Monocular Visual-Inertial-Pressure SLAM for Underwater Localization and 3D Mapping

Soutenance de thèse

Maxime Ferrera

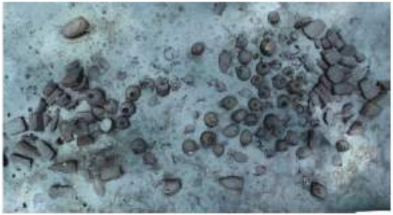
Encadrants : Julien Moras, Pauline Trouvé-Peloux (DTIS - IVA)
Directeur de thèse : Vincent Creuze (LIRMM - Université de Montpellier)

12 Décembre 2019

Introduction

Underwater Archaeology

- Many sites below 100 meters deep
- Not human-friendly environnements

Credit : DRASSM

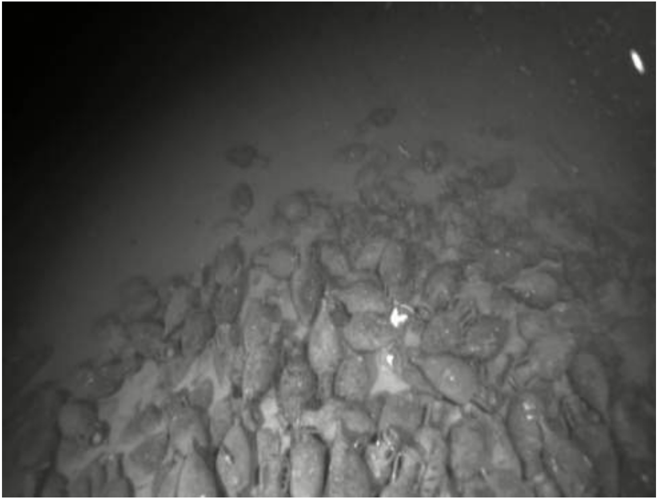
Introduction

Robots to the rescue

- ROV : Remotely Operated Vehicles
- ROVs are used for deep surveys

Introduction

Manual navigation is hard!



Credit : DRASSM

Introduction

Accurate localization in real-time is highly beneficial

- Assistance for efficient and safe piloting
- Autonomous navigation
- 3D reconstruction

Introduction

Accurate localization in real-time is highly beneficial

- Assistance for efficient and safe piloting
- Autonomous navigation
- 3D reconstruction

Underwater localization is tough

- GNSS-denied
- No easy access
- Requires 3D localization $\rightarrow$ 3D Orientation + 3D Position (Pose)

Context

ROVs for underwater archaeology

- Small / Lightweight ROVs
- Cost constraints

Context

ROVs for underwater archaeology

- Small / Lightweight ROVs
- Cost constraints

Existing technologies

- Accoustic sensors : USBL / SBL / LBL, Doppler Velocity Logs, Sonars
- Inertial Navigation Systems : high-end gyroscopes and accelerometers

Context

ROVs for underwater archaeology

- Small / Lightweight ROVs
- Cost constraints

Existing technologies

- Accoustic sensors : USBL / SBL / LBL, Doppler Velocity Logs, Sonars
- Inertial Navigation Systems : high-end gyroscopes and accelerometers

- Bulky and expensive

Localization from Vision

Localization from visual sensors

- Very popular in aerial / land robotics and AR / VR
- Cameras are cheap, lightweight and very informative

Localization from Vision

Localization from visual sensors

- Very popular in aerial / land robotics and AR / VR
- Cameras are cheap, lightweight and very informative

Monocular Visual SLAM

- SLAM : Simultaneous Localization And Mapping

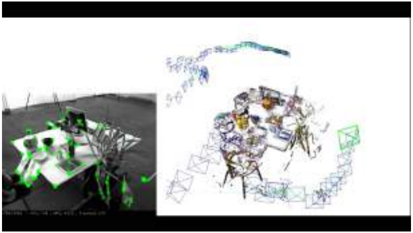
Localization from Vision

Localization from visual sensors

- Very popular in aerial / land robotics and AR / VR
- Cameras are cheap, lightweight and very informative

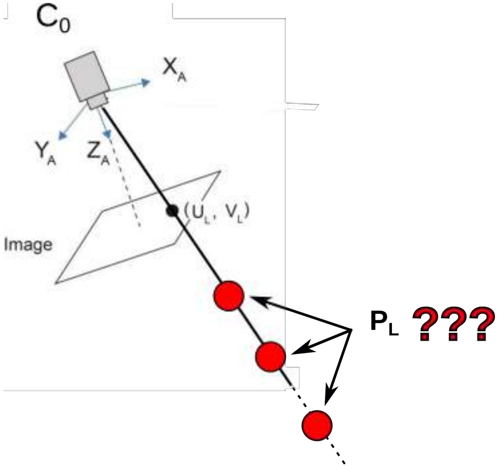
Monocular Visual SLAM

- SLAM : Simultaneous Localization And Mapping
- Visual SLAM : Use pixel correspondences between images



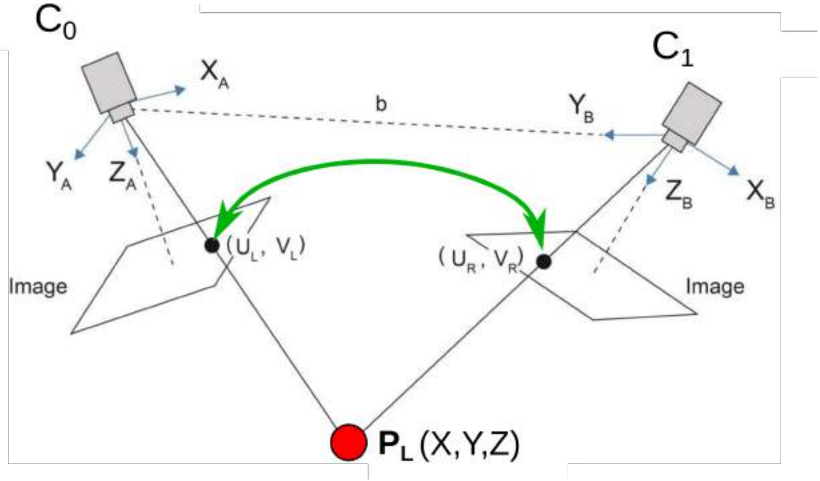
Localization from Vision

Single Image



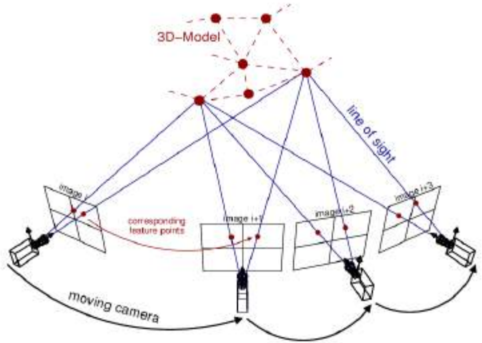
Localization from Vision

Multi-view



Localization from Vision

SLAM by Structure-from-Motion

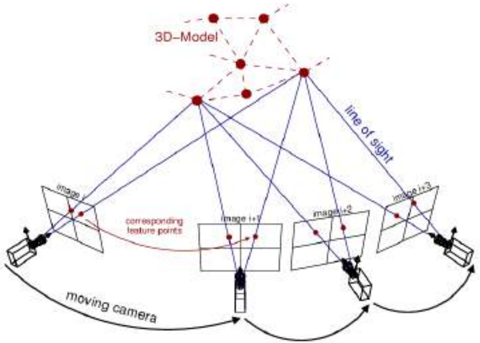


From pixel correspondences :

- Localization $\rightarrow$ 3D map
- 3D map $\rightarrow$ Localization

Localization from Vision

SLAM by Structure-from-Motion



From pixel correspondences :

- Localization $\rightarrow$ 3D map
- 3D map $\rightarrow$ Localization

▶ **Good features tracking is critical!**

State-of-the-art

Monocular Visual SLAM

- PTAM (Klein et al., 2007) : Use of keyframes for efficient optimization

State-of-the-art

Monocular Visual SLAM

- PTAM (Klein et al., 2007) : Use of keyframes for efficient optimization
- ORB-SLAM (Mur-Artal et al., 2015) : Use of descriptors for loop closure

State-of-the-art

Monocular Visual SLAM

- PTAM (Klein et al., 2007) : Use of keyframes for efficient optimization
- ORB-SLAM (Mur-Artal et al., 2015) : Use of descriptors for loop closure
- LSD-SLAM (Engel et al., 2014) / SVO (Forster et al., 2014) / DSO (Engel et al., 2017) : Joint tracking and pose estimation from the minimization of a photometric cost

State-of-the-art

Monocular Visual SLAM

- PTAM (KLEIN et al., 2007) : Use of keyframes for efficient optimization
- ORB-SLAM (MUR-ARTAL et al., 2015) : Use of descriptors for loop closure
- LSD-SLAM (ENGEL et al., 2014) / SVO (FORSTER et al., 2014) / DSO (ENGEL et al., 2017) : Joint tracking and pose estimation from the minimization of a photometric cost

▶ **Not designed for underwater environments**

State-of-the-art

Underwater Monocular Localization

- Use of a camera as a complementary sensor for loop detections (KIM et al., 2013)
- Visual Mosaicking (GARCIA et al., 2001; NICOSEVICI et al., 2009; SINGH et al., 2004)
- EKF based Visual SLAM (BURGUERA et al., 2015)
- Fusion with IMU and pressure sensor : EKF-based (SHKURTI et al., 2011), incremental positioning (CREUZE, 2017)

State-of-the-art

Underwater Monocular Localization

- Use of a camera as a complementary sensor for loop detections (Kim et al., 2013)
- Visual Mosaicking (Garcia et al., 2001; Nicosevici et al., 2009; Singh et al., 2004)
- EKF based Visual SLAM (Burguera et al., 2015)
- Fusion with IMU and pressure sensor : EKF-based (Shkurti et al., 2011), incremental positioning (Creuze, 2017)

▶ **Few works on keyframe-based 3D SLAM for underwater environments**

Thesis proposal

SLAM from a monocular vision-based system

- Convenient : double use of the ROV's camera
- Small size
- Low-cost
- 3D Reconstruction capability

Thesis proposal

SLAM from a monocular vision-based system

- Convenient : double use of the ROV's camera
- Small size
- Low-cost
- 3D Reconstruction capability

