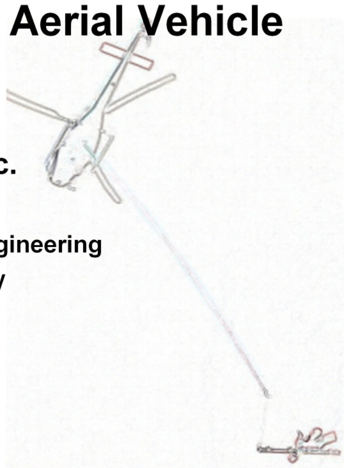
Carleton UNIVERSITY
Canada's Capital University

Obstacle Detection Using Monocular Camera for Low Flying Unmanned Aerial Vehicle

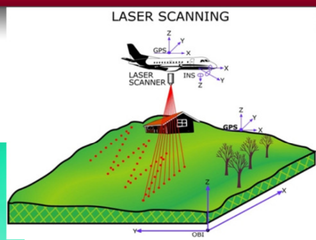


Fan Zhang, B.Sc.

Systems and Computer Engineering
Carleton University
January 2015

Carleton UNIVERSITY
Canada's Capital University

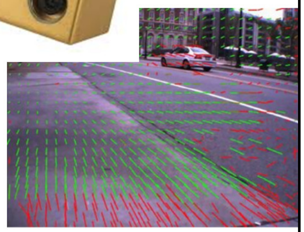
Introduction



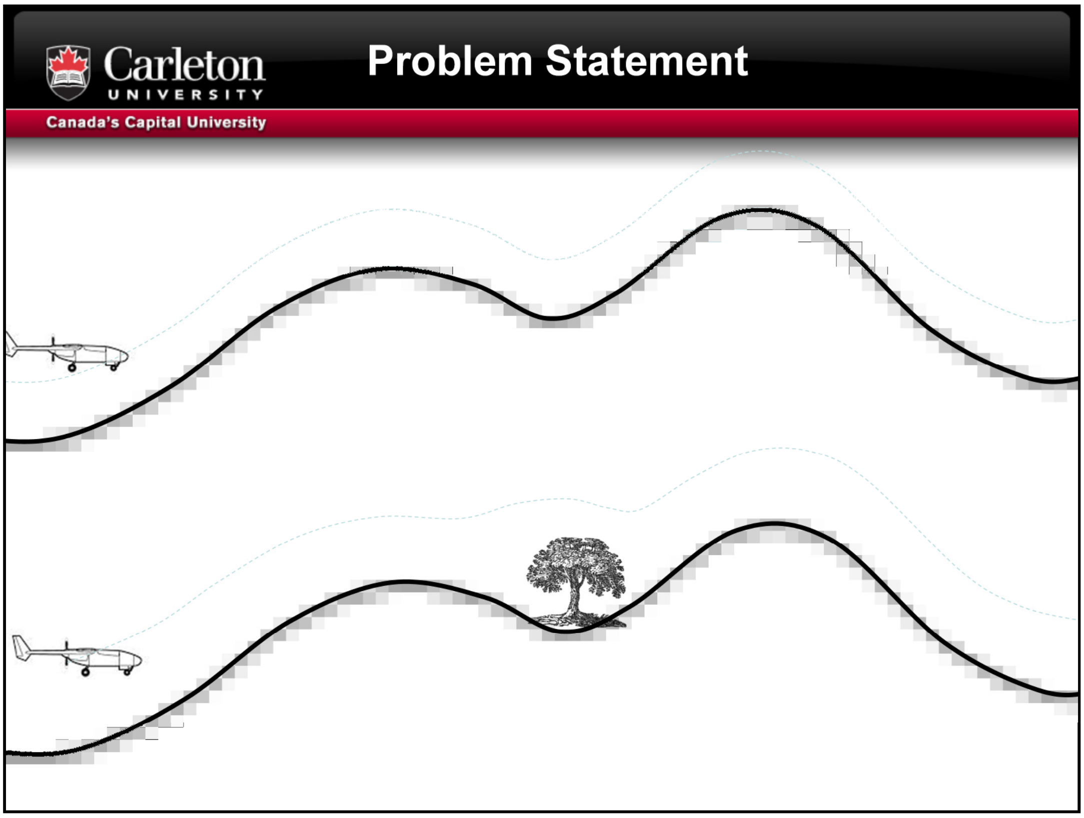
- **Bearing and Range Sensors**
  - Radar, Lidar, Sonar
    - *Require Scanning*
  - 3D Flash Lidar

- **Bearing Only Sensor**
  - Camera
    - *Binocular Configuration*
    - *Monocular Configuration*

