



GT2 Robotique Marine/Sous-marine (13/10/2021) Relocalisation de drone sous-marin grands fonds autonome (titre temporaire)

Vincent HUGEL
Claire DUNE
Aurélien ARNAUBEC
Ricard MARXER

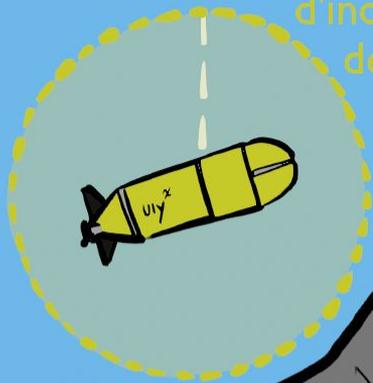
www.flotteoceanographique.fr

La Flotte océanographique française,
une très grande infrastructure de recherche opérée par l'Ifremer



6000 m de profondeur

~10-30 m
d'incertitude
de position



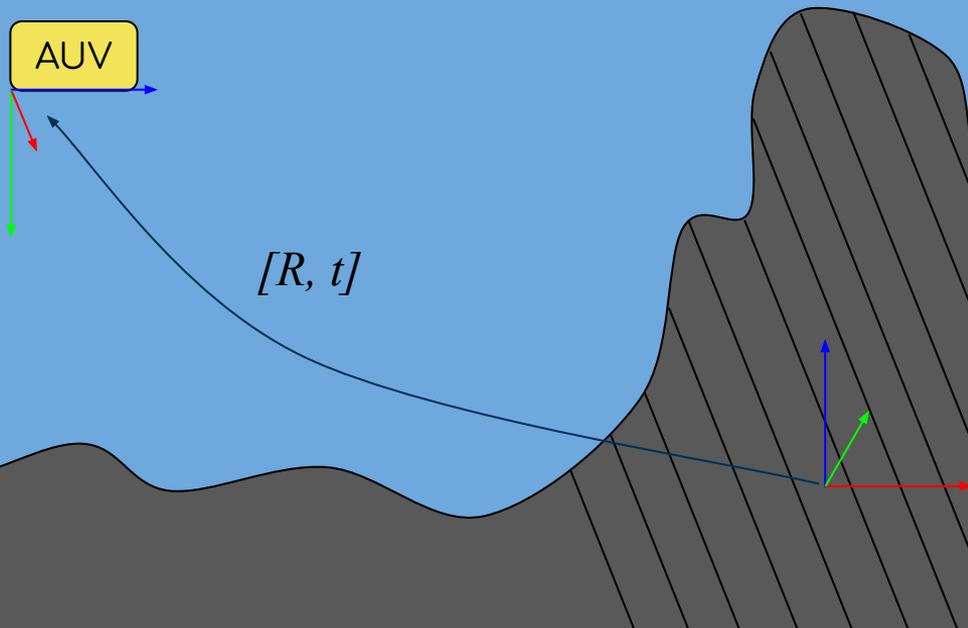
Zone cartographiée

~15 m



1. Place recognition

2. Relocalisation



Modalités

Plusieurs capteurs disponibles :

- Caméra(s)
- Sonars
- Température
- Magnétomètre
- Pression
- USBL (Ultra Short Base Line)
- INS (Inertial Navigation System)
- DVL (Doppler Velocity Log)

- ✓ Vision
- ? Multimodal



Base de données Tour Eiffel (Ifremer)

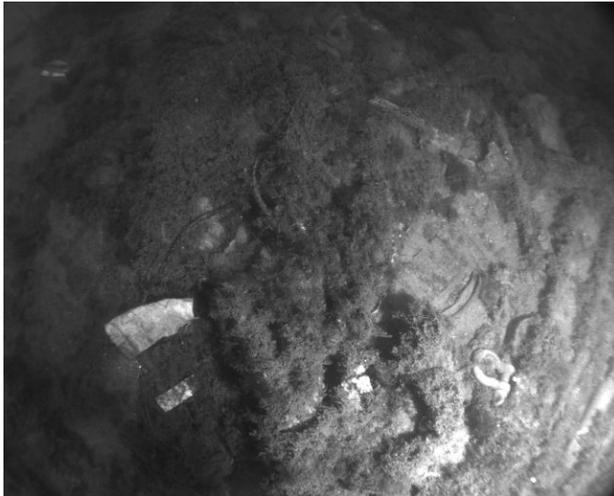
- 4 séquences étendues sur 6 ans
- Environ 4000 images par séquences
ex : 2015 : ~4800 images RGB 1920x1080px
2016 : ~3700 images RGB 1920x1080px



Base de données sous-marines

AQUALOC

- 3 différents sites
- Caméra monoculaire monochromatique
- IMU "low cost"
- Capteur de pression



Peu de datasets pour la relocalisation

- Nécessite plusieurs visites du même site
- Les données doivent être suffisantes pour faire une reconstruction 3D ou fournir une autre forme de vérité terrain
 - Quelle forme de vérité terrain ?