Monocular only

- No metric scale
- Fails if no visual information

Set of sensors

Camera

- Monochromatic Camera
- 20 Hz

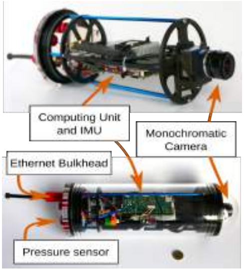
MEMS-IMU

- Inertial meas.
- 200 Hz
- High drift

Pressure Sensor

- Pressure meas. (bar) $\propto$ Depth (m)
- 5-10 Hz

Designed Systems



Size : 33.4 $\times$ 11.4 cm
Depth rated : 100 m

Size : 25.8 $\times$ 8.9 cm
Depth rated : 500 m

- **On-board computation**
  - $\rightarrow$ Autonomous and independent
  - $\rightarrow$ No bandwidth issue

- **Compact**
- **Low-cost :** $<$ 2.5 k€

Dataset

AQUALOC Dataset : `http://www.lirmm.fr/aqualoc/`

FIGURE – ROV Dumbo (DRASSM / LIRMM)

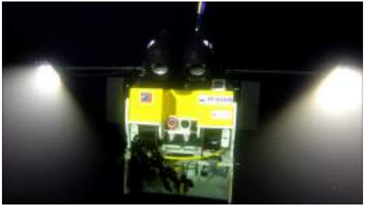
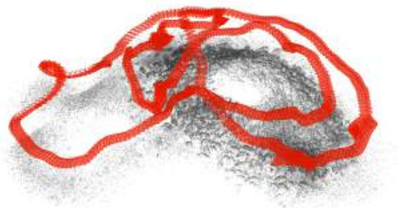


FIGURE – ROV Perseo (Copetech SM - Credit : DRASSM / F. Osada)

- 17 sequences
- Synchronized measurements
- Harbor & Archaeological sites
- Comparative baselines from offline photogrammetry



Problem Statement

1. **Underwater Features Tracking**
2. **Robust Underwater Monocular Visual SLAM**
3. **Multi-Sensors SLAM**
4. **Monocular Dense 3D Mapping**

Problem Statement

1. Underwater Features Tracking
2. **Robust Underwater Monocular Visual SLAM**
3. Multi-Sensors SLAM
4. Monocular Dense 3D Mapping

Problem Statement

1. **Underwater Features Tracking**
2. **Robust Underwater Monocular Visual SLAM**
3. **Multi-Sensors SLAM**
4. **Monocular Dense 3D Mapping**

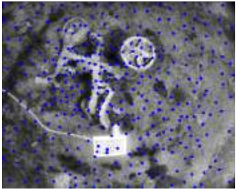
Problem Statement

1. Underwater Features Tracking
2. Robust Underwater Monocular Visual SLAM
3. Multi-Sensors SLAM
4. **Monocular Dense 3D Mapping**

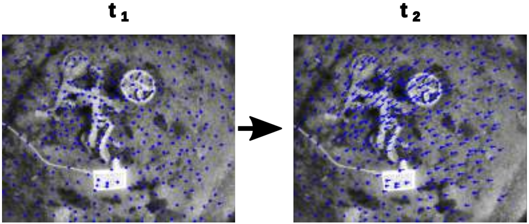
1. Underwater Features Tracking

Underwater Features Tracking

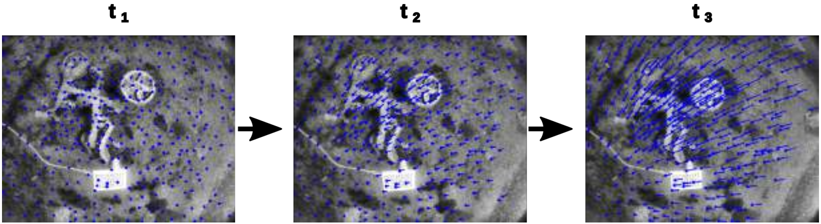
t_1



Underwater Features Tracking



Underwater Features Tracking



Underwater Features Tracking

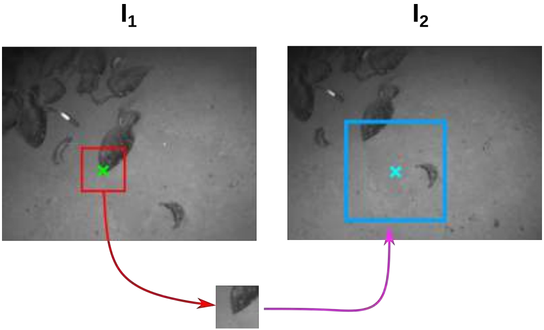
Challenging Imaging Conditions

Credit : DRASSM

Underwater Features Tracking

Direct methods

- Tracking by searching for photometric minima

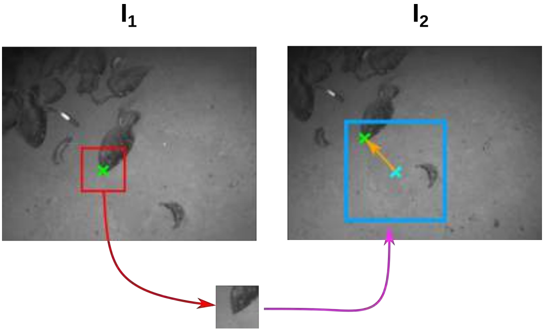


Underwater Features Tracking

Direct methods

- Tracking by searching for photometric minima
- Optical Flow (KLT) :

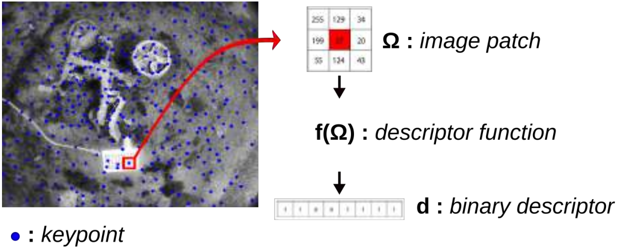
$$\underset{du,dv}{\arg\min} \sum_u \sum_v (\mathbf{I}_1(u,v) - \mathbf{I}_2(u+du, v+dv))^2$$



Underwater Features Tracking

Indirect methods

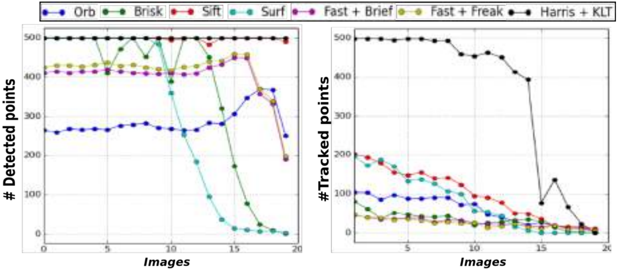
- Use descriptors (vectors)
- Similarity score between descriptors
- Descriptors : BRIEF, BRISK, FREAK, ORB, SURF, SIFT, ...



Underwater Features Tracking

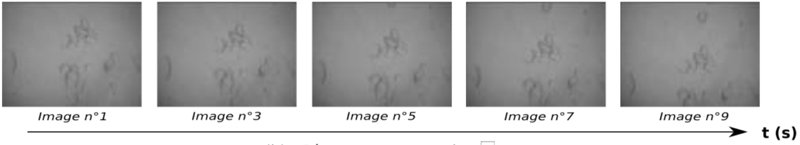
Evaluation of robustness to turbidity

(a) : TURBID dataset

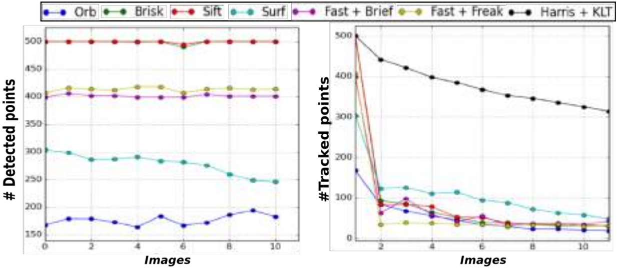


Underwater Features Tracking

Evaluation of tracking efficiency on a real sequence



(b) : Séquence sous-marine



Underwater Features Tracking

Conclusion

- Optical Flow (KLT) is very efficient
- Descriptors get too ambiguous for efficient tracking

2. Robust Underwater Monocular Visual SLAM

Robust Underwater Monocular Visual SLAM

- UW-VO : Keyframe-based monocular VSLAM
- Frame-to-frame features tracking from KLT
- Retracking mechanism

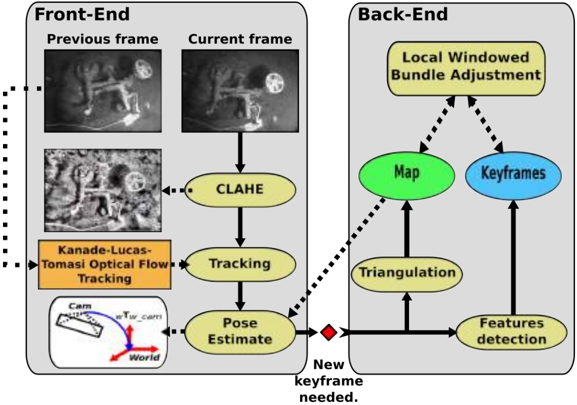
Robust Underwater Monocular Visual SLAM

Problem Statement

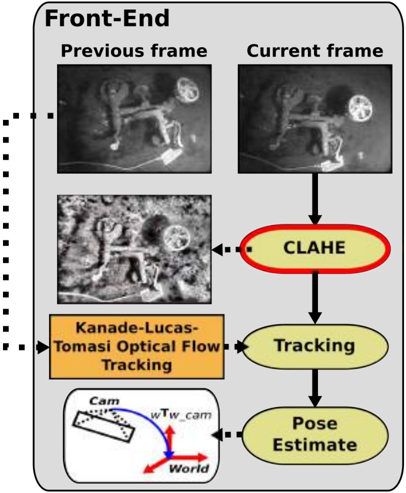
- Estimate the pose of the camera at each new image
- Pose : $\mathbf{X_i} = (\mathbf{R}, \mathbf{t}) \in \mathbb{SE}(3) \quad | \quad \mathbf{R} \in \mathbb{SO}(3) \quad \mathbf{t} \in \mathbb{R}^3$
- Estimate the position of 3D landmarks : $\mathbf{lm_j} \in \mathbb{R}^3$

Robust Underwater Monocular Visual SLAM

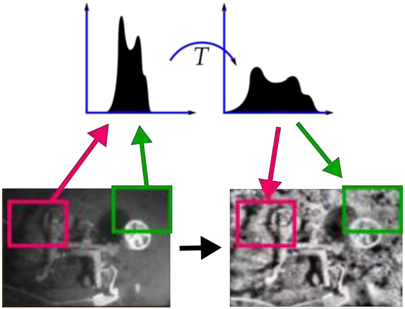
Tracking / Mapping : Two threads for efficient computation