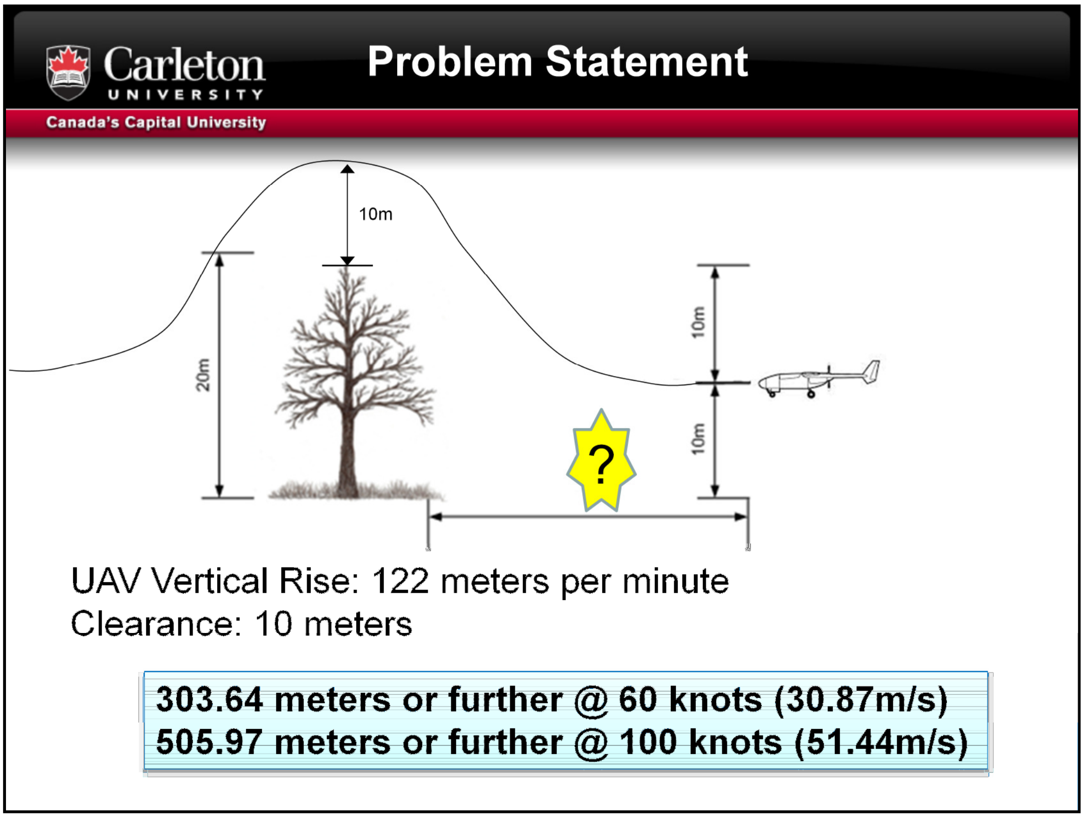
- **Computer Algorithm**
  - Machine Vision
  - Simultaneous Localization and Mapping (SLAM)



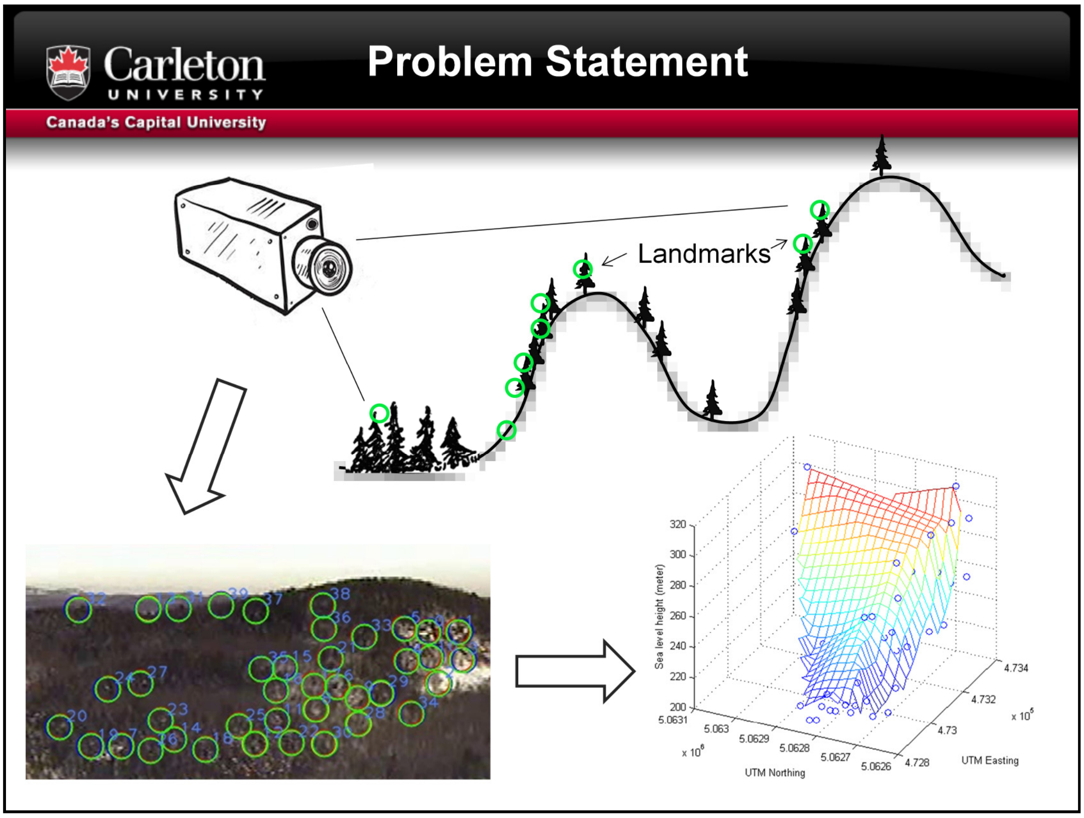
•In order for the UAV to avoid obstacle, we must detect, and know the distance and direction of the obstacle from the UAV.
•Many sensors are used in obstacle detection application. Some sensors are range and bearing sensor, meaning that they output range and direction measurement of the target, such as radar, lidar, and sonar. However, to get a 3D depth of the scene, mechanical scanning is required. For example, this picture illustrate a Lidar pointing downward, and performaning 2D scanning in order to get an elevation model of the terrain.
•3D flash Lidar(point) can obtain a depth measurement of the entire scene simultaneously by illuminating the entire scene with laser and detect the reflected laser with 2D imaging sensor. However, they are very pricy, which limits their use in commercial application.
•Another solution is to use video camera. They are inexpensive, small size, has many options to choose from. Because image sensors are bearing only sensor, in order to measure distance of a target, you need two views of the same target, and obtain depth measurement through triangulation. Same way as how human eye works. There are two type of configuration. Binocular configuration has two cameras sitting side by side, which mimic the human eye. Target distance can be calculated from the disparity of the target in the two images and the baseline of the two cameras. Monocular configuration use only one camera to capture an image sequence. By comparing the subsequence image frames, an optical flow image can be generated (this is an example of the optical flow image). The target position can be calculated using the optical flow result with the odometry measurement that provide UAV displacement information from frame to frame. When no odometry measurement is available, monocular camera can still obtain a structure of the scene, but cannot measure target distance.
•Using video camera to measure object distance requires machine vision algorithm. In this work, the machine vision algorithms processed the image sequence to extract and track visual features. On top of that, a SLAM framework is used to fuse output from machine vision algorithm with the odometry measurement to estimate UAV trajectory as well as mapping landmark from the scene.