Relocalisation visuelle

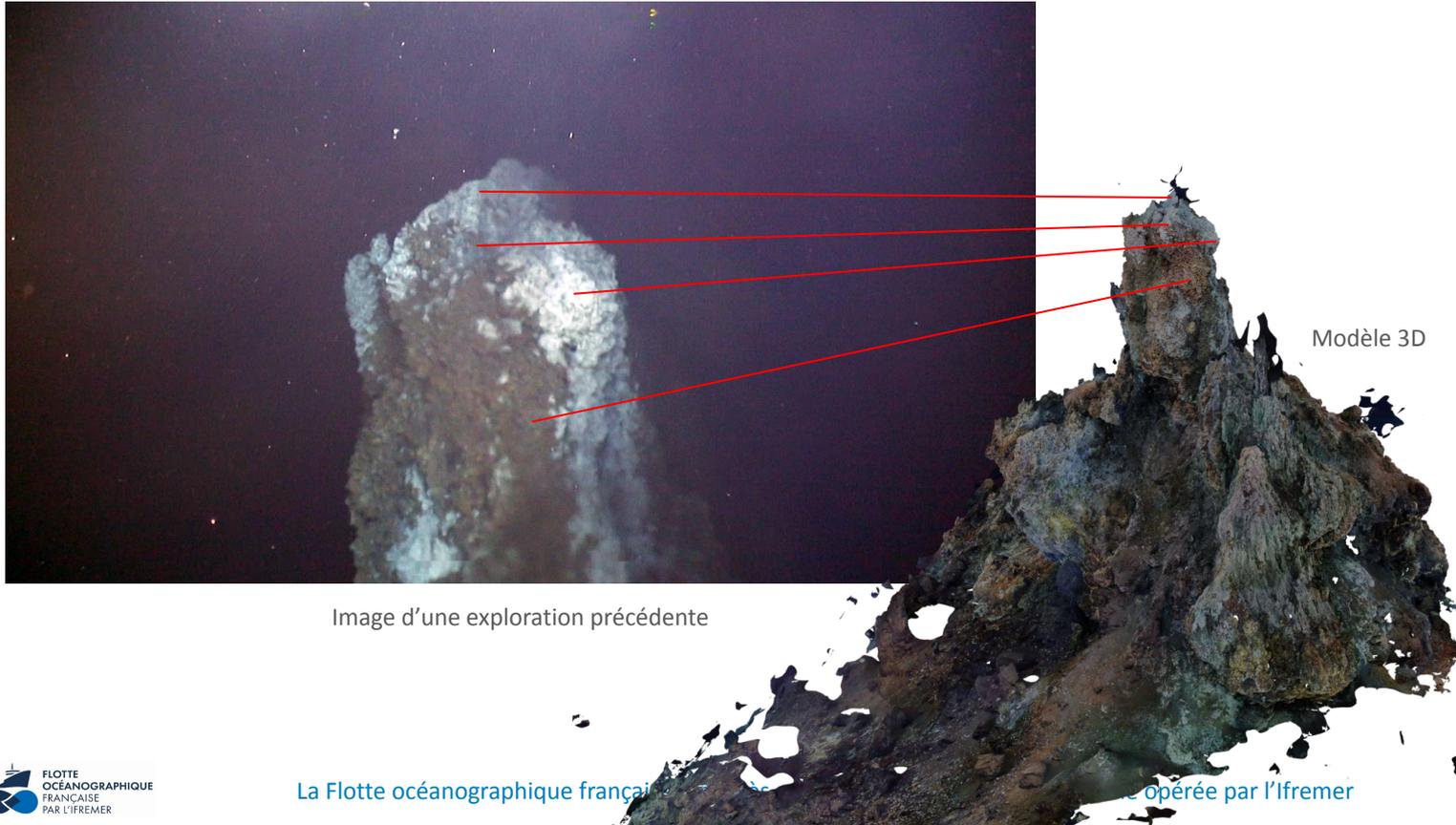
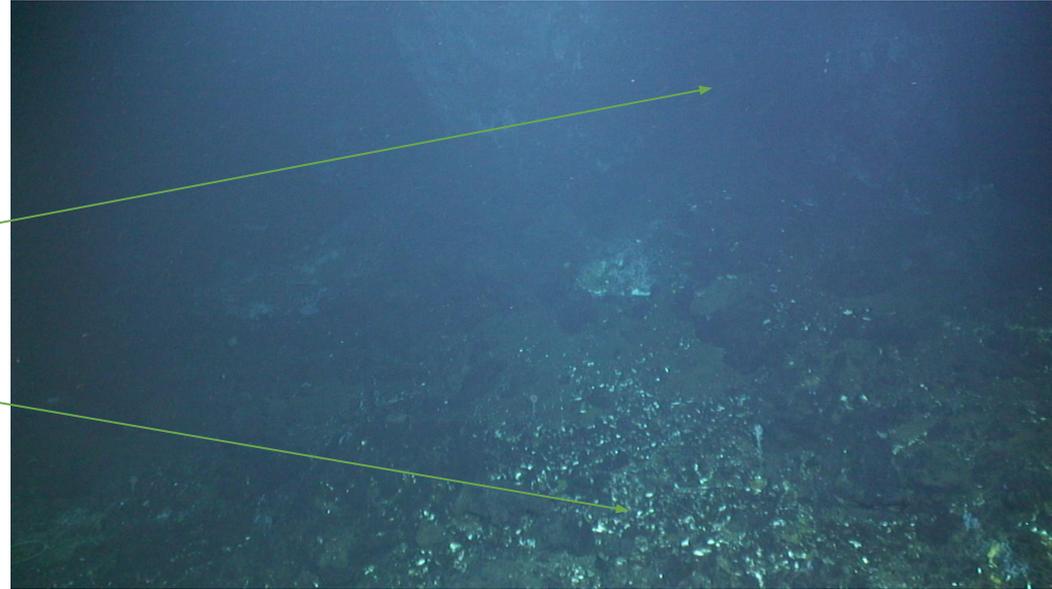


