An asymmetric real-time dense visual localisation and mapping system

Andrew I. Comport
Centre National de Recherche Scientifique (CNRS),
I3S - University of Nice Sophia Antipolis

Maxime Meilland, Patrick Rives
INRIA Sophia Antipolis, ARobAS

LDRMC, November 12, 2011

Introduction

Dense omnidirectional localisation and mapping

- A dense direct visual 3D SLAM approach [1, 2].
- Objective to densely map large scale environments.
- Complex scene geometry, uncertainty and occlusions.
- Fast real-time computation.
- A model for dynamic environments.

Solution:

- Ego-centric maps: RGB-D spherical panorama graph.
- Hybrid Model-Based and Visual-Odometry.
- Real-time asymmetric monocular camera localisation.

[1] A.I Comport, E. Malis & P. Rives, Accurate Quadrifocal Tracking for Robust 3D Visual Odometry, ICRA 07.

[2] M. Meilland, A.I Comport & P. Rives, A Spherical Robot-Centered Representation for Urban Navigation, IROS 10.

Summary

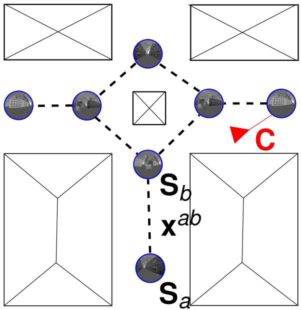
1. Scene Map Representation.
2. Acquisition system.
3. Automatic Dense Mapping.
4. Real-time localisation.
5. Conclusion.

Spherical ego-centered representation

Global Representation: Graph

$\mathcal{G} = \{\mathbf{S}_1, \ldots, \mathbf{S}_n; \mathbf{x}_1, \ldots, \mathbf{x}_m\}$,
$\mathbf{x}_n \in \mathbb{R}^6$: 6 d.o.f. twist between each sphere.

- Learning phase - high computation and low rate.
- A set of augmented spherical images sampled along a trajectory.
- Edges: $\mathbf{x}^{ab}$
- Nodes: $\mathbf{S}_{1\ldots n}$



Spherical ego-centered representation

Local representation: Augmented sphere

$$\mathbf{S} = \{\mathcal{I}_s, \mathcal{P}_{\boldsymbol{s}}, \mathbf{Z}_s, \mathbf{W}_s\}$$

Description

- $\mathcal{P}_{\boldsymbol{s}}$: Unit sphere sampling.
- $\mathcal{I}_s$: Photometric spherical image.
- $\mathbf{Z}_s$: Depth-map.
- $\mathbf{W}_s$: Saliency image (pixel ordering).

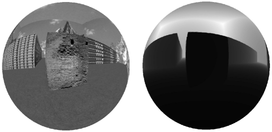


Figure: A Spherical image and its associated depth map.

Spherical imaging sensors

Omnidirectional camera [3]

- Poor and non uniform spatial resolution.
- Limited vertical FOV.

Image stitching [4]

- Multiple perspective cameras.
- High resolution panoramas.
- Unique center of projection **approximation**.
- Parallax artefacts.

Depth information **cannot** be extracted in a single frame.

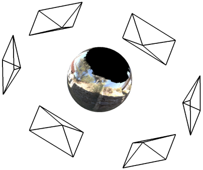
[3] S.K Nayar, Catadioptric omnidirectional camera, CVPR 97.
[4] R. Szeliski, Image alignment and stitching: a tutorial, 06

Multi-camera system

Multi-camera system with baselines

- 6 wide angle (125°) **stereo** cameras with wide baselines (65 cm).
- 360° of overlap.
- Stereo dense matching [5] for depth extraction.

Depth information **can** be extracted in a single frame - constrains general 6dof motion estimation.



[5] H. Hirschmuller, Stereo processing by semi-global matching and mutual information, PAMI 08.