For this research, it focus on small to medium size UAV conducting low altitude terrain following flight in natual scene. [click] The main application of this UAV is for low altitude high resolution airborne geological survey.
Digital elevation map is usually used to plan the flight path of a UAV. However its resolution is too low for obstacle detection. For example, when in this scenario, the UAV must be able to spot the obstacle, and plan a way around it. [click]
Only static scene is considered at this stage.

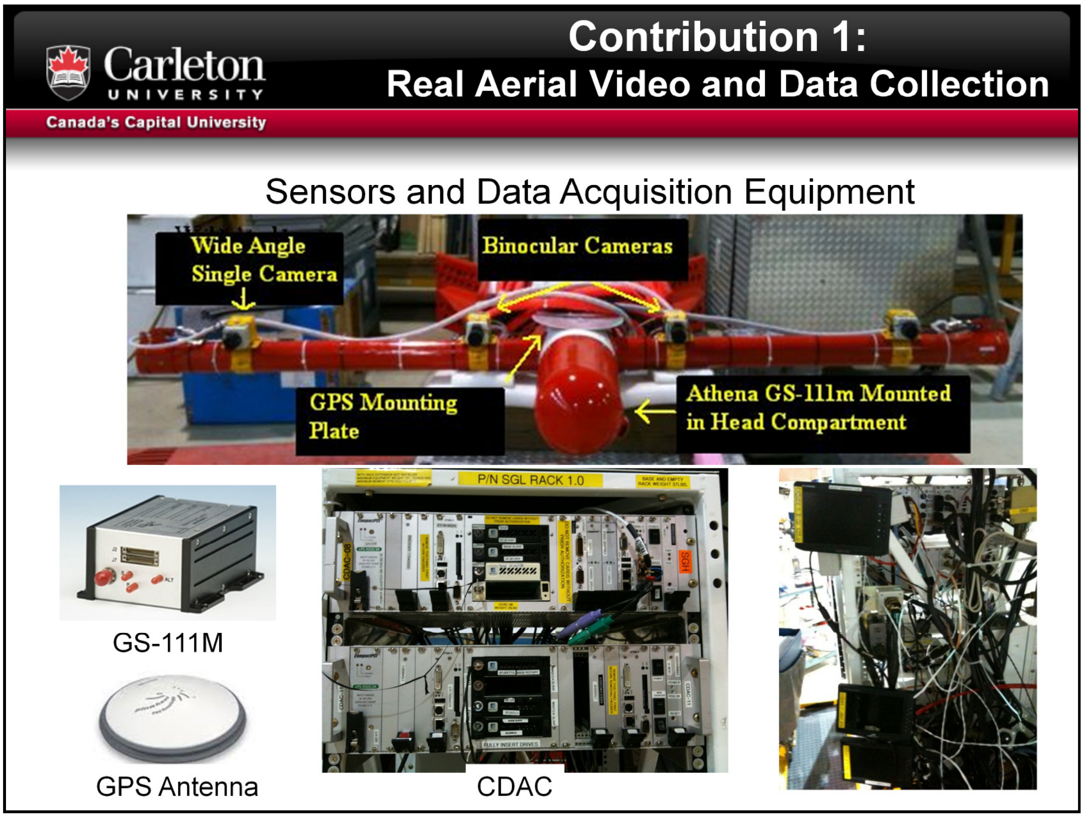


Then, how far do we need to detect the obstacle so that the UAV has enough time to avoid it. Based on the specification of the GeoSurv2, and giving about 10m of clearance from the obstacle, the UAV need to detect the obstacle at least 300 meters away when flying at 60 knots, and 500 meters away when flying at 100 knots