Image d'une exploration précédente

Modèle 3D

Difficultés en vision sous-marine

- Absorption de certaines longueurs d'ondes
- Plage de visibilité grandement réduite
- Aucune source de lumière en dehors de l'AUV
- Environnement soumis à des transformations
- Conditions variables (turbidité, salinité, ...)



Le deep learning en relocalisation visuelle

PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization

Alex Kendall

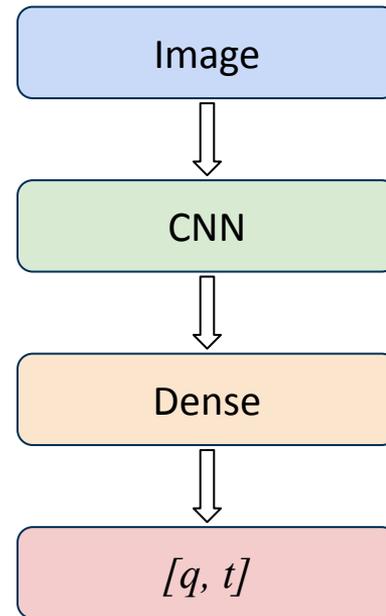
Matthew Grimes
University of Cambridge

Roberto Cipolla

agk34, mkg30, rc10001 @cam.ac.uk



Figure 1: **PoseNet: Convolutional neural network monocular camera relocalization.** Relocalization results for an input image (top), the predicted camera pose of a visual reconstruction (middle), shown again overlaid in red on the original image (bottom). Our system relocalizes to within approximately $2m$ and 6° for large outdoor scenes spanning $50,000m^2$. For an online demonstration, please see our project webpage: mi.eng.cam.ac.uk/projects/relocalisation/



