The project involves various phases to process and analyze video data. Below is a brief explanation of the four phases that currently exist:

1. Recording: In this phase, video data is recorded using two cameras (cap1 and cap2) capturing consecutive frames. The video files are read using OpenCV, and the frames are stored in **frame1** and **frame2**. The frames are then converted to grayscale for further processing. The ORB (Oriented FAST and Rotated BRIEF) feature detector is used to detect keypoints and compute descriptors for both frames.
2. Identification of Matches between Frames: ORB features in the two consecutive frames are matched to identify corresponding points between the frames. The feature matching is performed using the Brute-Force Matcher (**cv2.BFMatcher**) with the Hamming distance as the distance measure. The matched keypoints and descriptors are stored in **matches**.
3. Building a Single Cloud Based on Adjustments and Disparity Map Calculation: The function **depth\_from\_h264\_vectors** estimates the depth of each 3D point represented by a vector (x1, y1, x2, y2) between its 2D location in two consecutive frames. This depth estimation is based on the known height difference. The estimated depths are then used to build a 3D point cloud, and the function **topdown\_view** is used to generate a top-down view representation of the depth map.
4. Connecting Clouds: The phase of connecting clouds seems to be partially commented out and requires further implementation. It includes functions like **triangulate\_points**, which triangulates 2D keypoints from two frames to obtain 3D points.

Overall, the project appears to involve processing video data from two cameras, detecting and matching features, estimating depth, and constructing a 3D point cloud from the matched keypoints.