Sphere positioning

Accurate dense visual odometry [6], **direct** minimisation of intensity errors:

Multi-camera robust dense minimisation

$$\mathbf{e} = \sum_{i=1\ldots6} \rho\left(\mathcal{I}_i\left(w\left(\mathbf{T}(\mathbf{x}_i^c)\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}); \mathcal{P}_s^*, \mathbf{Z}_s\right)\right) - \mathcal{I}_s(\mathcal{P}_s^*, \mathbf{Z}_s)\right),$$

- $\mathcal{I}_s$: reference sphere intensities.
- $\mathcal{I}_i$: perspective images,
- $w(.)$ warping function
- $\widehat{T} \in \mathbb{SE}(3)$: initial motion estimation.
- $\mathbf{x} \in \mathbb{R}^6$: unknown 3D motion.
- $\rho(.)$: robust outlier rejection from M-estimation.

[6] A.I. Comport, E. Malis, and P. Rives, Real-time quadrifocal visual odometry, IJRR 2010.

New sphere selection

Accurate robust dense multi-camera visual odometry [6] :

Direct minimisation of intensity errors

$$\mathbf{e} = \sum_{i=1 \ldots 6} \rho\left(\mathcal{I}_i\left(w\left(\mathbf{T}(\mathbf{x}_i^c)\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}); \mathcal{P}_s^*, \mathbf{Z}_s\right)\right) - \mathcal{I}_s(\mathcal{P}_s^*, \mathbf{Z}_s)\right),$$

Improve depth maps over time

- Maintain as long as possible the reference sphere.
- Integrate the dense matching incrementally in time [7],

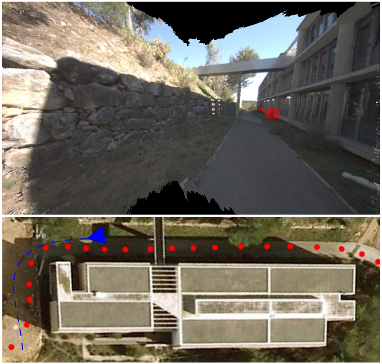
Robust update criteria: Median absolute deviation

$$\lambda < Median(\mathbf{e} - Median(\mathbf{e})).$$

[7] Tykkala, T.M. and Comport, A.I, A Dense Structure Model for Image Based Stereo SLAM, ICRA 11

Immersive navigation

- Virtual navigation using the spherical graph.
- Photo-realistic rendering.



Direct 3D Model-based (MB) tracking

The unknown 3D motion $\mathbf{x}$ between an augmented reference image $\boldsymbol{S} = \{\mathcal{I}^*, \mathcal{P}^*\}$ and the current camera $\mathcal{I}_t$ can be iteratively estimated by minimising a robust error between the warped image and the reference image:

$$\mathbf{e}_{MB} = \rho\left(\mathcal{I}_t\left(w(\mathcal{P}^*; \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}))\right) - \beta_{MB} - \mathcal{I}^*(\mathcal{P}^*)\right).$$

Dynamic environments - illumination change

- *Global* illumination - $\beta_{MB} = Median(\mathbf{e}_{MB})$ [8].
- *Local* illumination - robust diagonal weighting matrix $\mathbf{D}$ [9].

Inconvenient: Large environment changes

[8] Gonçalves, T. & Comport, A.I, Real-time Direct Tracking of Color Images in the Presence of Illumination Variation, ICRA 11.
[9]P.J. Huber, Robust Statistics, 1981.

Visual odometry(VO) tracking

A non-classic visual odometry approach - improves convergence speed and robustness to dynamic changes (current and previous image intensities are minimised with model geometry).

$$\mathbf{e}_{VO} = \rho\left(\mathcal{I}_t(w(\mathcal{P}^*; \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}))) - \beta_{VO} - \mathcal{I}^w_{t-1}(\mathcal{P}^*)\right),$$

Advantages

- 3D geometry is shared with the original 3D model.
- Very small local illumination changes can be expected between successive frames (ie. $\geq$ 20Hz) $\Rightarrow$ fast convergence.
- Still robust to global illumination changes (due to the global bias model).