Le deep learning en relocalisation visuelle

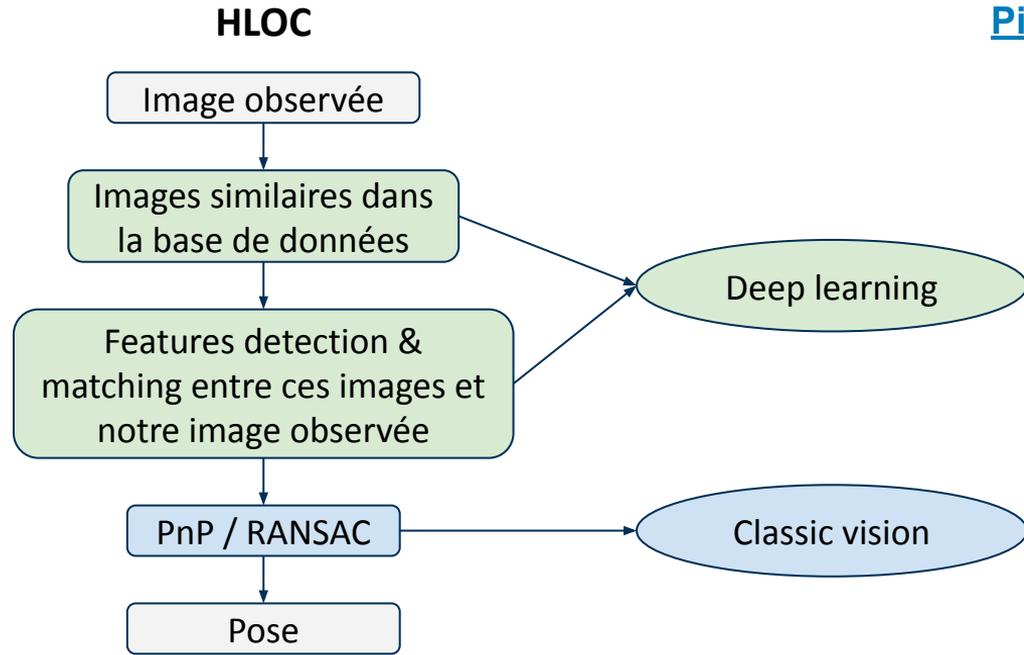


Images tirée du dataset introduit dans PoseNet

Le deep learning en relocalisation visuelle

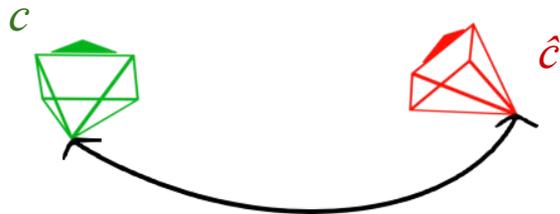
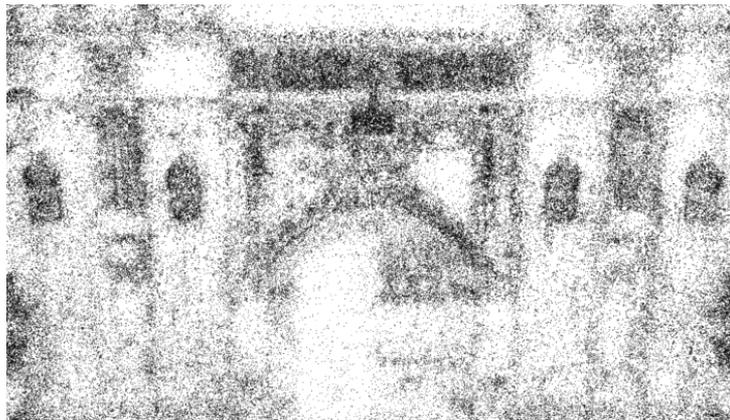


P-E Sarlin



PixLoc

Pose regression loss functions



$$L(c, \hat{c}) ?$$

Pose regression loss functions

- PoseNet [Kendall *et al.*]
- Visual Servoing [Bateux *et al.*]
- DSAC* [Brachmann *et al.*]
- DSAC [Brachmann *et al.*]
- Homoscedastic [Kendall *et al.*]
- Reprojection error [Kendall *et al.*]

$$\mathcal{L}_P = \|\hat{\mathbf{t}} - \mathbf{t}\|_2 + \beta \left\| \hat{\mathbf{q}} - \frac{\mathbf{q}}{\|\mathbf{q}\|} \right\|_2 \quad \text{Scene agnostic}$$

$$\text{loss}(\mathbf{I}) = \|\widehat{{}^{c_0}\mathbf{t}_c} - {}^{c_0}\mathbf{t}_c\|_2 + \beta \|\widehat{\theta\mathbf{u}} - \theta\mathbf{u}\|_2$$

$$\ell^{\text{Pose}}(\hat{\mathbf{h}}, \mathbf{h}^*) = \|\hat{\mathbf{t}} - \mathbf{t}^*\| + \gamma \angle(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*),$$

$$\ell_{\text{pose}}(\mathbf{h}, \mathbf{h}^*) = \max(\angle(\boldsymbol{\theta}, \boldsymbol{\theta}^*), \|\mathbf{t} - \mathbf{t}^*\|),$$

$$\mathcal{L}_\sigma(I) = \mathcal{L}_x(I) \exp(-\hat{s}_x) + \hat{s}_x + \mathcal{L}_q(I) \exp(-\hat{s}_q) + \hat{s}_q$$

