

Localisation end-to-end sur un terrain de football à l'aide de reconstruction 3D et d'apprentissage supervisé

Équipe Rhoban - Contact: Grégoire Passault - gregoire.passault@u-bordeaux.fr

Contexte

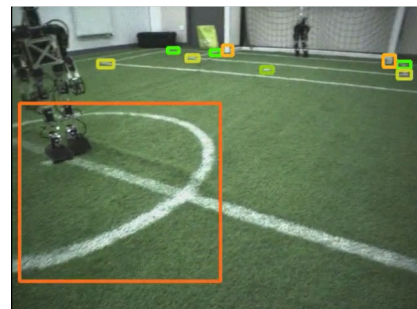
Rhoban est une équipe de robotique de l'université de Bordeaux qui développe des **robots humanoïdes de petite taille** (plateforme Sigmaban) et participe à la compétition de robotique RoboCup depuis 10 ans dans la ligue "humanoid soccer kid-size"¹. Dans cette dernière, des équipes robots humanoïdes doivent **jouer au football en totale autonomie sur un mini-terrain**. L'équipe rhoban est 5 fois championne du monde.



Localisation du robot sur le terrain

Aujourd'hui, les robots se localisent sur le terrain en détectant des points d'intérêt ("*features*") à l'aide d'un modèle de détection d'objets. Les bases des poteaux des cages, le cercle central, les intersections de lignes ("L" et "T"), et les marques de penalty sont ainsi repérés.

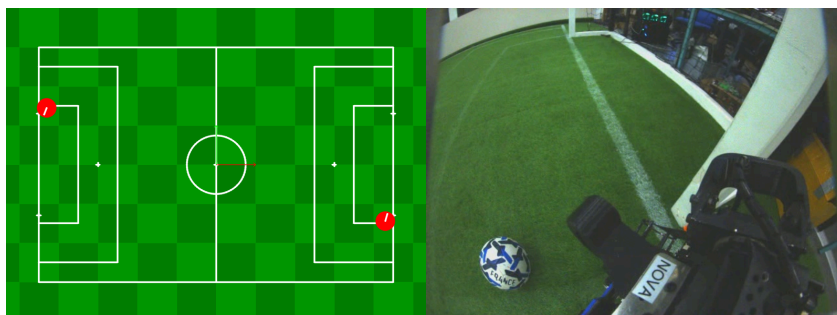
Ces éléments sont ensuite exploités à l'aide d'un filtre particulaire afin de localiser le robot sur le terrain.



Cette approche produit des estimations satisfaisantes, mais nécessite l'annotation manuelle de nombreuses images afin d'obtenir une vérité terrain.

Localisation *end-to-end*

La question posée ici est la suivante: *peut-on produire une estimation de la position directement à partir de l'image, au lieu de détecter des points d'intérêt* ? On obtiendrait alors une approche *end-to-end* ("de bout en bout")



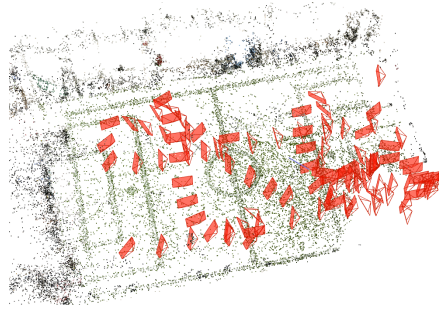
Exemple d'une des image du jeu de données (à droite), et de la vérité terrain associée (à gauche)

¹ <https://humanoid.robocup.org/>

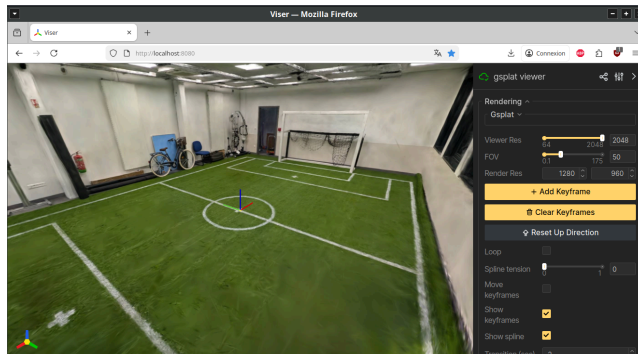
Approche proposée

L'approche proposée est la suivante:

1. **Obtenir une vidéo ou d'un ensemble de photos réalisées sur le terrain.**
Des fichiers seront fournis pour le projet.
2. **Estimer la position des prises de vue**
Ce problème est très bien résolu par des logiciels standard tels que COLMAP²



3. **Créer une reconstruction 3D photoréaliste**
Cette étape peut être résolue à l'aide de *gaussian splatting*, avec par exemple le projet *gsplat*³.



4. **Produire des images dont la pose caméra est connue**
Cette reconstruction 3D permettra alors de produire des images dont la position/orientation de la caméra seront connues. Cela permet de créer un *dataset*.
5. **Entraîner un modèle supervisé**
Il sera alors possible d'entraîner un modèle de prédiction de la position/orientation de la caméra en supervisé à l'aide de réseaux convolutionnels classiques.

L'équipe Rhoban peut fournir des fichiers (photos/vidéos) de terrains, ainsi que du code fonctionnel jusqu'à la partie 4, permettant aux étudiants du projet de se concentrer sur la partie "apprentissage supervisé". Davantage de reconstructions seront néanmoins sans doute nécessaires pour enrichir le jeu de données.

L'objectif étant de comparer plusieurs modèles de prédiction (architecture de réseau, fonction de coût, régularisation etc.).

² <https://colmap.github.io/>

³ <https://github.com/nerfstudio-project/gspat>