1. **Introduction**

In general, 3D dense point clouds of images are generated from a commonly used two-step method. In the first step, the *Structure from Motion (SfM)* approach is applied; and then, dense point clouds can be provided by using dense surface reconstruction method (Szeliski R., 2010). However, in this thesis, we will introduce how to extract images from video streams.

This research is based on PTAM (Parallel Tracking and Mapping) which selects key-frames from a video stream, employs full bundle adjustment to the selected key-frames; meanwhile, estimates the pose of the intermediate frames robustly in real-time (Klein G., Murray D., 2007). Nevertheless, PTAM is not adequate to our case which is a video stream from a UAV-based system. Therefore, the primary objective of this thesis is to improve PTAM algorithm to extract good frames that could be used for 3D reconstruction from a video stream captured by a UAV-based system.

2. **Overview of PTAM**

Figure 1 is the flowchart of PTAM which illustrates its working procedure. PTAM works well for a small AR workspace; however, a video stream acquired by the camera on UAV records the scenes of a relatively large region. Furthermore, blurred images cannot be avoided for the reason that oscillation and rotation of UAV are almost inevitable, but only FAST feature detector is employed in PTAM which is not capable of handling the extreme situations in our video stream. Therefore, robust feature detectors and descriptors should be integrated into the tracking system.
3. Methodology

The purpose of integrating robust feature detectors and descriptors is to extract stable and correct image correspondences from the continuous image sequences and make the tracking system more robust. This procedure can be divided into three parts shown...
in Figure 2. In this research, SIFT and SURF feature detectors and descriptors are integrated.

Figure 2. Feature matching procedure

4. Evaluation and analysis

4.1 Time evaluation results

Time evaluation is done by calculating the average time to extract one keypoint. The result is shown in Figure 3.

![Time for Detecting One Keypoint](image.png)

4.2 Robustness evaluation results

Robustness is compared among four different image sequences from the video stream. For each sequence, 1000 strongest feature points are chosen to establish the initial trails in the beginning; next, these initial trails are tracked in the following 30 frames, during this procedure, bad trails which does not fulfill the preset conditions are discarded while the position of good trails in the current frame are stored. When feature tracking is done for one frame, the inliers and outliers are evaluated by estimating the fundamental matrix for the good trails using RANSAC. In addition, the ratio between the number of inliers and the number of good trails is obtained.
4.2.1 UAV flying stably in one direction

Figure 4. Image sequence used for evaluation: UAV flying smoothly in one direction

Figure 5. UAV flying smoothly in one direction: “Number of Good-trails Comparison” plot & “Inlier Ratio Comparison” plot

From Figure 4, we can see that these frames have high similarity even between the first frame and the last frame. From Figure 5, we could not justify which algorithm is absolutely the best in this kind of circumstance.

4.2.2 UAV flying upwards stably

Figure 6. Image sequence used for evaluation: UAV flying upwards stably
Similar as the first situation, these three algorithms all perform well and it is difficult to say which performs best when the image scale changes and the images have low texture.

4.2.3 UAV shaking strongly

In spite of these images in Figure 8 coming from a video, they varies a lot even between
two adjacent images. According to Figure 9, we can draw the conclusion that SURF is the most robust algorithm to search for correspondences; FAST obviously has the worst performance.

4.2.3 UAV flying downwards unstably

From Figure 11, we can see that FAST is not capable to deal with this scenario obviously; while SIFT and SURF performed both well because they are invariant to scale and rotation. However, SURF held the best performance.

Figure 10. Image sequence used for evaluation: UAV moving downwards and shaking

Figure 11. UAV moving downwards and shaking: "Number of Good-trails Comparison" plot & "Inlier Ratio Comparison" plot

5. Conclusion

1) In ideal circumstances, i.e. UAV moving stably in invariant direction, FAST, SIFT and SURF have similar performance. They all can provide reasonable results and from the view of time, FAST used in PTAM is better than SIFT and SURF.

2) However, FAST utilized in PTAM is not capable of handling strong vibration
situations while SIFT and SURF can produce acceptable results. From the figures of results, we can see clearly that the system using FAST crushed in the early stage, but system applying SIFT and SURF can complete tracking and the inlier ratios are also acceptable.

3) To sum up, even though SIFT and SURF are time-consuming compared to FAST, they are more suitable to process the video stream taken by the UAV-based system.

**Bibliography**