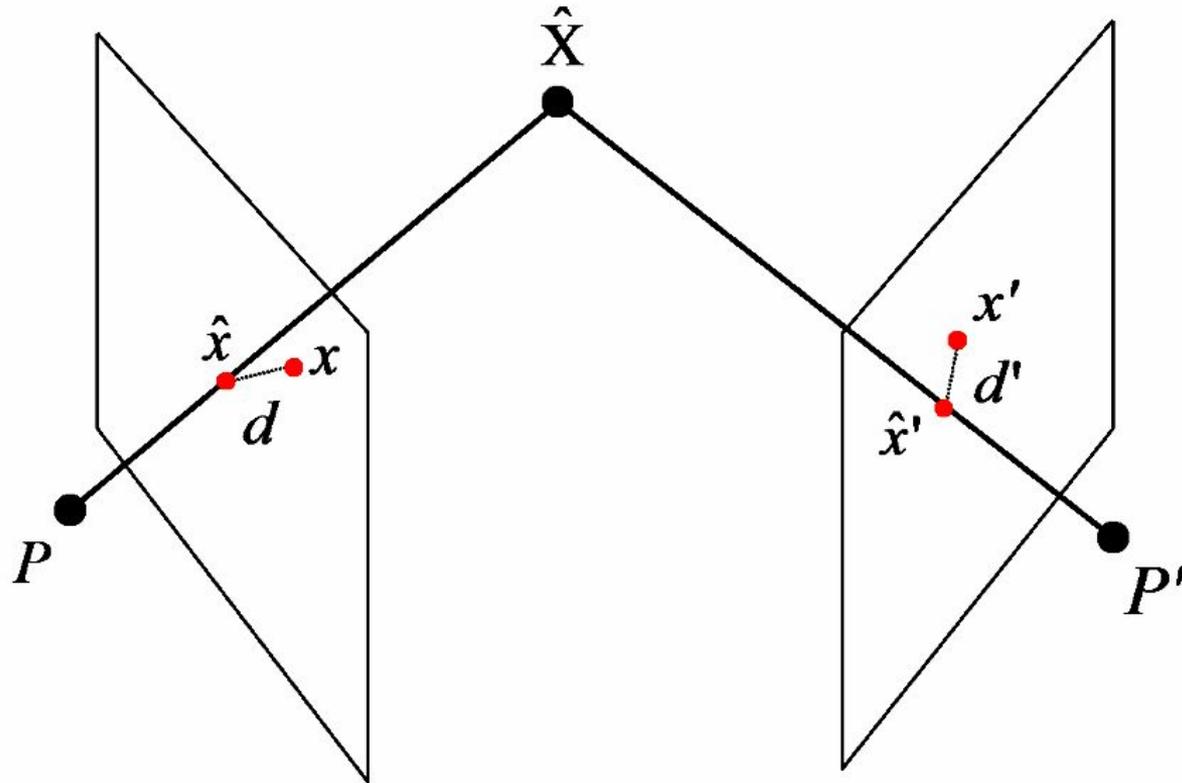
$$\mathcal{L}_G = \frac{1}{|\mathcal{G}|} \sum_{\mathbf{P}_i \in \mathcal{G}} \left\| \pi(\mathbf{t}, \mathbf{q}, \mathbf{P}_i) - \pi(\hat{\mathbf{t}}, \hat{\mathbf{q}}, \mathbf{P}_i) \right\|_1$$

Scene agnostic pose regression loss functions

PoseNet, DSAC, Bateux, DSAC*, homoscedastic :

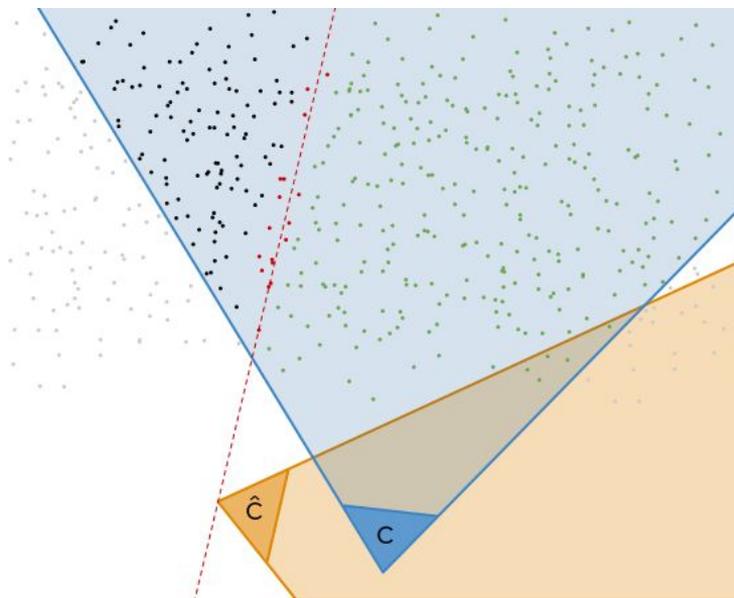
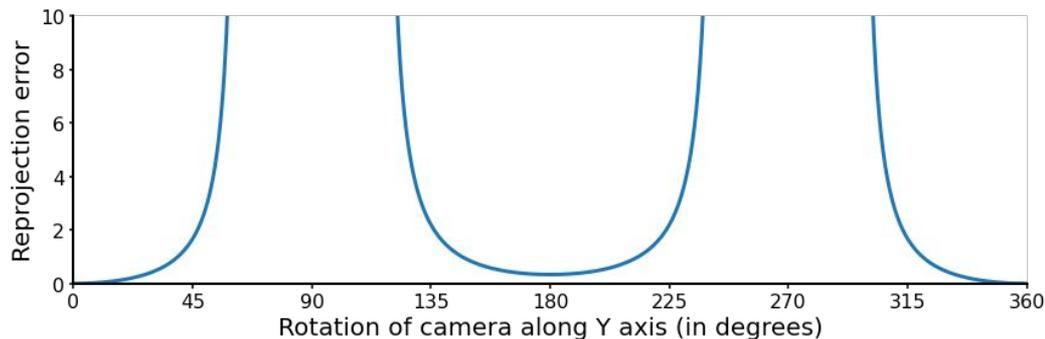
- Une pondération translation / rotation globale quelque soit l'observation
- Rotation et translation pas les mêmes unités ni ordre de grandeur dans $SE(3)$