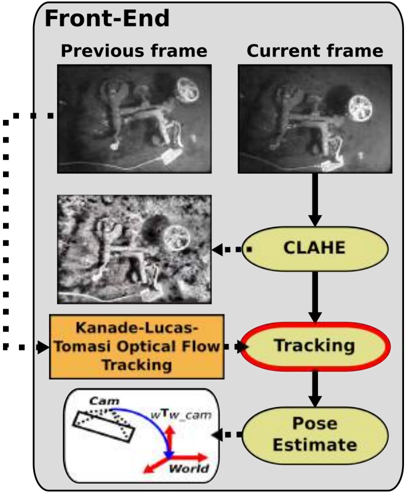
Robust Underwater Monocular Visual SLAM



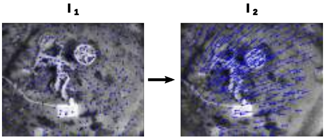
- **Pre-processing :** Contrast Local Adaptive Histogram Equalization (CLAHE)



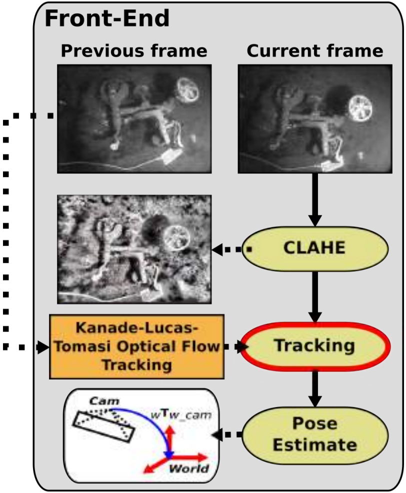
Robust Underwater Monocular Visual SLAM



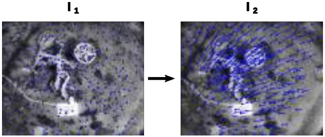
- **Features tracking :**
  Frame-to-frame KLT



Robust Underwater Monocular Visual SLAM



- **Features tracking :** Frame-to-frame KLT



- **KLT not robust to occlusions**

Robust Underwater Monocular Visual SLAM

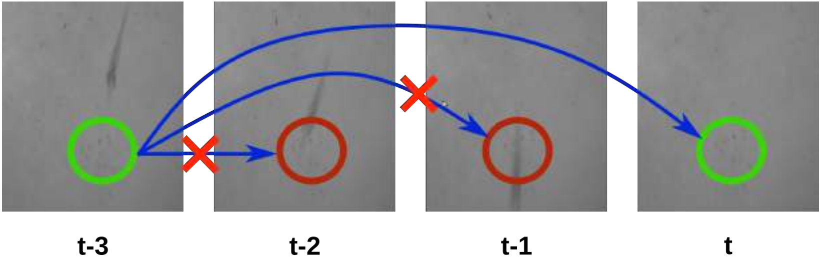
Many short occlusions due to moving fishes

Credit : DRASSM

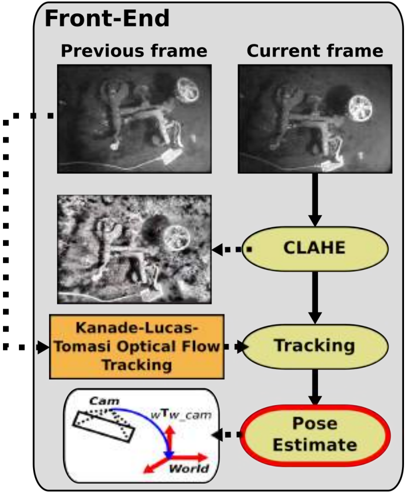
Robust Underwater Monocular Visual SLAM

Retracking mechanism

- Store the most recent images + lost features
- Multi-frame KLT retracking



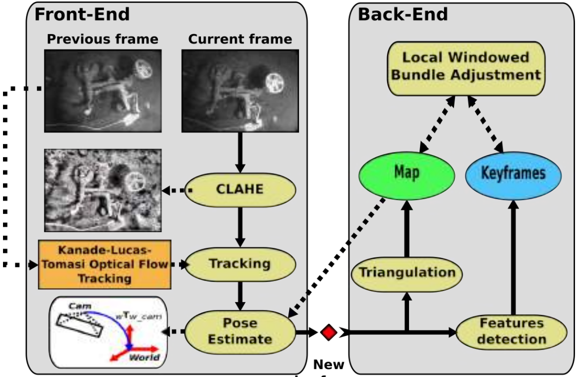
Robust Underwater Monocular Visual SLAM



- **Pose estimation :**
  - Use 2D / 3D observations

Robust Underwater Monocular Visual SLAM

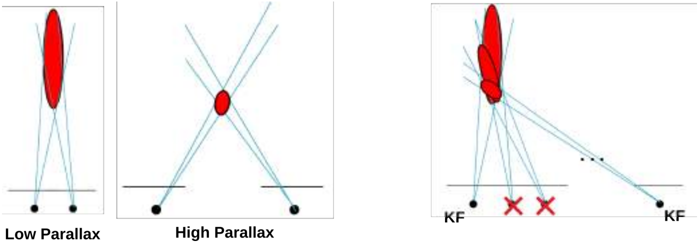
Keyframe selection decision



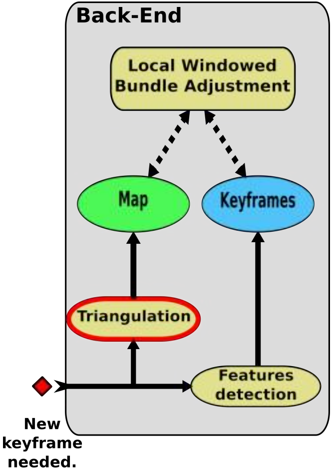
Robust Underwater Monocular Visual SLAM

Keyframe selection decision

- Enough motion since last keyframe



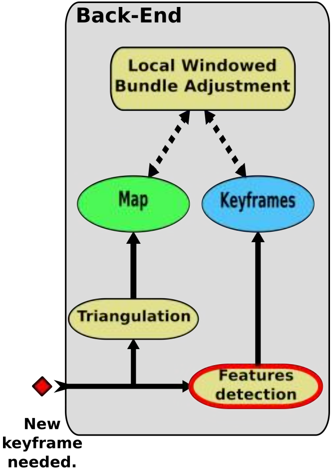
Robust Underwater Monocular Visual SLAM



Mapping thread

- Triangulation of new 3D points from 2D / 2D features between previous and current keyframes

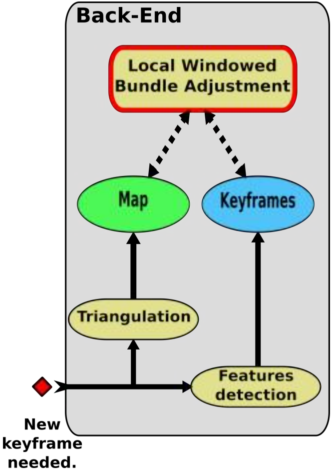
Robust Underwater Monocular Visual SLAM



Mapping thread

- Triangulation of new 3D points from 2D / 2D features between previous and current keyframes
- Detection of new 2D features to track (for next triangulation)

Robust Underwater Monocular Visual SLAM



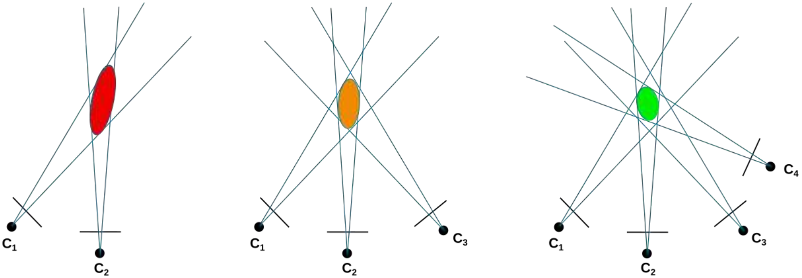
Mapping thread

- Triangulation of new 3D points from 2D / 2D features between previous and current keyframes
- Detection of new 2D features to track (for next triangulation)
- Optimization of the 3D map : Bundle Adjustment

Robust Underwater Monocular Visual SLAM

Bundle Adjustment

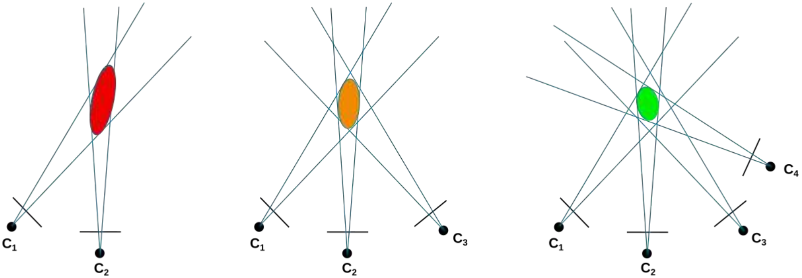
- Triangulation from two views not accurate because of noise



Robust Underwater Monocular Visual SLAM

Bundle Adjustment

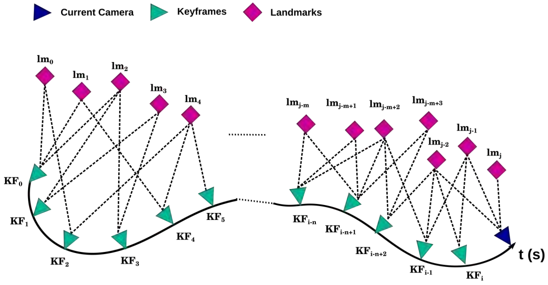
- Triangulation from two views not accurate because of noise
- Apply multi-view constraints for trajectory and 3D map optimization



Robust Underwater Monocular Visual SLAM

Bundle Adjustment

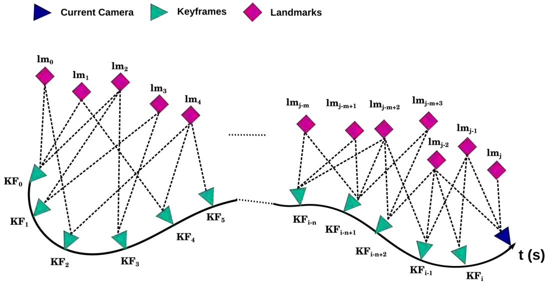
- Triangulation from two views not accurate because of noise
- Apply multi-view constraints for trajectory and 3D map optimization



Robust Underwater Monocular Visual SLAM

Bundle Adjustment

- Triangulation from two views not accurate because of noise
- Apply multi-view constraints for trajectory and 3D map optimization



