

Interaction naturelle en Réalité Augmentée mobile basée sur le Deep Learning

Contexte

La Réalité Augmentée (RA) est une expérience immersive et interactive qui requiert un moteur logiciel ou bien un SDK qui alimente l'application et rationalise le flux de donnée tout en optimisant les processus parallèles en temps réel. Un SDK de RA comporte des fonctions de reconnaissance, de rendu et de suivi pour permettre de créer des environnements hybrides et de développer divers scénarios applicatifs.

ARCore est un SDK développé par Google, qui a pour but de concevoir et déployer des applications de RA sur plateformes mobiles. ARCore réalise essentiellement les deux principes de la RA qui sont : la localisation temps réel et le recalage virtuel, c'est un SDK gratuit et disponible pour Android.

Afin d'unifier le développement des applications RA sur plateformes mobiles, Unity a créé une API commune, appelée AR Foundation, qui prend en charge plusieurs SDK de RA pour faciliter l'implémentation et le déploiement d'applications multiplateformes. En effet, AR Foundation comprend les principales fonctionnalités de ARKit (iOS), ARCore, de Magic Leap et de HoloLens, ainsi que des fonctionnalités de Unity pour créer des applications fiables et déployables sur différents OS et plateformes. Ce framework permet de tirer parti de toutes ces propriétés logicielles dans un flux de production unifié et offre les caractéristiques suivantes :

- Détecter des surfaces planes et construire des maillages de représentation temps réel.
- Utiliser des points de référence individuels pour générer des plans.
- Estimer et définir l'éclairage pour adapter les traitements et le rendu en conséquence.
- Raycasting envers des plans et des objets virtuels.

AR Foundation fonctionne avec Android et iOS avec un pipeline de construction et de test intuitive et facile à mettre en production. AR Foundation est basé sur Unity et permet de créer et d'exécuter la simulation virtuelle sur des plateformes mobiles en s'affranchissant des contraintes de configurations.

Dans les systèmes de RA, la détection d'objet est une étape fondamentale pour caractériser les cibles visuelles sur lesquelles il faudra ajouter les contenus virtuels. Cette tâche est complexe et nécessite une approche robuste et performante afin de caractériser les marqueurs visuels. L'apprentissage profond est utilisé pour la reconnaissance d'images en raison de sa capacité à apprendre et à extraire automatiquement des caractéristiques sémantiques des images.

Les modèles de deep learning ont des capacités de généralisation et peuvent apprendre à reconnaître des objets et des motifs complexes dans un environnement pouvant présenter des variations de conditions d'expérimentation.

Enfin, lorsque l'objet d'intérêt est identifié, la résolution de localisation 3D et le recalage virtuel permettent de placer l'entité de synthèse sur l'image afin d'augmenter et enrichir la perception de la scène réelle.

Objectifs du stage

Le système de RA doit permettre de détecter des images d'intérêt en temps réel et d'augmenter la scène avec des modèles 3D d'objets sur marqueur détectés pour vivre une expérience originale d'un mixage réel-virtuel.

La détection d'objet doit se faire à l'aide d'un apprentissage en deep learning pour identifier et localiser l'objet d'intérêt avec précision. La collecte de données, l'entraînement des modèles ainsi que paramétrage doivent être réalisés de manière judicieuse pour un résultat optimal.

L'objectif sera de superposer des modèles virtuels dynamiques et interactif pour expliquer et enrichir la perception à l'aide du réalisme et la qualité de la 3D. Ces objets 3D sont ajoutés à un support visuel 2D qui est un ensemble d'images de monuments pour une expérience touristique et culturelle.

L'application devra fournir une meilleure sensation de visualisation et d'interaction. La scène réelle est enrichie par des graphiques 3D qui seront manipulés par les gestes tactiles de la main dans un paradigme d'immersion et d'interaction inné aux systèmes de RA.

Étapes

- 1) Veille technologique sur les dataset et les techniques de scraping d'images
- 2) Revue documentaire sur les modèles de détection d'images.
- 3) Proposition, implémentation et optimisation du modèle de reconnaissance.
- 4) Augmentation virtuelle et interaction 3D naturelle.
- 5) Test et validation du système en conditions réelles de l'expérience-utilisateur.

Langage et plateformes

C#, Unity, ARCore, AR Foundation, Python, TensorFlow, PyTorch

Profil recherché et compétences requises

- Étudiant(e) en Master 2 Recherche dans le domaine Informatique.
- Formation en Computer Graphics (transformations projectives, modelview, shaders, mesh) et en Computer Vision (détection d'objet, analyse vidéo, tracking, géométrie multi-vues)
- Bonnes connaissances en Deep Learning (apprentissage/classification, ANN, CNN, Transformers, etc). Connaissances des frameworks TensorFlow et PyTorch.
- Compétences en programmation C#, Python.
- Compétences en conception et développement avec Unity.
- Motivation et intérêt pour la recherche et le développement.

Contacts

- Madjid MAIDI, ESME/IBISC, madjid.maidi@esme.fr
- Samir OTMANE, IBISC, Université d'Evry, samir.otmane@ibisc.univ-evry.fr

Références

- [1] J. Redmon, S. Kumar Divvala, R. B. Girshick and A. Farhadi. « You Only Look Once: Unified, Real-Time Object Detection ». IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016
- [2] S. Ren, K. He, R. B. Girshick and J. Sun. « Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks ». IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, N° 6, pp. 1137-1149, 2017
- [3] R. B. Girshick, J. Donahue, T. Darrell and J. Malik. « Rich Feature Hiérarchies for Accurate Object Detection and Semantic Segmentation ». IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587, 2014
- [4] R. Girshick. « Fast R-CNN », IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015
- [5] J. Guo, P. Chen, Y. Jiang, H. Yokoi and S. Togo. « Real-time Object Detection with Deep Learning for Robot Vision on Mixed Reality Device ». IEEE Global Conference on Life Sciences and Technologies (LifeTech), pp. 82-83, 2021
- [6] Unity Technologies, AR Foundation, <https://unity.com/fr/unity/features/arfoundation>, [Online; accessed: November, 2022].
- [7] Google, ARCore, <https://developers.google.com/ar>, [Online; accessed: November, 2022]
- [8] Apple, ARKit, <https://developer.apple.com/augmented-reality>, [Online; accessed: November, 2022]
- [9] PTC, Vuforia, <https://developer.vuforia.com>, [Online; accessed: November, 2022]
- [10] A. Samini, K. L. Palmerius, and P. Ljung, "A review of current, complete augmented reality solutions", 2021 International Conference on Cyberworlds (CW), 2021, pp. 49–56