Erreur de reprojection

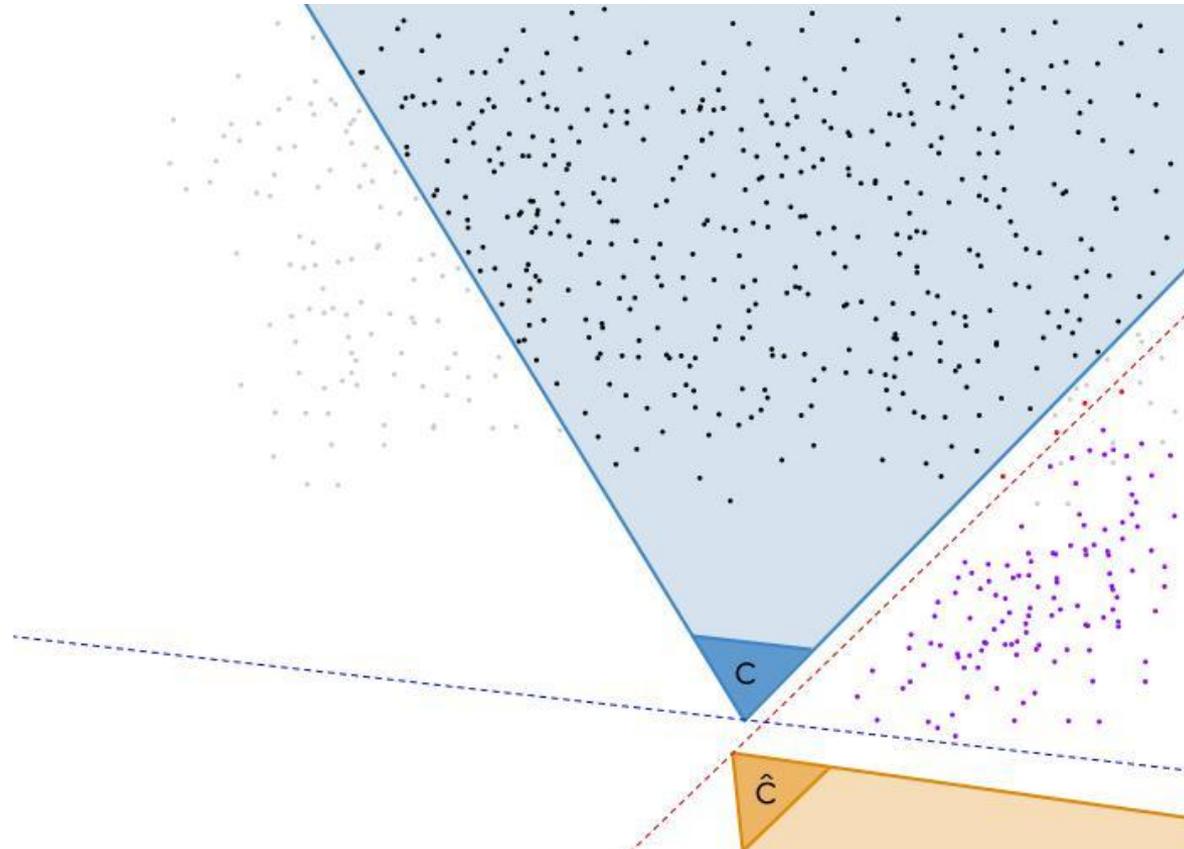


Erreur de reprojection

- Requiert une bonne initialisation des poses
- A un minimum local aussi large que le minimum global
- Points qui peuvent facilement se projeter à l'infini, gradient infini