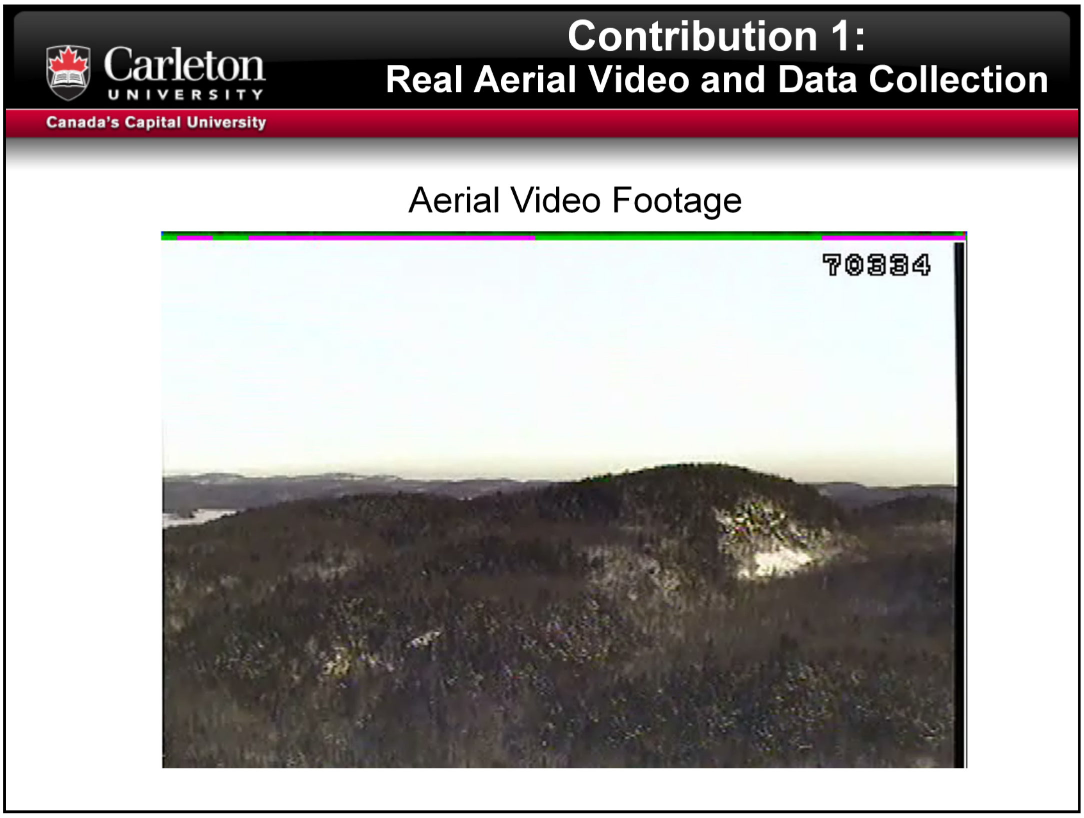


The solution proposed by the thesis, is to capture image sequence of the scene with a forward looking monocular camera. The machine vision algorithm then extract visual feature from the scene, and tracked them as long as they remained in the field of view. The physical point in 3D world that associated to the visual feature is called landmark. The SLAM algorithm can then use the displacement of these landmark from frame to frame, and along with the motion measurement of the camera, to estimate the trajectory of the camera, and the position of the landmarks. Then a sparse map of the scene can be generated from the landmarks position.

The first contribution of this thesis is that I have collected real aerial video and navigation data. Here is a video of the take off of the helicopter and the SUAS.
The test flight site is in the mountain area north of Gatineau. A simulated unmanned aircraft system (SUAS) was fabricated at SGL, and was used to carry all sensors. The SUAS was towed by a helicopter via a tow rope of 33 meters long. The helicopter flew a planned path at 100m above ground. The SUAS was approximately at 60-70 meters above ground.



Sensors mounted on the SUAS included one wide angle CCD camera with 6 mm focal length capturing monocular video at 30 fps, one GPS antenna mounted on the plate here, and one INS/GPS navigation unit Athena GS-111m mounted in the head compartment of the SUAS. There were also two narrow angle cameras install for capturing binocular video. These are installed to another research. Analog videos were sent to the helicopter through BNC cables. Navigation data were sent through RS485 signaling. All the videos and data were recorded with CDAC installed in the helicopter. There were three monitors installed in the helicopter so that the operator can monitor the video from the CCD camera.



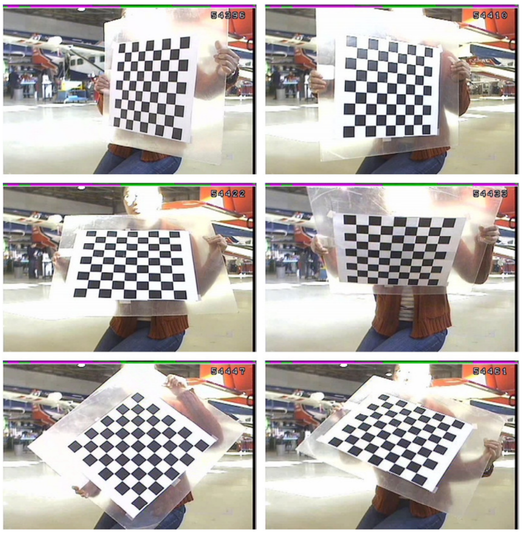
The video captured by the cameras was digitized to 480x720 resolution and time-stamped with GPS second on the image screen for post-flight synchronization with the inertial measurements. Here is a sample aerial video that we took.

