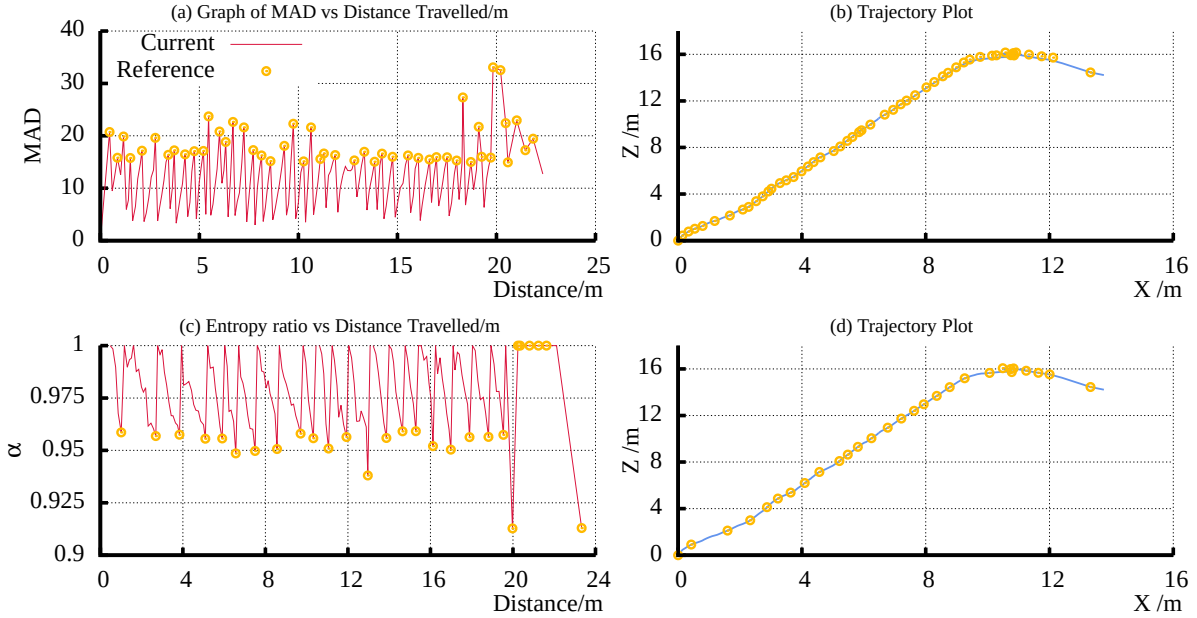

Figure 9: Comparison between MAD vs entropy ratio α using the same augmented *photometry + geometry* cost function.

ACKNOWLEDGEMENTS

This work has been developed at INRIA Sophia-Antipolis and extends a previous collaboration with the I3S/CNRS team. It is funded by EADS/Astrium under the contract N. 7128 and by the "European Regional Development Fund ERDF" through the Spanish Government under the contract DPI2011-25483.

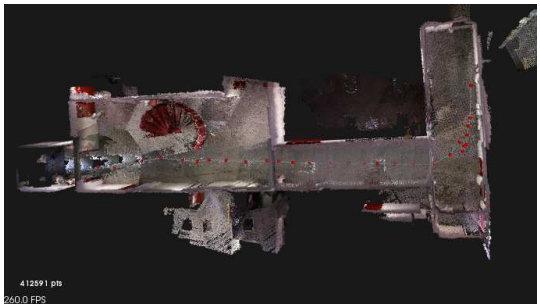

Figure 10: Top view of real trajectory with fusion of intensity and depth with entropy ratio α.

REFERENCES

Besl, P. and McKay, N. (1992). A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI*, 14(2):239–256.

Blanco, J.-L. (2010). A tutorial on se(3) transformation parameterizations and on-manifold optimization. Technical report, University of Malaga.

Chen, Y. and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155.

Dellaert, F. and Collins, R. (1999). Fast Image-based tracking by selective pixel integration. In *Proc. of the ICCV Workshop on frame-rate vision*.