- **Factor Graph $\rightarrow$ Maximum Likelihood Estimation**

Robust Underwater Monocular Visual SLAM

Bundle Adjustment : Maximum Likelihood Estimation

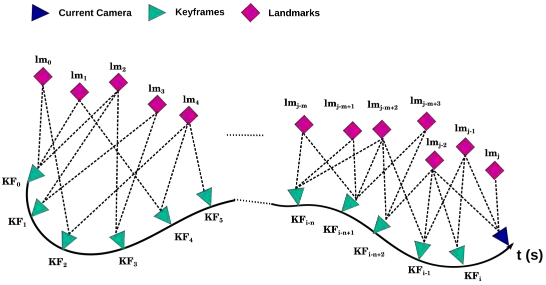
- Minimization of reprojection errors :

$$\chi^* = \underset{\chi}{\arg\min} \left(E_{visual} \left(\chi \right) \right) \quad \chi = \begin{bmatrix} \mathbf{X}_{KF_i} & \mathbf{lm}_j \end{bmatrix}^T$$

- Non-linear optimization solved with Levenberg-Marquardt

Robust Underwater Monocular Visual SLAM

Bundle Adjustment : Maximum Likelihood Estimation

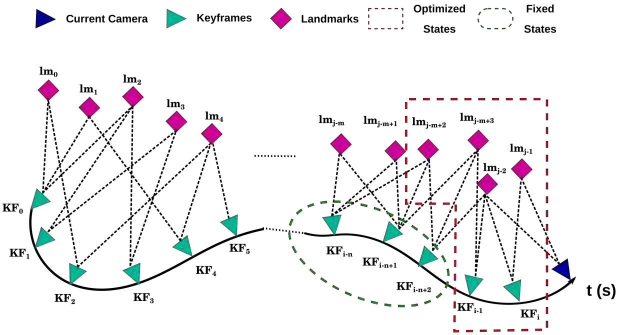


- **Full problem not tractable in real-time**

Robust Underwater Monocular Visual SLAM

Local Adaptive Windowed BA

- Optimize most recent keyframes and 3D landmarks only
- Monocular setup : scale unobservable ⇒ fix at least two keyframes



Robust Underwater Monocular Visual SLAM

Experimental Results

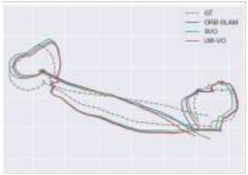
- Evaluation of UW-VO against ORB-SLAM, LSD-SLAM and SVO
- Video sequences acquired on a shipwreck (300 meters) by the DRASSM
- Monocular SLAM $\Rightarrow$ scaling w.r.t. groundtruth

Robust Underwater Monocular Visual SLAM

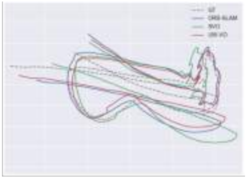
Experimental Results

Seq. #	Duration	Turbidity Level	Short Occlusions	Absolute Trajectory Error RMSE (in %)			
				LSD-SLAM	ORB-SLAM	SVO	UW-VO
1	4’	Low	Few	X	1.67	**1.63**	1.76
2	2’30”	Medium	Some	X	1.91	2.45	**1.73**
3	22”	High	Many	X	X	1.57	**1.04**
4	4’30”	Low	Many	X	**1.13**	X	1.58
5	3’15”	Medium	Many	X	1.94	X	**1.88**

TABLE – Sequences taken on a shipwreck (300 meters deep).



Sequence #1



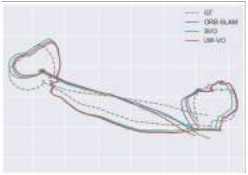
Sequence #2

Robust Underwater Monocular Visual SLAM

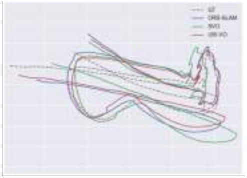
Experimental Results

				Absolute Trajectory Error RMSE (in %)			
Seq. #	Duration	Turbidity Level	Short Occlusions	LSD-SLAM	ORB-SLAM	SVO	UW-VO
1	4'	Low	Few	X	1.67	**1.63**	1.76
2	2'30"	Medium	Some	X	1.91	2.45	**1.73**
3	22"	High	Many	X	X	1.57	**1.04**
4	4'30"	Low	Many	X	**1.13**	X	1.58
5	3'15"	Medium	Many	X	1.94	X	**1.88**

TABLE – Sequences taken on a shipwreck (300 meters deep).



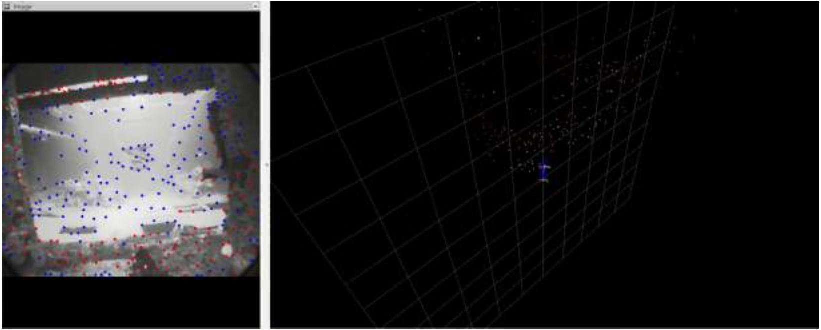
Sequence #1



Sequence #2

Robust Underwater Monocular Visual SLAM

UW-VO for localization during shipwreck exploration



Robust Underwater Monocular Visual SLAM

Conclusion

- Robust to underwater imaging conditions
- Accurate localization
- Real-time

Robust Underwater Monocular Visual SLAM

Conclusion

- Robust to underwater imaging conditions
- Accurate localization
- Real-time

- Monocular $\Rightarrow$ No scale

3. Multi-sensors SLAM

Tight Fusion : Insert other measurement modalities within the factor graph

Tight Fusion : Insert other measurement modalities within the factor graph

Fusion from Maximum Likelihood Estimation

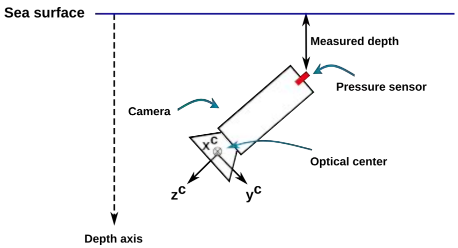
$$\mathcal{X}^* = \underset{\mathcal{X}}{\arg\min} \left(E_{visual}(\mathcal{X}) + E_{depth}(\mathcal{X}) + E_{IMU}(\mathcal{X}) \right)$$

- E_{visual} : Energy term based on visual measurements
- E_{depth} : Energy term based on pressure measurements
- E_{IMU} : Energy term based on inertial measurements

3.1. Visual-Pressure SLAM

Visual-Pressure SLAM

Visual-Pressure Setup



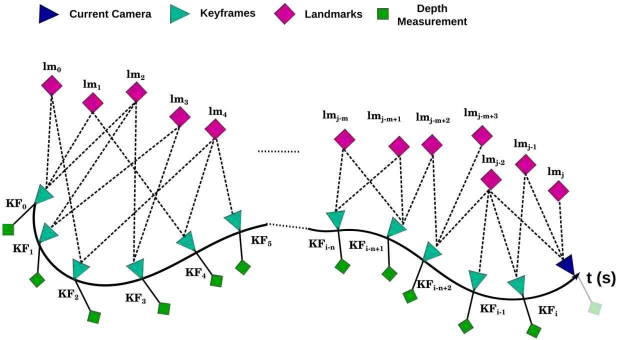
- Pressure measurements : pressure (Pa) $\propto$ depth (m)
- Depth variation from starting point :

$$\tilde{d}_i = {}_{\text{raw}}\tilde{d}_i - {}_{\text{raw}}\tilde{d}_0$$

Visual-Pressure SLAM

Strategy 1 : Integration of absolute depth measurements

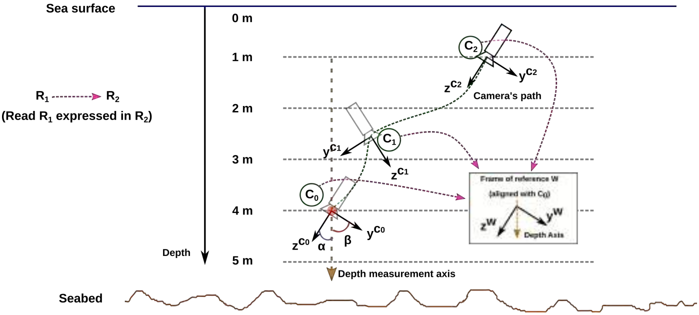
- Depth error term : $E_{depth}(\mathbf{X}_i) = \|\tilde{d}_i - \hat{t^z}_{Wc_i}\|^2_{\sigma^2_{depth}}$



Visual-Pressure SLAM

Visual-Pressure Fusion

- Misalignement issue!



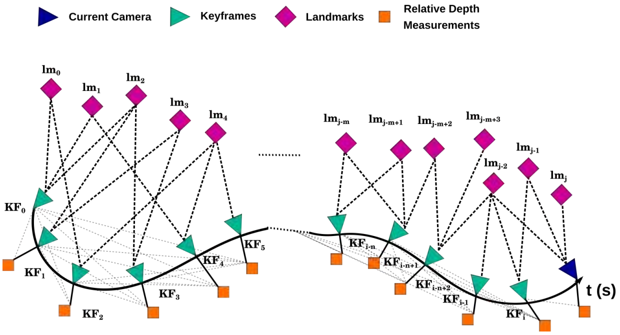
- Linear error along z^W : $\overline{d}_i = \overline{t^z}_{c_0 c_i} \cdot \cos(\alpha)$

Visual-Pressure SLAM

Strategy 2 : Integration of relative depth measurements

- Relative depth :

$$E_{depth}\left(\mathbf{X}_k, \mathbf{X}_i\right) = \left(\| \left(\tilde{d}_i - \tilde{d}_k\right) - \left(\hat{t}^z{}_{wc_i} - \hat{t}^z{}_{wc_k}\right) \|^2_{(2\cdot\sigma_{depth})^2} \right)$$



Visual-Pressure SLAM

Experimental Results

- **Init. Only** : UW-VO with scale factor estimation from 1st meas. only

$$\chi^* = \underset{\chi}{\arg\min} \left(E_{visual}(\chi) \right)$$

- **UW-VP** : Strategy 1 vs Strategy 2

$$\chi^* = \underset{\chi}{\arg\min} \left(E_{visual}(\chi) + \color{red}{E_{depth}}(\chi) \right)$$

- Dataset : sequences from AQUALOC
  - 7 sequences in a harbor
  - 2 sequences on a shipwreck (400 meters)

Visual-Pressure SLAM

Experimental Results

TABLE – Absolute trajectory errors (RMSE in m).