Carleton UNIVERSITY
Canada's Capital University

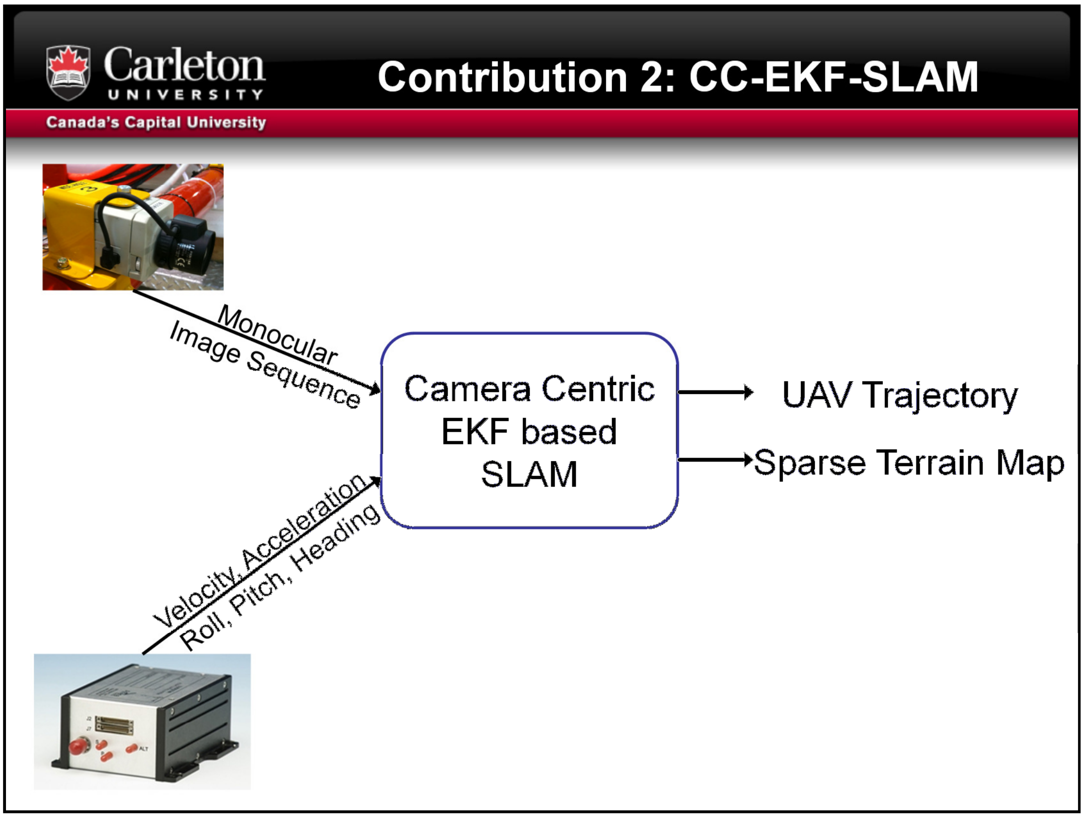
Contribution 1: Real Aerial Video and Data Collection

Camera Calibration

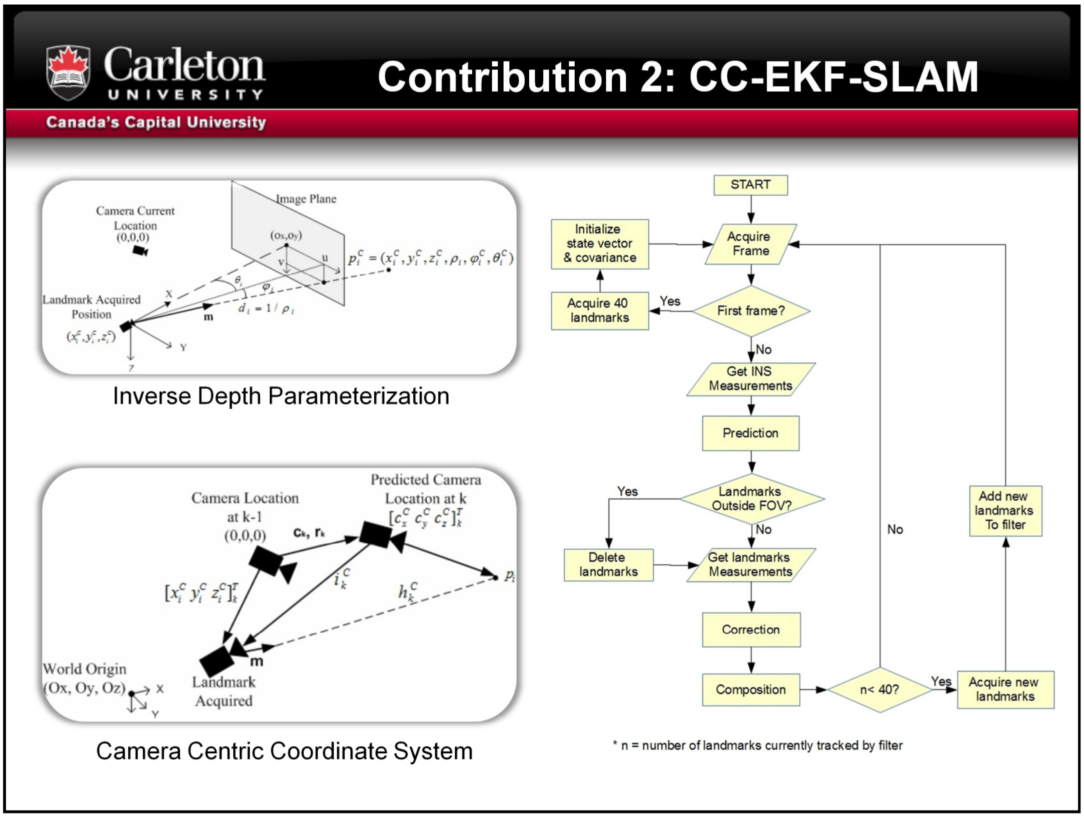
Parameter	Result
f_x	887.6 pixels
f_y	805.7 pixels
c_x	381.8 pixels
c_y	293.7 pixels
k_1	-0.102
k_2	-0.535
p_1	1.15e-003
p_2	8.40e-003