Fraundorfer, F. and Scaramuzza, D. (2012). Visual Odometry: PartII: Matching, Robustness, Optimisation and Applications. *IEEE Robotics and Automation magazine*, 19(2):78–90.

Harville, M., Rahimi, A., Darrell, T., Gordon, G., and Woodfill, J. (1999). 3D Pose Tracking with Linear Depth and Brightness Constraints. In *IEEE Intl. Conf. on Computer Vision, ICCV*, volume 1, pages 206–213.

Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D. (2012). RGB-D Mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments . *International Journal of Robotics Research*, 31(5):647–663.

Huber, P. (1981). *Robust Statistics*. New York, Wiley.

Kerl, C., Sturm, J., and Cremers, D. (2013). Dense Visual SLAM for RGB-D Cameras. In *Proc. of the Int. Conf. on Intelligent Robot Systems (IROS)*, Tokyo, Japan.

Kim, A. and Eustice, R. (2013). Real-time visual SLAM for autonomous underwater hull inspection using visual saliency. *IEEE Transactions on Robotics*, 29(3):719–733.

Konolige, K. and Agrawal, M. (2008). FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping. *IEEE Transaction on Robotics*, 24(5).

Konolige, K. and Bowman, J. (2009). Towards lifelong Visual Maps. In *International Conference on Intelligent Robots and Systems*.

Kretzschmar, H., Grisetti, G., and Stachniss, C. (2010). Lifelong map learning for graph-based slam in static environments. *Künstliche Intelligenz, KI*, 24(3):199–206.

Lui, W., Tang, T., Drummond, T., and Li, W. (2012). Robust Egomotion Estimation using ICP in Inverse Depth Coordinates. In *IEEE Intl. Conf. on Intelligent Robots and Systems, IROS*, Villamoura, Portugal.

Ma, Y., Soatto, S., Košecká, J., and Sastry, S. S. (2004). *An Invitation to 3-D Vision*. Springer.

Meilland, M. and Comport, A. (2013). Simultaneous superresolution tracking and mapping. In *IEEE Intl. Conf. on Robotics and Automation, ICRA*, Karslruhe, Germany.

Meilland, M., Comport, A., and Rives, P. (2010). A spherical robot-centered representation for urban navigation. In *Proc. of the Int. Conf. on Intelligent Robot Systems (IROS)*, Taipei, Taiwan.

Meilland, M., Comport, A. I., and Rives, P. (2011). Dense visual mapping of large scale environments for real-time localisation. In *IEEE International Conference on Intelligent Robots and Systems*, pages 4242 –4248.

Newcombe, R., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A., Kohli, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *IEEE Intl. Symp. on Mixed and Augmented Reality, ISMAR*.

Rahimi, A., Morency, L., and Darrell, T. (2001). Reducing Drift in Parametric Motion Tracking. In *IEEE Intl. Conf. on Computer Vision, ICCV*.

Royer, E., Lhuillier, M., M., D., and J-M., L. (2007). Monocular vision for mobile robot localization and autonomous navigation. *International Journal of Computer Vision*, 74(3):237 – 260.

Strasdat, H., Montiel, J., and Davison, A. (2010). Scale Drift-Aware Large Scale Monocular SLAM. In *Robots Science and Systems, RSS*.

Tykkälä, T., Audras, C., and Comport, A. (2011). Direct Iterative Closest Point for Real-time Visual Odometry. In *IEEE Intl. Conf. on Computer Vision Workshops, ICCV*.

Tykkälä, T., Hartikainen, H., Comport, A., and Kämäräinen, J.-K. (2013). RGB-D tracking and reconstruction for TV broadcasts. In *Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, Barcelona, Spain.

Wang, Q., Zhang, W., and Tang, X. (2006). Real Time Bayesian 3-D Pose Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(12):1533–1541.

Zhang, Z. (1992). On local matching of free-form curves. In *Proc. of British Machine Vision Conference, BMVC*, pages 347–356.