		Absolute Trajectory Error (m)		
		Init. Only	*UW-VP*	
Seq.	Length (m)	UW-VO	Strat. 1	Strat. 2
# 1	39.3	1.01	0.55	**0.53**
# 2	75.6	1.70	1.23	**0.40**
# 3	23.6	0.52	0.30	**0.26**
# 4	55.8	X	X	X
# 5	28.5	0.96	0.19	**0.11**
# 6	19.5	0.17	0.11	**0.06**
# 7	32.9	X	X	X
# A	41.2	0.96	0.58	**0.52**
# B	65.4	1.3	0.99	**0.90**

Visual-Pressure SLAM

Conclusion

- Recovery of the scale factor
- Improved localization accuracy

Visual-Pressure SLAM

Conclusion

- Recovery of the scale factor
- Improved localization accuracy

- Misalignement issue has to be taken care of!

Visual-Pressure SLAM

Conclusion

- Recovery of the scale factor
- Improved localization accuracy

- Misalignement issue has to be taken care of!

- Still fully dependent on vision

3.2. Visual-Inertial-Pressure SLAM

Visual-Inertial-Pressure SLAM

Low-cost MEMS-IMU Model

- Angular Velocity measurements :

$$\tilde{\omega}_B(t) = \omega_B(t) + \mathbf{b}^g(t) + \eta^g$$

- Linear Acceleration measurements :

$$\tilde{\mathbf{a}}_B(t) = \mathbf{R}_{WB}(t)^T \cdot (\mathbf{a}_W(t) - \mathbf{g}_W) + \mathbf{b}^a(t) + \eta^a$$

Visual-Inertial-Pressure SLAM

Low-cost MEMS-IMU Model

- Angular Velocity measurements :

$$\tilde{\omega}_B(t) = \omega_B(t) + \mathbf{b}^g(t) + \eta^g$$

- Linear Acceleration measurements :

$$\tilde{\mathbf{a}}_B(t) = \mathbf{R}_{WB}(t)^T \cdot (\mathbf{a}_W(t) - \mathbf{g}_W) + \mathbf{b}^a(t) + \eta^a$$

- ▶ Measurements corrupted by time-varying biases and zero-mean gaussian noise

Visual-Inertial-Pressure SLAM

IMU Measurements

- Motion estimations from IMU meas. $\Rightarrow \mathbf{R}_{WBi}, \mathbf{v}_{WBi}, \mathbf{p}_{WBi}$

Visual-Inertial-Pressure SLAM

IMU Measurements

- Motion estimations from IMU meas. $\Rightarrow \mathbf{R}_{WBi}, \mathbf{v}_{WBi}, \mathbf{p}_{WBi}$

- Motion information at high rates (200 Hz)

Visual-Inertial-Pressure SLAM

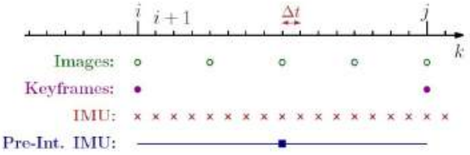
IMU Measurements

- Motion estimations from IMU meas. $\Rightarrow \mathbf{R}_{WBi}, \mathbf{v}_{WBi}, \mathbf{p}_{WBi}$
  - Motion information at high rates (200 Hz)
  - But big drift because of varying biases and noise

Visual-Inertial-Pressure SLAM

IMU Preintegration

- Summarize intra-keyframe IMU measurements as one measurement :



- Relative motion measurements : $\Delta\tilde{\mathbf{R}}_{BiBj}, \Delta\tilde{\mathbf{p}}_{BiBj}, \Delta\tilde{\mathbf{v}}_{BiBj}$

- Easy insertion in the Factor Graph formulation

Visual-Inertial-Pressure SLAM

New state to estimate :

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{R}_{WBi} & \mathbf{p}_{WBi} & \mathbf{v}_{WBi} & \mathbf{b}_i^g & \mathbf{b}_i^a \end{bmatrix}^T$$

Visual-Inertial-Pressure SLAM

New state to estimate :

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{R}_{WBi} & \mathbf{p}_{WBi} & \mathbf{v}_{WBi} & \mathbf{b}_i^g & \mathbf{b}_i^a \end{bmatrix}^T$$

IMU Preintegration : Relative errors between keyframes

$$\mathbf{e}_{\Delta\mathbf{R}_{BiBj}} = \hat{\mathbf{R}}_{BiBj} \boxminus \Delta\tilde{\mathbf{R}}_{BiBj}$$
$$\mathbf{e}_{\Delta\mathbf{p}_{BiBj}} = \hat{\mathbf{p}}_{BiBj} - \Delta\tilde{\mathbf{p}}_{BiBj}$$
$$\mathbf{e}_{\Delta\mathbf{v}_{BiBj}} = \hat{\mathbf{v}}_{BiBj} - \Delta\tilde{\mathbf{v}}_{BiBj}$$

$$\mathbf{e}_{\Delta\mathbf{b}^g_{BiBj}} = \hat{\mathbf{b}}^g_{Bj} - \hat{\mathbf{b}}^g_{Bi}$$
$$\mathbf{e}_{\Delta\mathbf{b}^a_{BiBj}} = \hat{\mathbf{b}}^a_{Bj} - \hat{\mathbf{b}}^a_{Bi}$$

Random-walk biases

Visual-Inertial-Pressure SLAM

New state to estimate :

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{R}_{WBi} & \mathbf{p}_{WBi} & \mathbf{v}_{WBi} & \mathbf{b}_i^g & \mathbf{b}_i^a \end{bmatrix}^T$$

IMU Preintegration : Relative errors between keyframes

$$\mathbf{e}_{\Delta \mathbf{R}_{BiBj}} = \hat{\mathbf{R}}_{BiBj} \boxminus \Delta\tilde{\mathbf{R}}_{BiBj}$$

$$\mathbf{e}_{\Delta \mathbf{p}_{BiBj}} = \hat{\mathbf{p}}_{BiBj} - \Delta\tilde{\mathbf{p}}_{BiBj}$$

$$\mathbf{e}_{\Delta \mathbf{v}_{BiBj}} = \hat{\mathbf{v}}_{BiBj} - \Delta\tilde{\mathbf{v}}_{BiBj}$$

$$\mathbf{e}_{\Delta \mathbf{b}^g_{BiBj}} = \hat{\mathbf{b}}^g_{Bj} - \hat{\mathbf{b}}^g_{Bi}$$

$$\mathbf{e}_{\Delta \mathbf{b}^a_{BiBj}} = \hat{\mathbf{b}}^a_{Bj} - \hat{\mathbf{b}}^a_{Bi}$$

Random-walk biases

IMU Energy term

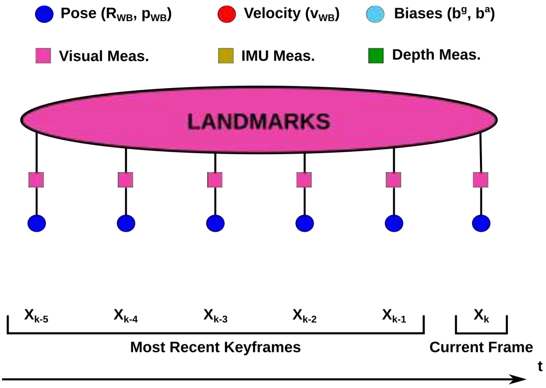
$$E_{IMU}(\chi) = \sum_{\mathfrak{K}^*} \left(\mathbf{e}_{imu}(\mathbf{X}_i, \mathbf{X}_j)^T \cdot {\mathbf{\Sigma}^{imu}_{BiBj}}^{-1} \cdot \mathbf{e}_{imu}(\mathbf{X}_i, \mathbf{X}_j) \right)$$

$$\mathbf{e}_{imu}(\mathbf{X}_i, \mathbf{X}_j) = \begin{bmatrix} \mathbf{e}_{\Delta \mathbf{R}_{BiBj}} & \mathbf{e}_{\Delta \mathbf{p}_{BiBj}} & \mathbf{e}_{\Delta \mathbf{v}_{BiBj}} & \mathbf{e}_{\Delta \mathbf{b}^g_{BiBj}} & \mathbf{e}_{\Delta \mathbf{b}^a_{BiBj}} \end{bmatrix}^T$$

Visual-Inertial-Pressure SLAM

Visual-Inertial-Pressure Optimization

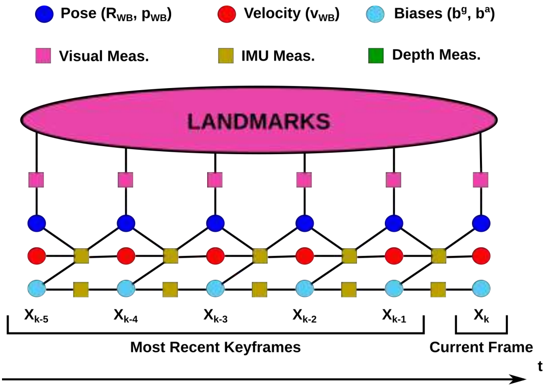
$$\mathcal{X}^* = \underset{\mathcal{X}}{\arg\min} \left(E_{visual}\left(\mathcal{X}\right)\right)$$



Visual-Inertial-Pressure SLAM

Visual-Inertial-Pressure Optimization

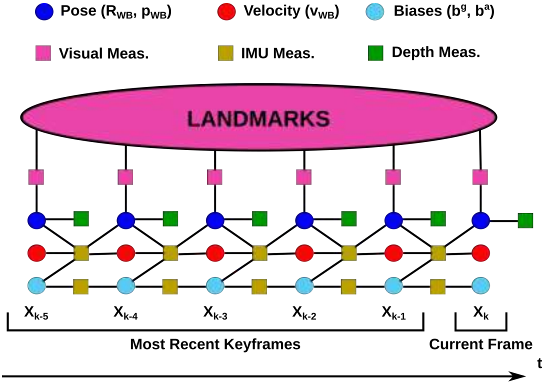
$$\chi^* = \arg\min_{\chi} \left(E_{visual}(\chi) + E_{IMU}(\chi) \right)$$



Visual-Inertial-Pressure SLAM

Visual-Inertial-Pressure Optimization

$$\chi^* = \arg\min_{\chi} \left(E_{visual}(\chi) + E_{IMU}(\chi) + E_{depth}(\chi) \right)$$



Visual-Inertial-Pressure SLAM

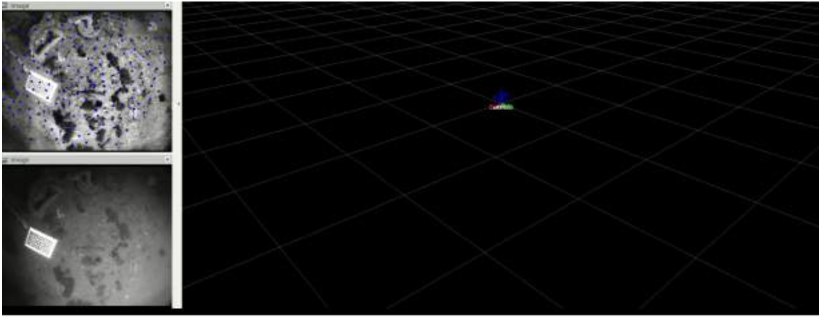
Experimental Results

TABLE – Absolute trajectory errors (RMSE in m).

