Fluid Unidirectional Navigation

Street View Interpolation via Planar Reprojection

Research Project Proposal

Maxim Quénel Pierre Louis Brun Mame Alé Seye Denis Bereziuc

November 30, 2025

1 Introduction and Motivation

Current virtual navigation systems, exemplified by platforms like *Google Street View*, operate primarily through discrete transitions between static panoramas. This discontinuity creates a visual break that impairs immersion and the kinetic perception of movement.

The objective of this research is to develop a **Novel View Synthesis** system capable of generating a continuous video stream, simulating fluid advancement between two distant geographic points.

For this initial phase, the scope is restricted to unidirectional navigation (frontal view aligned with the road axis). This approach, comparable to a dashcam effect, allows for isolating the 3D reconstruction problem by avoiding the complex distortions inherent in spherical 360° projections.

2 Problem Definition

The central challenge lies in generating photo-realistic intermediate frames between two source images separated by a significant distance.

By limiting the field of view to a fixed frontal perspective, we simplify the geometry of the problem. However, three major technical challenges must be solved:

- **Monocular Depth Estimation:**
  Inferring the 3D geometric structure of the scene from unique RGB information.
- **Occlusion and Disocclusion Management:**
  Handling the appearance of zones previously hidden by foreground objects during the virtual movement of the camera.
- **Temporal Consistency:**
  Guaranteeing an imperceptible and geometrically valid transition between the projection of the starting image and that of the arrival image.

3 Dataset and Acquisition

The system relies on dynamic data generation via the Google Maps API. Unlike methods exploiting full panoramas, we prioritize the extraction of standard rectilinear views.

The process queries metadata to locate the next node, then calculates the precise heading to align the virtual camera's optical axis with the direction of the displacement vector. This alignment constraint ensures that the optical center of the target image coincides with the vanishing point of the source image.

4 General Methodology

The proposed architecture is based on explicit 3D reprojection, made possible by the unidirectional movement constraint. The processing pipeline is broken down into four key steps:

4.1 Dense Depth Estimation

Instead of relying on active sensors (LiDAR), we utilize state-of-the-art deep learning models specialized in relative or metric monocular depth estimation. The goal is to obtain a dense and precise *depth map* for each frontal RGB image.

4.2 3D Warping and Geometric Reprojection

This step transforms 2D data into a usable 3D structure:

- **Back-projection:** Conversion of source image pixels into a 3D point cloud, using the depth map and the camera's intrinsic parameters.
- **Rigid Transformation:** Application of a translation along the depth axis Z to simulate camera movement.
- **Reprojection:** Projection of the transformed 3D points onto the new target image plane.

4.3 Interpolation and Bidirectional Fusion

To mitigate geometric artifacts and holes caused by forward movement, we adopt a bidirectional synthesis strategy:

- Generation of a **"Forward"** view (projection of image $A \to B$).
- Generation of a **"Backward"** view (retro-projection of image $B \to A$).

The two synthesized views are then merged via linear weighting or a confidence map to ensure a fluid transition.

4.4 Post-processing and Inpainting

Finally, to correct residual artifacts (particularly on image borders or areas of strong disocclusion), image filling algorithms (*Inpainting*) are applied. These methods, whether based on pixel diffusion techniques or generative neural networks, allow for the reconstruction of missing texture in a manner consistent with the visual context.