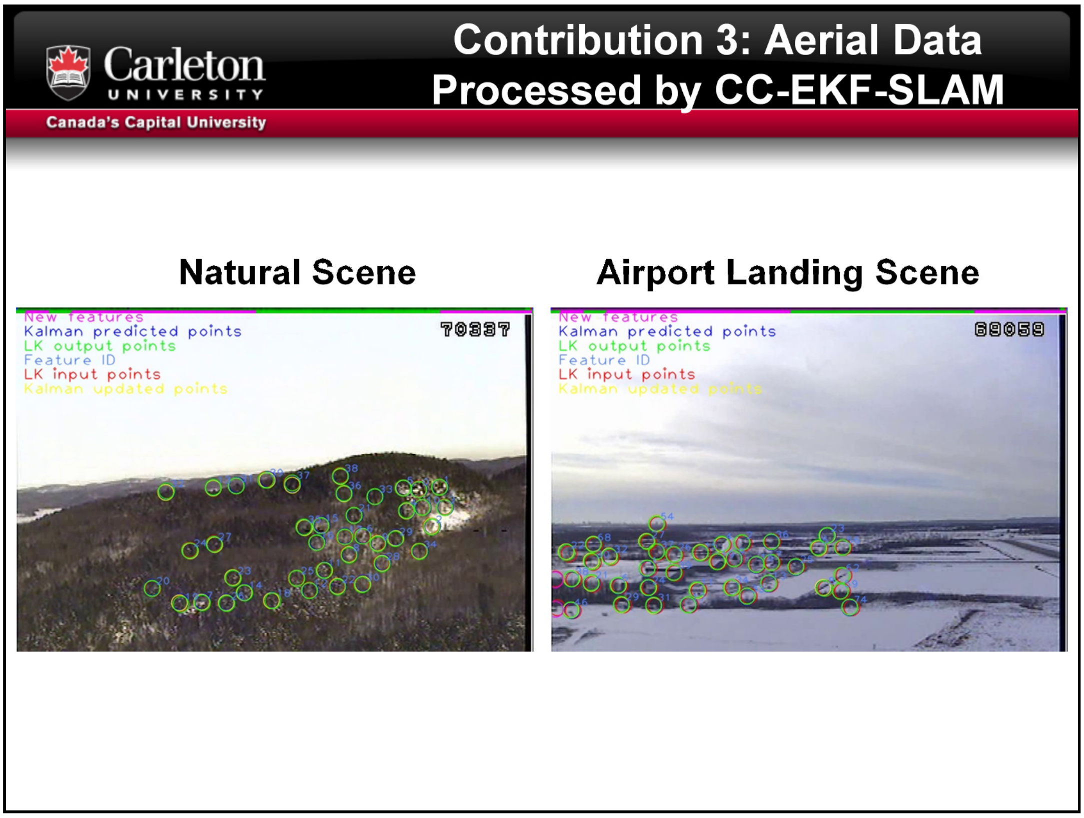
A camera calibration was done right after the flight. The calibration algorithm takes different view of the calibration target which is a black and white checkerboard, calculate the camera intrinsic and extrinsic parameters. The results are listed here.