		Absolute Trajectory Error (m)	
Seq.	Length (m)	UW-VP	UW-VIP
# 1	39.3	0.49	**0.42**
# 2	75.6	**0.36**	0.37
# 3	23.6	**0.25**	0.26
# 4	55.8	X	**1.56**
# 5	28.5	0.13	**0.09**
# 6	19.5	**0.04**	0.06
# 7	32.9	X	**1.16**
# A	41.2	**0.34**	0.36
# B	65.4	0.72	**0.69**

Visual-Inertial-Pressure SLAM

UW-VIP for localization with short loss of visual information



Visual-Inertial-Pressure SLAM

Conclusion

- Robust to short loss of visual information

Visual-Inertial-Pressure SLAM

Conclusion

- Robust to short loss of visual information
- Factor graph formulation could be used to fuse even more sensors!

4. Monocular Dense 3D Mapping

Monocular Dense 3D Mapping

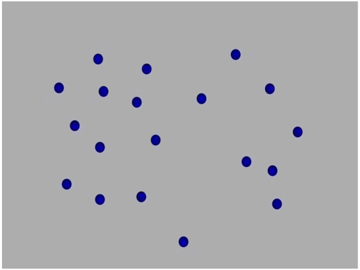
Dense 3D Mapping

- Densify the sparse 3D measurements
- Make use of optimized states : keyframes + 3D landmarks

Monocular Dense 3D Mapping

Depth Map Densification

- Find 3D features nearest-neighbors from 2D Delaunay triangulation

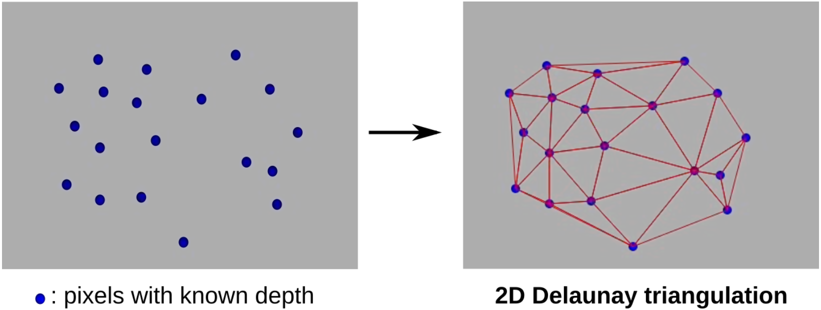


● : pixels with known depth

Monocular Dense 3D Mapping

Depth Map Densification

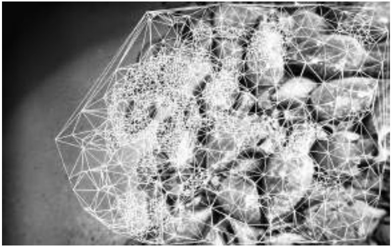
- Find 3D features nearest-neighbors from 2D Delaunay triangulation
- Depth value interpolation from Delaunay triangles



Monocular Dense 3D Mapping

Depth Map Densification

- Find 3D features nearest-neighbors from 2D Delaunay triangulation
- Depth value interpolation from Delaunay triangles

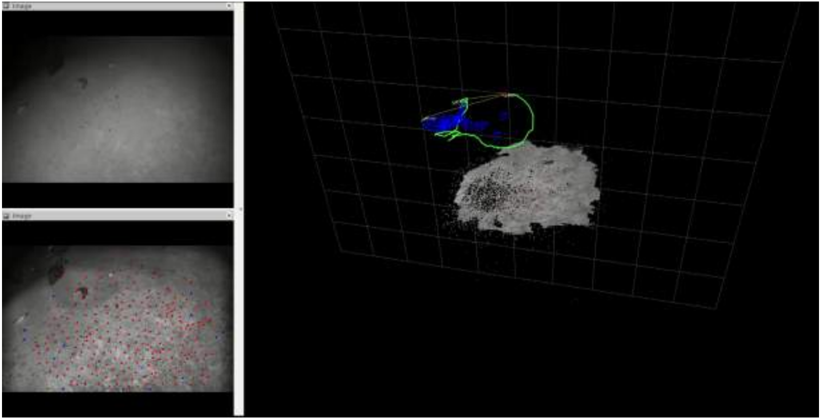


(a) 2D Delaunay triangulation.

(b) 2D densified depth map.

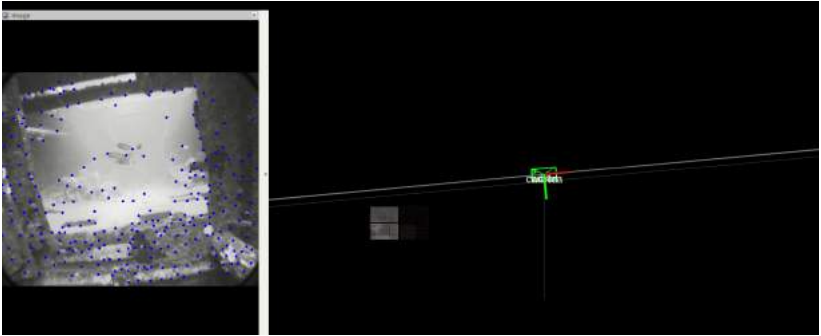
Monocular Dense 3D Mapping

Online 3D Reconstruction



Monocular Dense 3D Mapping

Online 3D Reconstruction in Complex Environment



Monocular Dense 3D Mapping

Conclusion

- Dense 3D reconstruction from monocular camera
- Real-time dense 3D reconstruction (but delayed)

Conclusion

Conclusion

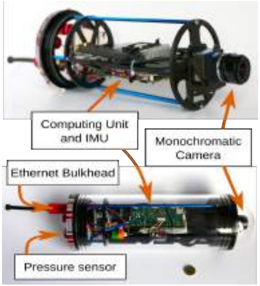
Contributions

- KLT is well suited for VSLAM tasks on underwater images
- Robust underwater monocular VSLAM method : UW-VO
- Tight fusion framework for Visual-Inertial-Pressure SLAM
- Dense 3D reconstruction module for monocular setup

Conclusion

Experimental Validation

- Algorithms validated on the Tegra TX2
- All the methods run in real-time
- Release of a public dataset : AQUALOC



Conclusion

Perspectives

- Add loop closure for drift reduction and relocalization
  $\rightarrow$ Online Bag of Words (Angeli et al., 2008 ; Garcia-Fidalgo et al., 2018 ; Nicosevici et al., 2012)
- Binocular SLAM extension
  $\rightarrow$ increased robustness
- Integration of the SLAM estimates in ROV's command :
  - Servoing
  - Autonomous navigation
  - Automatic photogrammetry

Conclusion

Publications & Dépôt logiciel

Journal Papers

Maxime FERRERA, Vincent CREUZE, Julien MORAS et Pauline TROUVÉ-PELOUX (2019a). "AQUALOC : An Underwater Dataset for Visual-Inertial-Pressure Localization.". In : **The International Journal of Robotics Research**

Maxime FERRERA, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2019b). "Real-Time Monocular Visual Odometry for Turbid and Dynamic Underwater Environments". In : **Sensors**. T. 19. 3

International Conference Papers

Maxime FERRERA, Julien MORAS, Pauline TROUVÉ-PELOUX, Vincent CREUZE et Denis DÉGEZ (2018a). "The Aqualoc Dataset: Towards Real-Time Underwater Localization from a Visual-Inertial-Pressure Acquisition System". In : **IROS Workshop - New Horizons for Underwater Intervention Missions : from Current Technologies to Future Applications**

National Conference Papers

Maxime FERRERA, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2018b). "Odométrie Visuelle Monoculaire en Environnement Sous-Marin". In : **Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)**

Maxime FERRERA, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2017). "Localisation autonome basée vision d'un robot sous-marin et cartographie de précision". In : **ORASIS**

Conclusion

Thank you for your attention!

ANGELI, Adrien, David FILLIAT, Stéphane DONCIEUX et Jean-Arcady MEYER (2008). “A fast and incremental method for loop-closure detection using bags of visual words”. In : **IEEE Transactions on Robotics**, p. 1027-1037.

BURGUERA, A., F. BONIN-FONT et G. OLIVER (2015). “Trajectory-Based Visual Localization in Underwater Surveying Missions”. In : **Sensors**. T. 15. 1, p. 1708-1735.

CREUZE, Vincent (2017). “Monocular Odometry for Underwater Vehicles with Online Estimation of the Scale Factor”. In : **IFAC 2017 World Congress**.

ENGEL, J., T. SCHOPS et D. CREMERS (2014). “LSD-SLAM: Large-Scale Direct Monocular SLAM”. In : **European Conference on Computer Vision (ECCV)**. Zurich, Switzerland, p. 834-849.

ENGEL, Jakob, Vladlen KOLTUN et Daniel CREMERS (2017). “Direct sparse odometry”. In : **IEEE transactions on pattern analysis and machine intelligence** 40.3, p. 611-625.

FERRERA, Maxime, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2017). “Localisation autonome basée vision d’un robot sous-marin et cartographie de précision”. In : **ORASIS**.

FERRERA, Maxime, Julien MORAS, Pauline TROUVÉ-PELOUX, Vincent CREUZE et Denis DÉGEZ (2018a). “The Aqualoc Dataset: Towards Real-Time Underwater Localization from a Visual-Inertial-Pressure Acquisition

System”. In : **IROS Workshop - New Horizons for Underwater Intervention Missions : from Current Technologies to Future Applications**.

FERRERA, Maxime, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2018b). “Odométrie Visuelle Monoculaire en Environnement Sous-Marin”. In : **Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)**.

FERRERA, Maxime, Vincent CREUZE, Julien MORAS et Pauline TROUVÉ-PELOUX (2019a). “AQUALOC : An Underwater Dataset for Visual-Inertial-Pressure Localization.”. In : **The International Journal of Robotics Research**.

FERRERA, Maxime, Julien MORAS, Pauline TROUVÉ-PELOUX et Vincent CREUZE (2019b). “Real-Time Monocular Visual Odometry for Turbid and Dynamic Underwater Environments”. In : **Sensors**. T. 19. 3.

