AR Journey Into Movies

Mixed Reality Project Proposal
Supervised by: Xudong Jiang
September 27, 2025

GROUP MEMBERS

Li Wa Tang Zhihao Cai Yiwei Pang Junyi Huang

I. DESCRIPTION OF THE PROJECT

The project ”AR journey into movies” aims to develop an AR application that brings movie scenes into their original filming locations through precise visual localization and pose alignment. As input, the system will use multi-angle photographs taken at the film’s shooting locations, original movie frames, and real-time camera footage from AR-enabled smartphones. Using Hierarchical Localization techniques [1]–[3], the application will align the movie scene with the user’s real-world perspective. The expected output will be an immersive AR visualization that overlays movie imagery onto the live environment, allowing users to explore and experience movies scenes directly within their real surroundings.

II. WORK PACKAGES AND TIMELINE

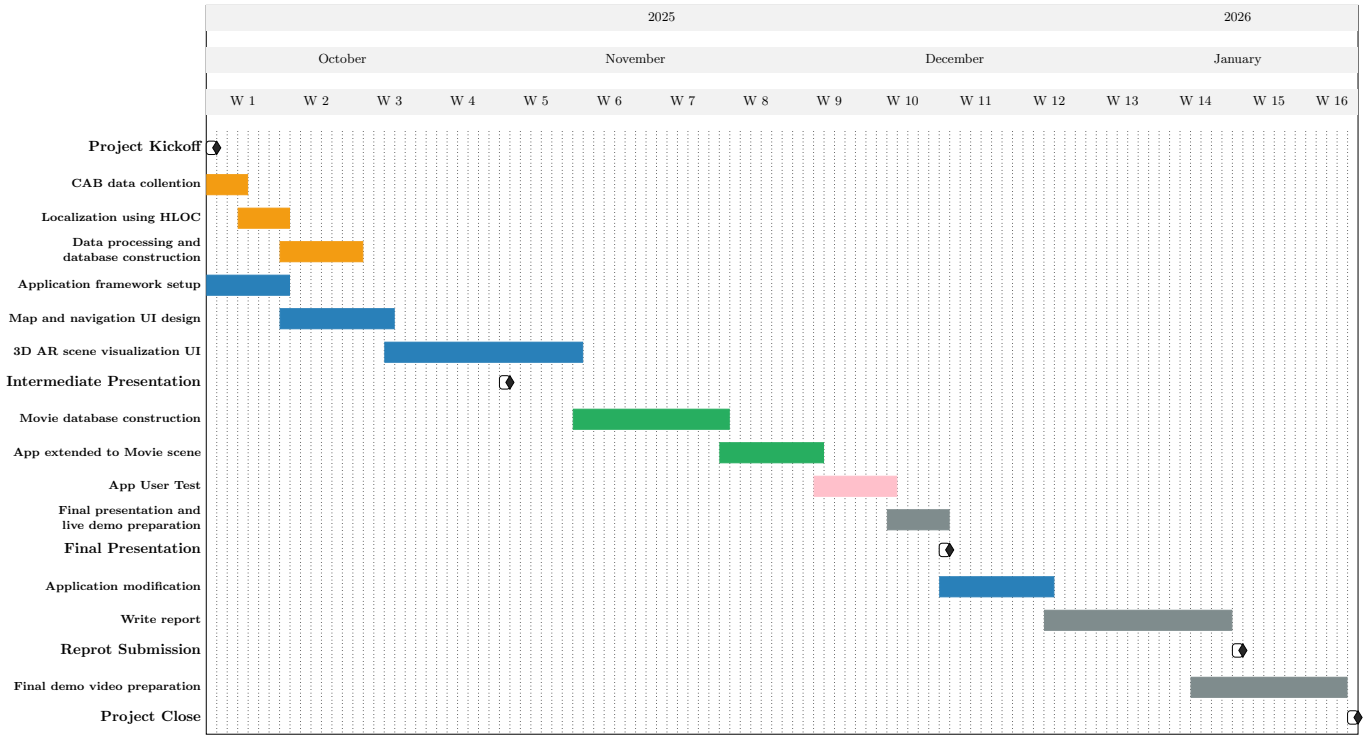

Fig. 1: Timeline of Work Packages

CAB demo stage:
Offline part (Responsible: Junyi Huang, Yiwei Pang):

- Data Collection:
  Capture 10-20 images for the mapping database and several query images from several directions of the main entrance of the CAB building of ETHZ, with an Android phone that supports Google ARCore. Then, using the ARCore framework (C/C++) functions to acquire the accurate camera intrinsics on PC.
  Outcomes: demo image dataset and camera intrinsics.

- Hierarchical-Localisation (HLOC) Based Localisation Pipeline [2] Construction:
  Apply the HLOC pipeline, including reference SfM reconstruction (3D mapping with mapping database images) and query localisation (with query images). Feature extraction using superpoint_aachen [1], image pairs matching using SuperGlue [3] (Python).
  Outcomes: the camera position in the SfM world coordinate system and orientation.
- Data processing and Database construction:
  Aligh SfM coordinates with real-world geographic coordinates (WGS84) by using EXIF information of multiple mapping images in the demo database (Python). Then, package the query images'poses, and 3D point geographic coordinates into JSON for later AR overlay and UI.
  Outcomes: query camera's WGS84 coordinate and pose in JSON file.

User Interface (Responsible: Li Wa Tang, Zhihao Cai):

- Visual Guidance and Fine Localization
  On the Android client, developed with Unity (C#), we will render the target camera frustum in the AR world to guide the user towards the correct filming viewpoint. Additional visual cues inspired by Google Maps Live View will be provided, including directional arrows, textual hints (e.g., "Move left 2m" / "Turn 15°"), and optional polylines to indicate the path. Once the user's device pose approaches the target frustum pose, the system will switch into a fine localization mode. In this stage, the selected movie frame will be overlaid as a semi-transparent static image on the live camera feed, allowing the user to manually refine the alignment.
  Outcomes: intuitive movement guidance, stable overlay aligned with the scene, and an immersive AR experience combining automatic guidance and user refinement.

Movie scene stage: Then we will extend the database and mapping pipeline to real-world movie filming locations for future experiments and demonstrations.

III. Outcomes and Demonstration

The expected outcome of our project is an immersive AR application that allows users to experience actual movie scenes in their phones. More specifically, our outcome will contain the followint four parts:

1) Utilize pre-built 3D models based on HLOC technology to align them with the real-world coordinate system.
2) Estimate the camera pose based on input movie scene images and align them with the corresponding real-world scenes.
3) Render the selected movie frames as synchronously as possible onto the real-world view.
4) Provide an immersive AR experience by guiding users to precise locations via tetrahedral prompts, then transitioning seamlessly to 3D scenes. This enables users to interact with scenes on their mobile devices while moving around the real-world location.

To better evaluate the performance of our project, we designed the following experiments:

1) Evaluate the accuracy of alignment by performing some necessary error analyses(e.g. the alignment error between the projected movie frame and the real-world scene and the pose difference between ARCore's live tracking and HLOC's computed camera position).
2) If feasible, seek participants to test the application at one or more film location and collect feedback on immersion and usability.

At the end of the semester, we plan to present our project through a pre-recorded video demonstrating the application running at a real location of a movie scene(e.g., users raise their phones toward the scene, and the movie content is overlaid in AR).

References

[1] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description.
[2] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From Coarse to Fine: Robust Hierarchical Localization at Large Scale.
[3] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning Feature Matching with Graph Neural Networks.