Thermal Multi-View Reconstruction

Aayushi Kedar Barve

Professor : Alcherio Martinoli
Assistant(s) : Alexander Wallen Kiessling

Conventional 3D reconstruction pipelines rely on RGB imagery, which degrades under low lighting, featureless surfaces, or visually obscured environments such as smoke-filled or dark scenes. Thermal infrared cameras capture emitted heat signatures rather than reflected visible light, making them an attractive alternative in these conditions. The objective of this project was to evaluate the effectiveness of classical Structure from Motion (SfM) methods on thermal images.

Two camera calibration methods were evaluated — a mask-based technique using centroids of bright blobs visible through cut-outs on a heated background, and a standard chessboard-based method. The mask-based method proved more robust under high lens distortion and emerged as the preferred approach.

For detecting correspondences across viewpoints, 7 detector-descriptor pairs (SIFT, ORB, AKAZE, KAZE, BRISK, FAST+BRIEF, FAST+FREAK) and 6 preprocessing methods (linear normalization, log normalization, histogram equalization, CLAHE, Gaussian sharpening, Gamma correction) were evaluated using average feature count, pseudo precision and recall based on RANSAC inliers, good matches and total keypoints — since ground truth correspondences cannot be assumed available for arbitrary thermal scenes.

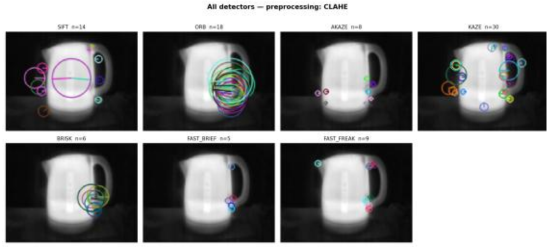

Fig 1. Features detected with different detectors for CLAHE

The lack of features as seen in Fig 1. and Table 1. reaching a maximum of 30 per image on average, paired with the fact that the detected matches are not geometrically consistent enough, made classical feature detection and matching on thermal imagery insufficient for subsequent accurate pose estimation from the detected matches. Deep learning-based methods such as SuperPoint and AnyThermal are also not suitable for this task since the former was primarily trained and designed for RGB images, and the latter generates patch-based descriptors that are not localized enough for pose estimation. Incremental SfM and global SfM techniques (with averaged world frame rotations and translation estimates using different viewpoint pairs) were also compared, with the latter generating a lower estimation error compared to the ground truth at the cost of being more computationally involved.

Preprocessing	Detector	Avg. Features/Img	Precision	Recall
CLAHE	KAZE	29.1	0.448	0.173
Hist Eq	SIFT	22.0	0.371	0.160
Gauss+Sharpen	KAZE	21.0	0.392	0.187
Log Norm	KAZE	18.1	0.380	0.221
Gamma(0.5)	KAZE	18.1	0.352	0.199

Table 1. Feature detector evaluation results for the 5 best combinations

The ground truth poses were then used to evaluate the performance of dense reconstruction methods. Voxel carving and depth-based dense reconstruction using photometric error (where per-pixel depth was estimated by minimizing over a cost volume for a range of depth hypotheses projected onto other camera views with respect to a reference image) were tested and both produced successful reconstructions of a heated electric kettle. However, both methods are not identically generalizable across different scenes and work because of the high thermal gradients available due to the contrast of the heated electric kettle against the room-temperature background.

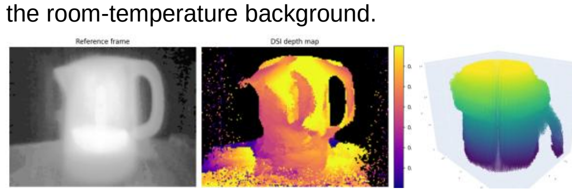

Fig 2. (a) Depth map for an example reference image and (b) Voxel carving reconstruction

In conclusion, 3D reconstruction with thermal-only information is challenging due to the lack of features and geometrically consistent matches. The primary bottleneck is camera pose estimation, and supplementary sensors such as IMUs can be used to aid in motion estimation between camera views. Other promising directions of investigation include visual-thermal fusion and the design of different approaches to find correspondences across thermal-only images outside of classical feature detection pipelines.