FORSTER, C., M. PIZZOLI et D. SCARAMUZZA (2014). “SVO: Fast semi-direct monocular visual odometry”. In : **2014 IEEE International Conference on Robotics and Automation (ICRA)**.

GARCIA, Rafael, Xevi CUFI et Marc CARRERAS (2001). “Estimating the motion of an underwater robot from a monocular image sequence”. In : **Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)**. T. 3. IEEE, p. 1682-1687.

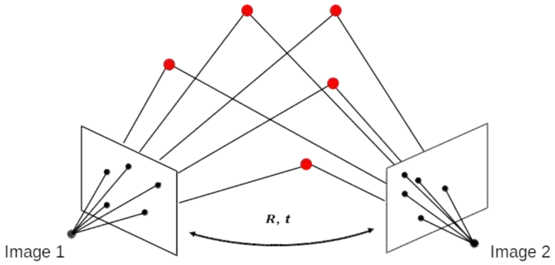
- Garcia-Fidalgo, Emilio et Alberto Ortiz (2018). “iBoW-LCD : An Appearance-Based Loop-Closure Detection Approach Using Incremental Bags of Binary Words”. In : **IEEE Robotics and Automation Letters** 3.4, p. 3051-3057. doi : 10.1109/LRA.2018.2849609.
- Kim, Ayoung et Ryan M Eustice (2013). “Real-time visual SLAM for autonomous underwater hull inspection using visual saliency”. In : **IEEE Transactions on Robotics** 29.3, p. 719-733.
- Klein, G. et D. Murray (2007). “Parallel Tracking and Mapping for Small AR Workspaces”. In : **IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)**. Nara, Japan, p. 225-234.
- Mur-Artal, R., J. M. M. Montiel et J. D. Tardós (2015). “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”. In : **IEEE Transactions on Robotics (T-RO)**. T. 31. 5, p. 1147-1163.
- Nicosevici, Tudor, Nuno Gracias, Shahriar Negahdaripour et Rafael Garcia (2009). “Efficient three-dimensional scene modeling and mosaicing”. In : **booktitle of Field Robotics**.
- Nicosevici, Tudor et Rafael Garcia (2012). “Automatic visual bag-of-words for online robot navigation and mapping”. In : **IEEE Transactions on Robotics** 28.4, p. 886-898.
- Shkurti, Florian, Ioannis Rekleitis, Milena Scaccia et Gregory Dudek (2011). “State estimation of an underwater robot using visual and inertial

information". In : **2011 IEEE/RSJ International Conference on Intelligent Robots and Systems**. IEEE, p. 5054-5060.

SINGH, Hanumant, Jonathan HOWLAND et Oscar PIZARRO (2004). "Advances in large-area photomosaicking underwater". In : **IEEE Journal of Oceanic Engineering** 29.3, p. 872-886.

Monocular Initialization

- 2D/2D initialisation from Essential Matrix
- Up-to-scale transformation ($\|\mathbf{t}\| = ?$)



- Scale arbitrary fixed : $\|\mathbf{t}\| = 1$

Bundle Adjustment

- Optimization of the keyframes and 3D landmarks :

$$\chi = \left\{\mathbf{X}_{\mathbf{KF_i}}, \mathbf{lm_j}\right\} \quad \chi = \left[\mathbf{X}_{cur} \quad \mathbf{X}_{KF_i} \quad \mathbf{lm}_j\right]^T$$

- Minimization of the reprojection errors :

$$\underset{\chi}{\arg\min} = \sum_i \sum_j \rho\left(\mathbf{e_{ij}}^T \cdot \mathbf{\Sigma_{ij}^{-1}} \cdot \mathbf{e_{ij}}\right)$$

$$\mathbf{e_{ij}} = \mathbf{x_{ij}} - proj(\mathbf{X}_{\mathbf{KF_i}}, \mathbf{lm_j}) : \text{Reprojection error}$$

$$\rho(\cdot) : \text{Huber norm}$$

Bundle Adjustment

- Levenberg-Marquardt Algorithm :

$$\left(\underbrace{\mathbf{J}_{\delta\chi}^{T}(\chi)\cdot\mathbf{\Sigma}_{visual}^{-1}\cdot\mathbf{J}_{\delta\chi}(\chi)}_{\mathbf{A}}+\lambda\cdot\mathrm{diag}(\mathbf{A})\right)\delta\chi=-\mathbf{J}_{\delta\chi}^{T}(\chi)\cdot\mathbf{\Sigma}_{visual}^{-1}\cdot e(\chi)$$

Bundle Adjustment

- Levenberg-Marquardt Algorithm :

$$\left(\underbrace{\mathbf{J}_{\delta\chi}^{T}(\chi) \cdot \mathbf{\Sigma}_{visual}^{-1} \cdot \mathbf{J}_{\delta\chi}(\chi)}_{\mathbf{A}} + \lambda \cdot \mathrm{diag}(\mathbf{A}) \right) \delta\chi = -\mathbf{J}_{\delta\chi}^{T}(\chi) \cdot \mathbf{\Sigma}_{visual}^{-1} \cdot e(\chi)$$

- Pose defined on $\mathbb{SE}(3)$!

Bundle Adjustment

- Levenberg-Marquardt Algorithm :

$$\left(\underbrace{\mathbf{J}_{\delta\chi}^{T}(\chi)\cdot\mathbf{\Sigma}_{visual}^{-1}\cdot\mathbf{J}_{\delta\chi}(\chi)}_{\mathbf{A}}+\lambda\cdot\mathrm{diag}(\mathbf{A})\right)\delta\chi=-\mathbf{J}_{\delta\chi}^{T}(\chi)\cdot\mathbf{\Sigma}_{visual}^{-1}\cdot e(\chi)$$

- Pose defined on $\mathbb{SE}(3)$!

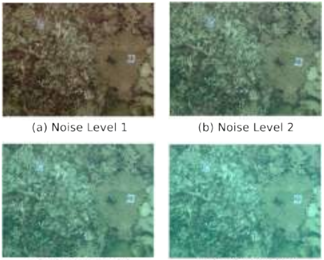
- On-manifold optimization :

$$proj(\mathbf{X}_i\boxplus\boldsymbol{\delta}\mathbf{X}_i,\mathbf{lm}_j\boxplus\boldsymbol{\delta}\mathbf{lm}_j)\rightarrow\mathbf{X}_i\in\mathbb{SE}(3)\text{ and }\boldsymbol{\delta}\mathbf{X}_i\in\mathbb{R}^6$$

$$\mathbf{X}_i\leftarrow\mathbf{X}_i\cdot\underbrace{Exp(\boldsymbol{\delta}\mathbf{X}_i)}_{\mathfrak{se}(3)}\quad;\quad\mathbf{lm}_j\leftarrow\mathbf{lm}_j+\boldsymbol{\delta}\mathbf{lm}_j$$

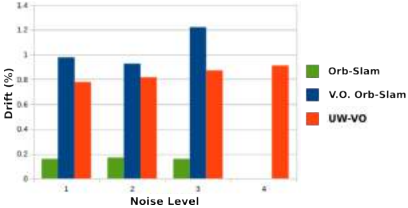
Robust Underwater Monocular Visual SLAM

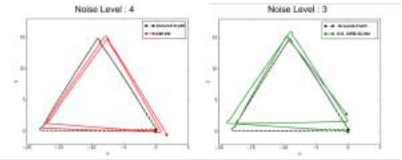
Experimental Results : Synthetic Turbid Sequences



(a) Noise Level 1 (b) Noise Level 2

(c) Noise Level 3 (d) Noise Level 4





(a) (b)

- ORB-SLAM : Loop closing feature
- ORB-SLAM : Not robust to mid-level and high-level of turbidity
- UW-VO : Better accuracy in terms of pure localization

Visual-Pressure SLAM

Is the depth axis observable ?

Visual-Pressure SLAM

Is the depth axis observable ?

$\Rightarrow$ **Relaxation of the gauge constraints**

Visual-Pressure SLAM

Is the depth axis observable ?

$\Rightarrow$ **Relaxation of the gauge constraints**

- Monocular SLAM : no scale $\rightarrow$ fix at least two keyframes in BA

Visual-Pressure SLAM

Is the depth axis observable ?

$\Rightarrow$ **Relaxation of the gauge constraints**

- Monocular SLAM : no scale $\rightarrow$ fix at least two keyframes in BA
- SLAM with scale : fix one keyframe in BA $\rightarrow$ fix the localization frame

Visual-Pressure SLAM

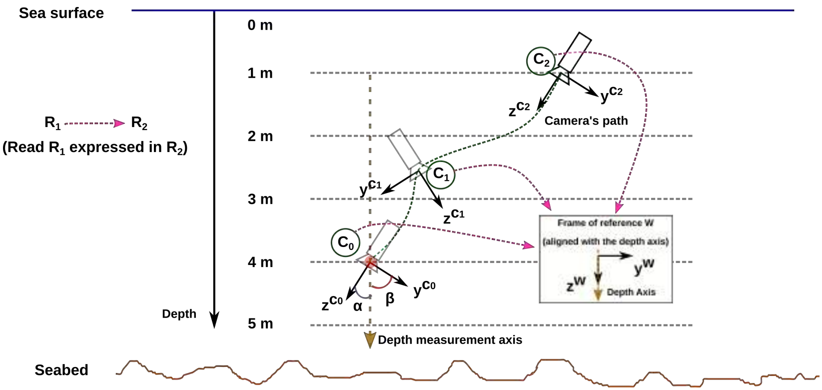
Is the depth axis observable ?

$\Rightarrow$ **Relaxation of the gauge constraints**

- Monocular SLAM : no scale $\rightarrow$ fix at least two keyframes in BA
- SLAM with scale : fix one keyframe in BA $\rightarrow$ fix the localization frame
- **Free Gauge :** no fixed state in BA during initialization phase

Visual-Pressure SLAM

Relaxation of the gauge constraints



Visual-Pressure SLAM

Experimental Results

TABLE – Absolute trajectory errors (RMSE in m) on the Harbor dataset.