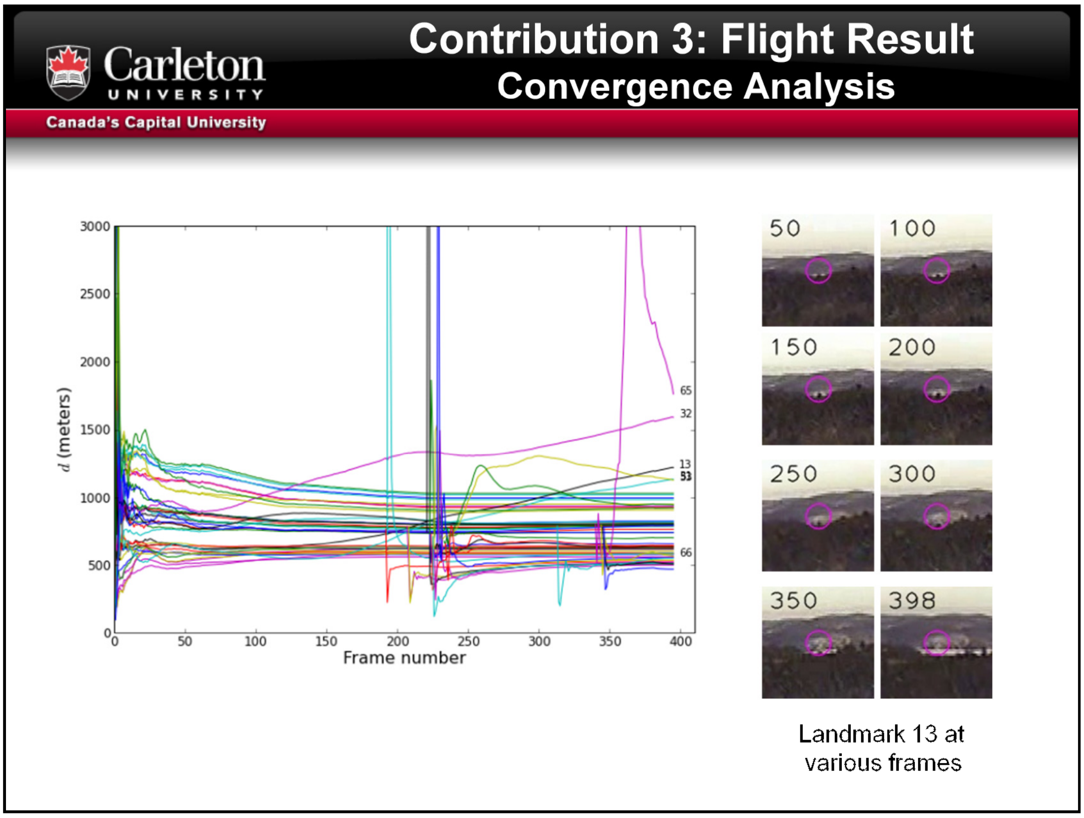
2nd contribution is the implementation of the CC-EKF-SLAM algorithm. The algorithm used an extended Kalman filter based SLAM algorithm to fuse monocular image sequence with inertial measurements, and generate estimation on UAV trajectory and landmark positions.



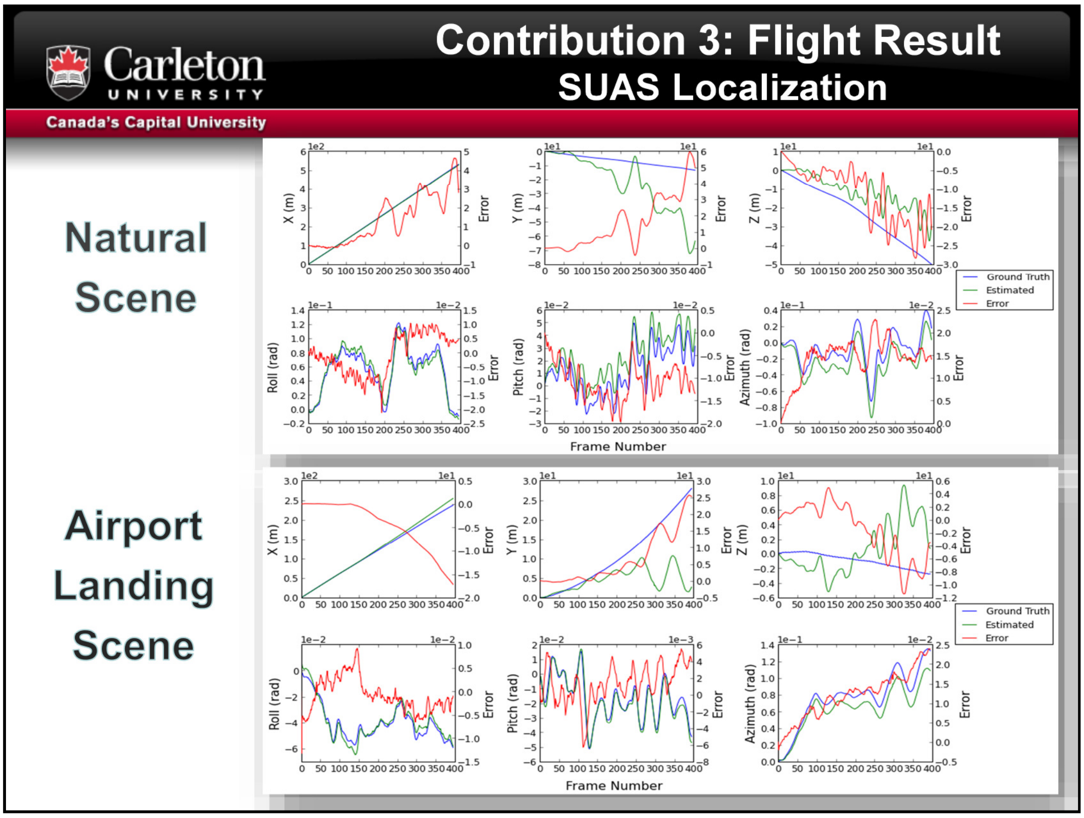
The algorithm has a few unique features. 1st, all landmarks represented in inverse depth parameterization which improve the linearity for distant objects. 2nd, all parameters are presented in camera reference frame at all time to improve the consistency property of the extended Kalman filter when mapping large area. The algorithm is setup to remove old landmark and acquire new landmark in order to achieve continuous operation.