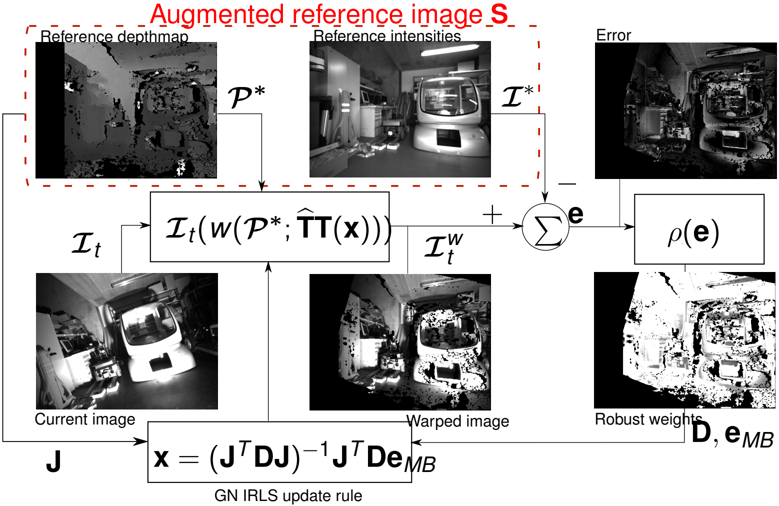
however **visual odometry drifts** with time.

Hybrid model-based and visual odometry (H) tracking

Proposed method: Global minimisation of the error functions, combines the advantages of both techniques.

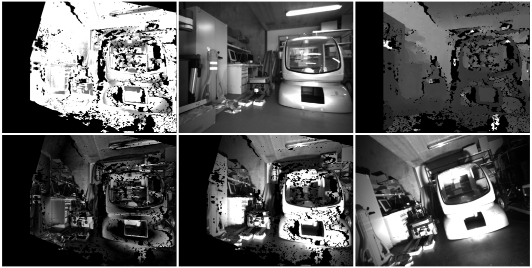
$$\mathbf{e}_H = \begin{bmatrix} \mathbf{e}_{MB} & \mathbf{e}_{VO} \end{bmatrix}^T .$$

- **Fast** convergence (due to VO).
- **No drift** since raw sensor measurement is maintained in the minimisation process (due to MB).


Augmented reference image S
Reference depthmap
Reference intensities
Error
$\mathcal{P}^*$
$\mathcal{I}^*$
$-$
$+$
e
$\mathcal{I}_t(w(\mathcal{P}^*; \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x})))$
$\rho(\mathbf{e})$
$\mathcal{I}_t$
$\mathcal{I}_t^w$
Current image
Warped image
Robust weights
$\mathbf{D}, \mathbf{e}_{MB}$
J
$\mathbf{x} = (\mathbf{J}^T\mathbf{D}\mathbf{J})^{-1}\mathbf{J}^T\mathbf{D}\mathbf{e}_{MB}$
GN IRLS update rule

Tracking in dynamic environments

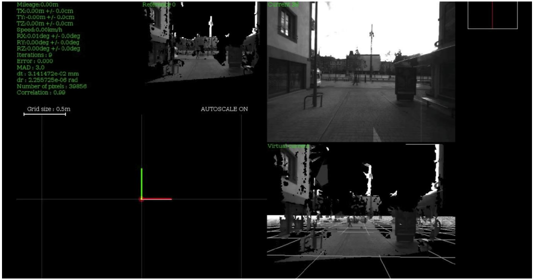
- Local reflections and illumination variation.
- Global illumination change.
- Intensity saturation.
- Varying focal length.



- Kinect based localisation and mapping at 30Hz [10].
- Direct Iterative Closest Point [11] (tomorrow).

