Project Proposal

1. **Team Members:**
    - Kishore Shenoy
    - Rajeshwari Devaramani
2. **Problem Definition:** What are you trying to solve or do? Why is this a problem that is useful or interesting to study or experiment with? (2-3 bullet points)
    - Generate Stereoscopic views from single camera video.
    - The work will be useful to generate 3D footage for movies while allowing artists to use gear of their choice.

3. **Project Idea:** What do you plan on accomplishing in your project scope? How will your idea solve the problem proposed in your problem definition? (2-3 bullet points)
    - Since depth data is inherently present in a motion video, why can't we use that instead of needing stereo-camera to do that?
    - Structure from Motion >> 3D Reconstruction (triangulation/learning based methods) >> Stereoscopic view.
    - The plan is to implement a non-learning-based 3D reconstruction, but nothing set in sand. We want it to be physically based. From there, 3D points to stereo based on camera matrix (multiplication).
    - Although there might be limitations based on the type of motion, we can still use methods like GenAI on the output to “filter” out errors or fill up “holes”.

4. **Related Works:** Briefly describe what work has been done in this area and how they currently approach the problem. List any research papers, existing codebases, or miscellaneous articles/websites that you found that relate to your work. (minimum 5 references; at least 3 should be research papers)
    - Researchers have devised algorithms to estimate depth from single-camera videos using methods like monocular depth estimation networks, structure from motion, and optical flow. They've also utilized deep learning for both depth estimation and view synthesis. Additionally, they've employed Structure from Motion techniques along with 3D reconstruction methods to reconstruct scene geometry and camera poses from single-camera videos.
    - https://arxiv.org/abs/1806.07381v1
    - https://www.cs.jhu.edu/~misha/ReadingSeminar/Papers/Triggs00.pdf
    - https://arxiv.org/pdf/2306.06360.pdf
    - https://www.mdpi.com/2076-3417/10/9/3101
    - https://www.mdpi.com/2072-4292/15/11/2740

5. **Datasets:** How will you benchmark your approach or theory? Describe and link each dataset you will use in your project, and explain why you selected the dataset. (min 1 dataset)
    - For benchmarking the approach of generating stereoscopic views from single-camera videos, we will use controlled datasets that provide ground truth camera coordinates and scene information.
    - **https://github.com/ZheC/GTA-IM-Dataset**
    - The GTA-IM Dataset is suitable for our approach because it provides ground truth camera coordinates, which are essential for evaluating the accuracy of 3D reconstruction and view synthesis. The synthetic nature of the dataset allows for precise control over scene complexity, motion patterns, and camera parameters, facilitating systematic evaluation and comparison of different algorithms

6. **Proposed Experiments:** What experiments will your team perform in your project? Roughly describe what ideas you will try. (2-3 bullet points)
    - First, we will try to perform 3D reconstruction using motion + camera location.
    - We will attempt to create (stereo) images from (point, color) 3D data.
    - Also, try to search across the video for visual data of features from the required viewport in order to extrapolate/generate the stereoscopic view.
    - If both of the above points don't work, we'd have to create polygons to render the simulated view as seen by the other eye.
    - In order to get a clean data, we will attempt to source it from games, say Grand Theft Auto, where we can obtain actual spatial data, camera coordinates and polygons. This should also aid in training if we choose to go down that path.

7. **Compute Resources:** What compute resources (GPUs, CPUs, TPUs) does your team plan to use? The type of compute resources can limit the scope of your project, so we want teams to pick appropriate projects for the resources they have accessible.
    - Some kind of a GPU that can handle training an ML model.
    - A CUDA process to perform fast matrix operations.