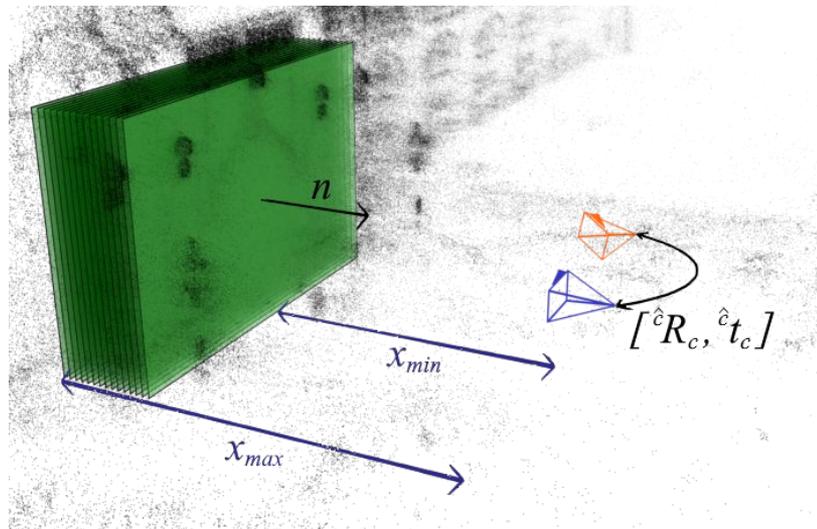
Erreur de reprojektion



Homography-based loss function

✗ Reprojection de points
d'une vue à l'autre

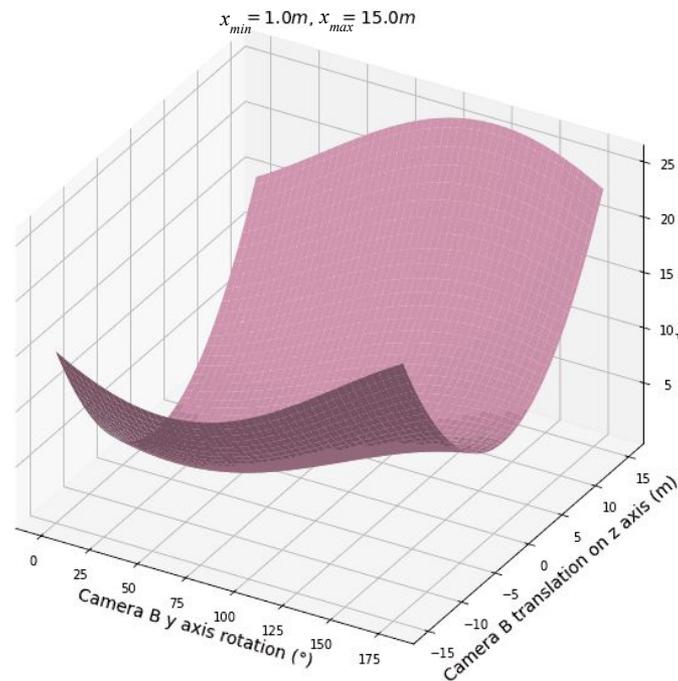
✓ Homographies entre les deux vues
caméras induites par des plans parallèle à
la vue de notre caméra vérité de terrain



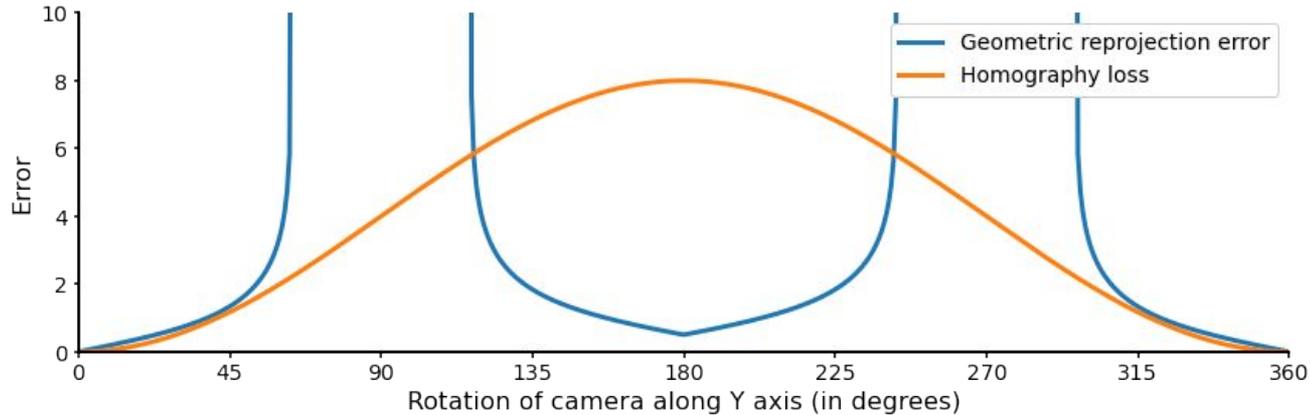
Homography-based loss function

Approxime l'erreur de reprojection

$$\mathcal{L}_H = \frac{1}{x_{max} - x_{min}} \int_{x_{min}}^{x_{max}} \|I - H\|_F^2 dx$$