		Absolute Trajectory Error (m)				
		Init. Only	*Regular*		*Free Gauge*	
Seq.	Length (m)	UW-VO	Strat. 1	Strat. 2	Strat. 1	Strat. 2
# 1	39.3	1.01	0.55	0.53	0.56	**0.49**
# 2	75.6	1.70	1.23	0.40	1.08	**0.36**
# 3	23.6	0.52	0.30	0.26	**0.23**	0.25
# 4	55.8	X	X	X	X	X
# 5	28.5	0.96	0.19	**0.11**	0.12	0.13
# 6	19.5	0.17	0.11	0.06	**0.04**	**0.04**
# 7	32.9	X	X	X	X	X
# A	41.2	0.96	0.58	0.52	0.44	**0.34**
# B	65.4	1.3	0.99	0.90	1.01	**0.72**

Visual-Inertial-Pressure SLAM

Low-cost MEMS-IMU Model

- Angular Velocity measurements :

$$\tilde{\omega}_B(t) = \omega_B(t) + \mathbf{b}^g(t) + \eta^g$$

- Linear Acceleration measurements :

$$\tilde{a}_B(t) = \mathbf{R}_{WB}(t)^T \cdot (a_W(t) - \mathbf{g}_W) + \mathbf{b}^a(t) + \eta^a$$

- Measurements at high rates (200 Hz) but big drift because of varying biases

Visual-Inertial-Pressure SLAM

New state to estimate :

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{R}_{WBi} & \mathbf{p}_{WBi} & \mathbf{v}_{WBi} & \mathbf{b}_i^g & \mathbf{b}_i^a \end{bmatrix}^T$$

Visual-Inertial-Pressure SLAM

New state to estimate :

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{R}_{WBi} & \mathbf{p}_{WBi} & \mathbf{v}_{WBi} & \mathbf{b}_i^g & \mathbf{b}_i^a \end{bmatrix}^T$$

IMU Measurements :

$$\mathbf{R}_{WBj} = \mathbf{R}_{WBi} \cdot \prod_{k=i}^{j-1} \cdot \operatorname{Exp}\left(\left(\tilde{\omega}_k - \mathbf{b}_k^g - \eta^g\right) \cdot \Delta t_{kk+1}\right)$$

$$\mathbf{v}_{WBj} = \mathbf{v}_{WBi} + \mathbf{g}_W \cdot \Delta t_{ij} + \sum_{k=i}^{j-1} \mathbf{R}_{WB_k} \cdot \left(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \eta^a\right) \cdot \Delta t_{kk+1}$$

$$\mathbf{p}_{WBj} = \mathbf{p}_{WBi} + \frac{1}{2} \cdot \mathbf{g}_W \cdot \Delta t_{ij}^2 + \sum_{k=i}^{j-1} \left[\mathbf{v}_{WB_k} \cdot \Delta t_{kk+1} + \frac{1}{2} \cdot \mathbf{R}_{WB_k} \cdot \left(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \eta^a\right) \cdot \Delta t_{kk+1}^2 \right]$$

Visual-Inertial-Pressure SLAM

Issues

- Measurements depend on optimized states!
- Every measurements have to be re-computed when states change

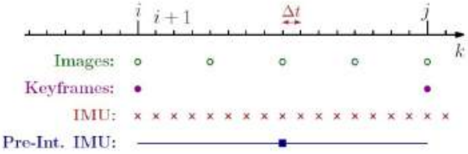
Visual-Inertial-Pressure SLAM

Issues

- Measurements depend on optimized states!
- Every measurements have to be re-computed when states change

IMU Preintegration

- Summarize intra-keyframe IMU measurements as one measurement :



- Remove dependency on optimized states
- Easy insertion in the Factor Graph formulation

Visual-Inertial-Pressure SLAM

IMU Preintegration

$$
\begin{aligned}
\Delta\tilde{\mathbf{R}}_{BiBj} &\doteq \mathbf{R}_{BiW} \cdot \mathbf{R}_{WBj} \\
&= \prod_{k=i}^{j-1} \cdot \mathrm{Exp}\left(\left(\tilde{\omega}_k - \mathbf{b}^g_k - \eta^g \right) \cdot \Delta t_{kk+1} \right) \\
\Delta\tilde{\mathbf{v}}_{BiBj} &\doteq \mathbf{R}_{BiW} \cdot \left(\mathbf{v}_{WBj} - \mathbf{v}_{WBi} - \mathbf{g}_W \cdot \Delta t_{ij} \right) \\
&= \sum_{k=i}^{j-1} \Delta\mathbf{R}_{BiB_k} \cdot \left(\tilde{\mathbf{a}}_k - \mathbf{b}^a_k - \eta^a \right) \cdot \Delta t_{kk+1} \\
\Delta\tilde{\mathbf{p}}_{BiBj} &\doteq \mathbf{R}_{BiW} \cdot \left(\mathbf{p}_{WBj} - \mathbf{p}_{WBi} - \mathbf{v}_{WBi} \cdot \Delta t_{ij} - \frac{1}{2} \cdot \mathbf{g}_W \cdot \Delta t_{ij}^2 \right) \\
&= \sum_{k=i}^{j-1} \left[\Delta\mathbf{v}_{WB_k} \cdot \Delta t_{kk+1} + \frac{1}{2} \cdot \Delta\mathbf{R}_{BiB_k} \cdot \left(\tilde{\mathbf{a}}_k - \mathbf{b}^a_k - \eta^a \right) \cdot \Delta t_{kk+1}^2 \right]
\end{aligned}
$$

Visual-Inertial-Pressure SLAM

IMU Preintegrated Measurements

$$\Delta\tilde{\mathbf{R}}_{BiBj} = \Delta\tilde{\mathbf{R}}_{BiBj} \cdot \mathrm{Exp}\left(\frac{\partial\Delta\tilde{\mathbf{R}}_{BiBj}}{\partial\mathbf{b}^g} \cdot \delta\mathbf{b}^g\right)$$

$$\Delta\tilde{\mathbf{v}}_{BiBj} = \Delta\tilde{\mathbf{v}}_{BiBj} + \frac{\partial\Delta\tilde{\mathbf{v}}_{BiBj}}{\partial\mathbf{b}^g} \cdot \delta\mathbf{b}^g + \frac{\partial\Delta\tilde{\mathbf{v}}_{BiBj}}{\partial\mathbf{b}^a} \cdot \delta\mathbf{b}^a$$

$$\Delta\tilde{\mathbf{p}}_{BiBj} = \Delta\tilde{\mathbf{p}}_{BiBj} + \frac{\partial\Delta\tilde{\mathbf{p}}_{BiBj}}{\partial\mathbf{b}^g} \cdot \delta\mathbf{b}^g + \frac{\partial\Delta\tilde{\mathbf{p}}_{BiBj}}{\partial\mathbf{b}^a} \cdot \delta\mathbf{b}^a$$

Biases Evolution :

$$\mathbf{b}^g(t+\Delta t) = \mathbf{b}^g(t) + \Big\rangle_{\boldsymbol{b_g}} \quad , \Big\rangle_{\boldsymbol{b_g}} \sim \mathcal{N}(\mathbf{0}_{3\times1}, \mathbf{I}_{3\times3} \cdot \sigma^2_{b_g})$$

$$\mathbf{b}^a(t+\Delta t) = \mathbf{b}^a(t) + \Big\rangle_{\boldsymbol{b_a}} \quad , \Big\rangle_{\boldsymbol{b_a}} \sim \mathcal{N}(\mathbf{0}_{3\times1}, \mathbf{I}_{3\times3} \cdot \sigma^2_{b_a})$$

Visual-Inertial-Pressure SLAM

IMU Preintegrated Measurements

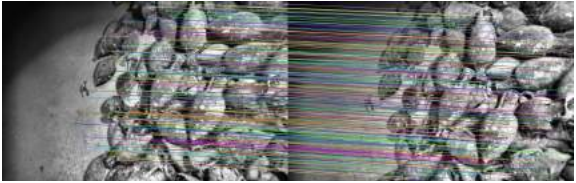
$$
\begin{aligned}
\Delta\tilde{\mathbf{R}}_{BiBj} &= \Delta\tilde{\mathbf{R}}_{BiBj} \cdot \operatorname{Exp}\left(\frac{\partial \Delta\tilde{\mathbf{R}}_{BiBj}}{\partial \mathbf{b}^g} \cdot \delta\mathbf{b}^g\right) \\
\Delta\tilde{\mathbf{v}}_{BiBj} &= \Delta\tilde{\mathbf{v}}_{BiBj} + \frac{\partial \Delta\tilde{\mathbf{v}}_{BiBj}}{\partial \mathbf{b}^g} \cdot \delta\mathbf{b}^g + \frac{\partial \Delta\tilde{\mathbf{v}}_{BiBj}}{\partial \mathbf{b}^a} \cdot \delta\mathbf{b}^a \\
\Delta\tilde{\mathbf{p}}_{BiBj} &= \Delta\tilde{\mathbf{p}}_{BiBj} + \frac{\partial \Delta\tilde{\mathbf{p}}_{BiBj}}{\partial \mathbf{b}^g} \cdot \delta\mathbf{b}^g + \frac{\partial \Delta\tilde{\mathbf{p}}_{BiBj}}{\partial \mathbf{b}^a} \cdot \delta\mathbf{b}^a
\end{aligned}
$$

Biases Evolution :

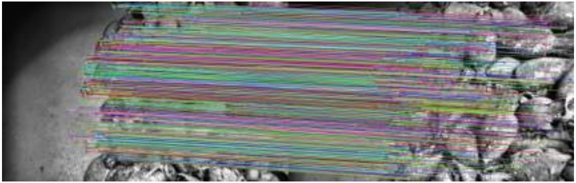
$$
\begin{aligned}
\mathbf{b}^g(t+\Delta t) &= \mathbf{b}^g(t) + \boldsymbol{b_g} \quad , \boldsymbol{b_g} \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \mathbf{I}_{3\times 3} \cdot \sigma_{b_g}^2) \\
\mathbf{b}^a(t+\Delta t) &= \mathbf{b}^a(t) + \boldsymbol{b_a} \quad , \boldsymbol{b_a} \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \mathbf{I}_{3\times 3} \cdot \sigma_{b_a}^2)
\end{aligned}
$$

Monocular Dense 3D Mapping

1. 3D Augmentation : Denser KLT tracking between last two optimized keyframes



(a) Initial set of 3D correspondences.



(b) Augmented set of 3D correspondences.

Monocular Dense 3D Mapping

3. Dense 3D Meshing

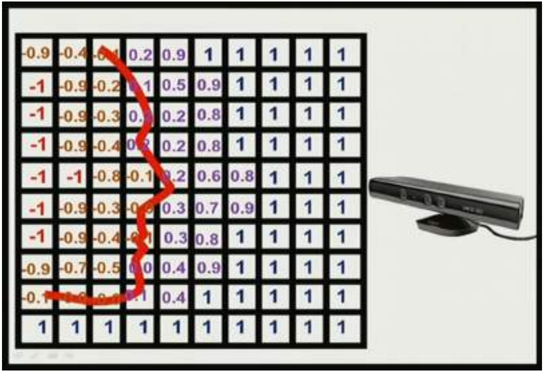


FIGURE – Example of a 2D TSDF grid (image taken from `http://pointclouds.org/documentation/tutorials/using_kinfu_large_scale.php`).