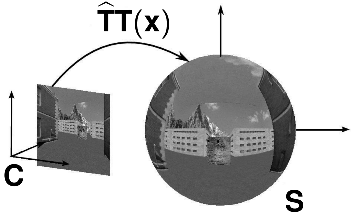
[10] C. Audras, A.I Comport, M. Meilland, and P. Rives. Real-time dense RGB-D localisation and mapping. ACRA 11
[11] Tykkälä, T.M. and Comport, A.I. Direct Iterative Closest Point for Real-time Visual Odometry CVVT-ICCV 11

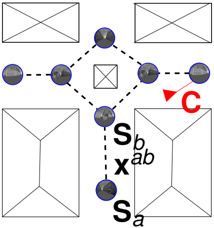
- Real-world complex environment with pedestrians, cars, trams, and illumination change.
- Dense and fast model acquisition.



Online localisation

Real time localization and navigation of a vehicle using a **single** camera.





Monocular direct 3D image registration

$$\mathbf{e} = \rho\left(\mathcal{I}_t\left(w\left(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}); \mathcal{P}_s^*, \mathbf{Z}_s, \mathbf{W}_s\right)\right) - \mathcal{I}_s(\mathcal{P}_s^*, \mathbf{Z}_s, \mathbf{W}_s)\right).$$

Real-time localisation results

- 300 augmented spheres extracted from the graph.
- Real-time **monocular** camera (robot) localisation at 45 Hz.

Conclusion and Future works

Conclusion

A dense large scale scene representation for asymmetric real-time localisation.

- An ego-centric omnidirectional graph approach for large scale mapping.
- Fast and robust real-time asymmetric localisation.
- A hybrid approach for dynamic environments.
- Dense autonomous navigation.

Future works

- Studies on optimising and extending the representation for long term mapping.

Open postdoc position

- A new postdoc position is open for autonomous navigation from dense maps - details to follow shortly at: http://www.i3s.unice.fr/ comport
- Please send an email if you are interested: comport@i3s.unice.fr

References I

- A. Comport, E. Malis, and P. Rives, "Accurate quadrifocal tracking for robust 3d visual odometry," in *IEEE Conference on Robotics and Automation*, April 2007, pp. 40–45.
- M. Meilland, A. Comport, and P. Rives, "A spherical robot-centered representation for urban navigation," in *IEEE Int. Conf. on Intelligent Robots and Systems*, 2010, pp. 5196 –5201.
- S. Nayar, "Catadioptric omnidirectional camera," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 482–. [Online]. Available: http://portal.acm.org/citation.cfm?id=794189.794460

References II

- R. Szeliski, “Image alignment and stitching: a tutorial,” *Found. Trends. Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2006.
- H. Hirschmuller, “Stereo processing by semiglobal matching and mutual information,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 328–341, 2008.
- A. Comport, E. Malis, and P. Rives, “Real-time quadrifocal visual odometry,” *In The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 245–266, 2010.

References III

- T. Tykkala and A. Comport, “A dense structure model for image based stereo slam,” in *IEEE International Conference on Robotics and Automation, ICRA’11*, Shanghai, China, May 9-13 2011. [Online]. Available: http://www.i3s.unice.fr/~comport/publications/2011_ICRA_Tykkala.pdf
- T. Gonçalves and A. Comport, “Real-time direct tracking of color images in the presence of illumination variation,” in *IEEE International Conference on Robotics and Automation*, May 2011.
- P. Huber, *Robust Statistics*. New york, Wiley, 1981.

References IV

C. Audras, A. Comport, M. Meilland, and P. Rives, "Real-time dense rgb-d localisation and mapping," in *To Appear Australian Conference on Robotics and Automation*, Monash University, Australia, December 7-9 2011. [Online]. Available: http://www.i3s.unice.fr/~comport/publications/2011_ACRA_Comport.pdf

T. Tykkälä and A. Comport, "Direct iterative closest point for real-time visual odometry," in *The Second international Workshop on Computer Vision in Vehicle Technology: From Earth to Mars in conjunction with the International Conference on Computer Vision*, Barcelona, Spain, November 6-13 2011. [Online]. Available: http://www.i3s.unice.fr/~comport/publications/2011_CCVT_Tykkala.pdf