

Navigation Fluide Unidirectionnelle

Interpolation de Vues Street View par Reprojection Planaire

Proposition de Projet de Recherche

Maxim Quénel Pierre Louis Brun Mame Alé Seye Denis Bereziuc

30 Novembre 2025

1 Introduction et Motivation

Les systèmes de navigation virtuelle actuels, illustrés par des plateformes comme *Google Street View*, fonctionnent principalement par transitions discrètes entre des panoramas statiques. Cette discontinuité crée une rupture visuelle qui altère l'immersion et la perception cinétique du déplacement.

L'objectif de cette recherche est de développer un système de **Synthèse de Vues** (*Novel View Synthesis*) capable de générer un flux vidéo continu, simulant un avancement fluide entre deux points géographiques distants.

Pour cette phase initiale, le périmètre est restreint à une navigation unidirectionnelle (vue frontale alignée sur l'axe de la route). Cette approche, comparable à un effet de caméra embarquée, permet d'isoler la problématique de la reconstruction 3D en s'affranchissant des distorsions complexes propres aux projections sphériques à 360°.

2 Définition du Problème

Le défi central réside dans la génération de trames intermédiaires photo-réalistes entre deux images sources séparées par une distance significative.

En limitant le champ de vision à une perspective frontale fixe, nous simplifions la géométrie du problème. Cependant, trois défis techniques majeurs doivent être résolus :

- **Estimation de Profondeur Monoculaire :**
Inférer la structure géométrique 3D de la scène à partir d'une unique information RGB.
- **Gestion des Occlusions et Disocclusions :**
Gérer l'apparition de zones précédemment cachées par des objets au premier plan lors du mouvement virtuel de la caméra.
- **Cohérence Temporelle :**
Garantir une transition imperceptible et géométriquement valide entre la projection de l'image de départ et celle de l'image d'arrivée.

3 Dataset et Acquisition

Le système s'appuie sur une génération de données dynamiques via l'API Google Maps. Contrairement aux méthodes exploitant des panoramas complets, nous privilégions l'extraction de vues rectilinéaires standard.

Le processus interroge les métadonnées pour localiser le nœud suivant, puis calcule le cap précis pour aligner l'axe optique de la caméra virtuelle avec la direction du vecteur de déplacement. Cette contrainte d'alignement assure que le centre optique de l'image cible coïncide avec le point de fuite de l'image source.

4 Méthodologie Générale

L'architecture proposée repose sur une reprojection 3D explicite, rendue possible par la contrainte de mouvement unidirectionnel. La pipeline de traitement se décompose en quatre étapes clés :

4.1 Estimation Dense de la Profondeur

Au lieu de s'appuyer sur des capteurs actifs (LiDAR), nous utilisons des modèles d'apprentissage profond de pointe spécialisés dans l'estimation de profondeur monoculaire relative ou métrique. L'objectif est d'obtenir une *depth map* dense et précise pour chaque image frontale RGB.

4.2 Warping 3D et Reprojection Géométrique

Cette étape transforme les données 2D en structure 3D exploitable :

- **Back-projection** : Conversion des pixels de l'image source en un nuage de points 3D, en utilisant la carte de profondeur et les paramètres intrinsèques de la caméra.
- **Transformation rigide** : Application d'une translation sur l'axe de profondeur Z pour simuler le mouvement de la caméra.
- **Reprojection** : Projection des points 3D transformés sur le nouveau plan image cible.

4.3 Interpolation et Fusion Bidirectionnelle

Pour pallier les artefacts géométriques et les trous causés par le mouvement avant, nous adoptons une stratégie de synthèse bidirectionnelle :

- Génération d'une vue "**Forward**" (projection de l'image $A \rightarrow B$).
- Génération d'une vue "**Backward**" (rétro-projection de l'image $B \rightarrow A$).

Les deux vues synthétisées sont ensuite fusionnées via une pondération linéaire ou une carte de confiance pour assurer une transition fluide.

4.4 Post-traitement et Inpainting

Enfin, pour corriger les artefacts résiduels (notamment sur les bordures de l'image ou les zones de forte disocclusion), des algorithmes de remplissage d'image (*Inpainting*) sont appliqués. Ces méthodes, qu'elles soient basées sur des techniques de diffusion de pixels ou des réseaux de neurones génératifs, permettent de reconstruire la texture manquante de manière cohérente avec le contexte visuel.