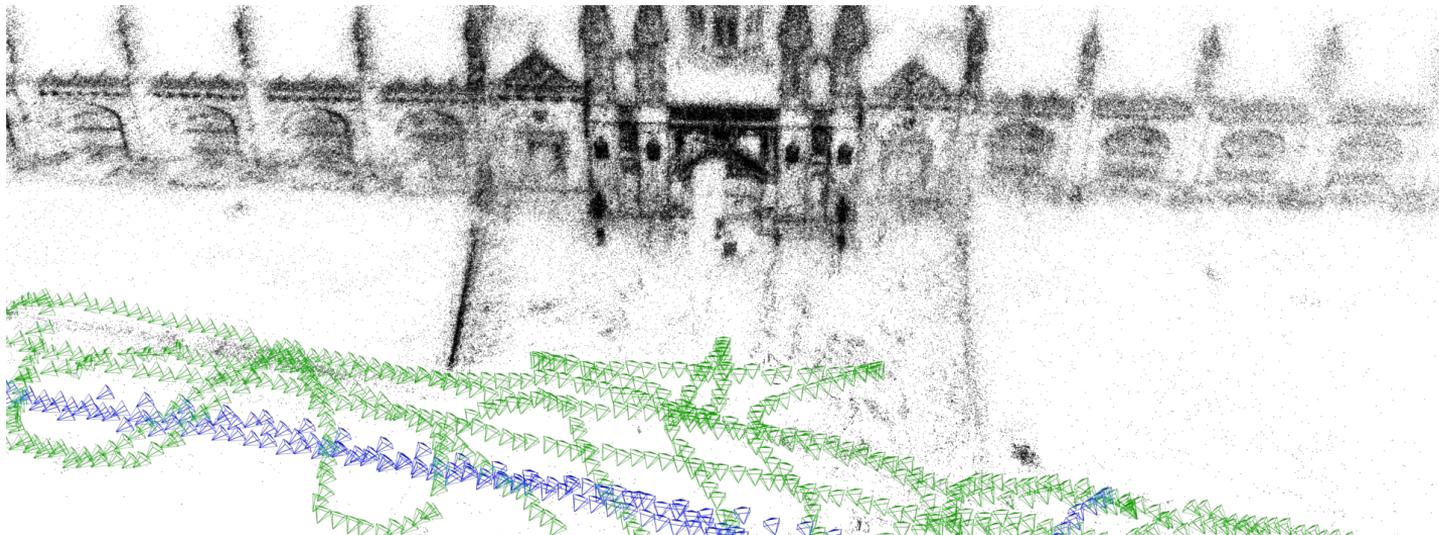


Homography-based loss function



Cambridge dataset

- 6 scènes
- Entre 5 et 15 séquences par scènes
- Entre 200 et 1000 images par séquence



Synthèse des résultats



Scene	PoseNet ($\beta = 500$)	Homoscedastic	MaxError	Geometric reprojection	Global homography	Local homography
Great Court	311, 4.77m , 2.51°	196 , 6.54m, 3.18°	7.92e+4, 9.22m, 29.0°	3.58e+4, 18.25m, 10.31°	3.81e+4, 11.7m, 8.20°	3.46e+4, 14.1m, 9.74°
King's College	75.7, 0.76m, 0.90°	41.8, 0.87m, 1.15°	6.74e+4, 1.21m, 6.33°	34.4, 0.64m , 0.89°	29.8, 0.80m, 0.97°	26.6 , 0.94m, 0.98°
Old Hospital	141, 2.26m, 3.28°	106, 1.97m, 3.20°	1.22e+4, 2.57m, 4.88°	72.3, 1.61m , 2.36°	76.7, 2.38m, 3.35°	51.6 , 1.79m, 2.80°
Shop Façade	603, 0.79m , 4.15°	195, 1.22m, 4.57°	4.90e+4, 1.46m, 7.12°	203, 1.14m, 4.36°	185, 1.37m, 4.58°	157 , 1.20m, 4.19°
St Mary's Church	443, 1.69m, 4.32°	308, 1.60m, 4.06°	1.52e+4, 1.62m, 9.06°	2112, 1.73m, 4.48°	198, 1.56m , 3.82°	174 , 1.70m, 3.90°
Street	3496, 23.2m, 29.4°	4060, 17.3m, 35.3°	8.51e+4, 16.1m, 43.6°	<i>diverges</i>	1437, 16.4m, 30.1°	1312 , 12.5m , 28.3°

Mean squared reprojection error (px), distance médiane entre les poses, distance angulaire médiane entre les poses

Distance entre les poses : $\left\| \hat{t} - t \right\|_2$

Distance angulaire : θ de (u, θ)

Travaux futurs

- Benchmark des méthodes de relocalisation en sous-marin
- Deep underwater features
- Multi-modal ?

Merci pour votre attention !