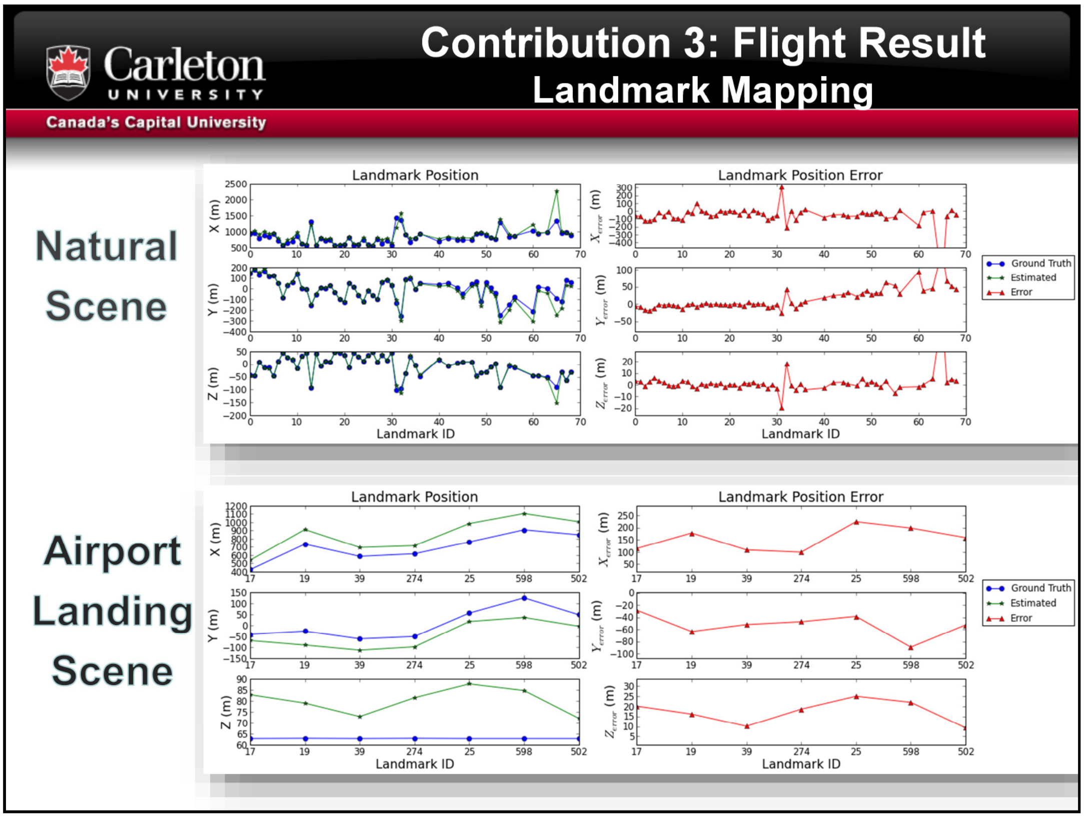
The 3rd contribution is that the aerial video and data were processed by CC-EKF-SLAM algorithm, and the results agreed with the ground truth in general. Two pieces of video were processed, one is a natual scene video, and the other one is airport landing scene. Each piece is 400 frames long. The airport landing video was processed because in natural scene, it is difficult to judge the correctness of the correspondence between estimated landmark and the ground truth landmark. In airport landing scene, manual correspondence can be made by choosing visual feature on man-made objects.



First the convergence behavior was analyzed. Most landmarks converged very quickly, within 30 frames. The landmarks that are drifting away are those located on the hill top. Because the small error allowed in the visual tracking algorithm, these visual feature changed slightly from frame to frame, and eventually became something different. For example, the initial view of landmark 13 is at the hill top. At the end of the video, it has moved to the mountain behind the lake.



For the accuracy of SUAS localization, the estimated poses generally agree with the ground truth. The position on the Y and Z axis sees more drift, and the position on the X axis has very good accuracy. At the same time, it can be observed that there is correlation between the position error and the rotation of the SUAS. The position error on Y is correlated to rotation around Z, and position error on Z is correlated to the rotation around Y.



•For accuracy of landmark mapping
•In both videos, average landmark distance from the SUAS is around 1000meters, with error within +/- 100meter. A good indicator that the algorithm is able to map object around 1000m range.
•Landmark with ID bigger than 40 has offsets. These are the landmarks initialized after the 1st frame. The error analysis revealed the cause of these offset error.
•In airport landing video, all landmarks has a stable offset error regardless when they were initialized. Since all landmarks are located at the corner of the image plane, these error were likely caused by lens distortion.

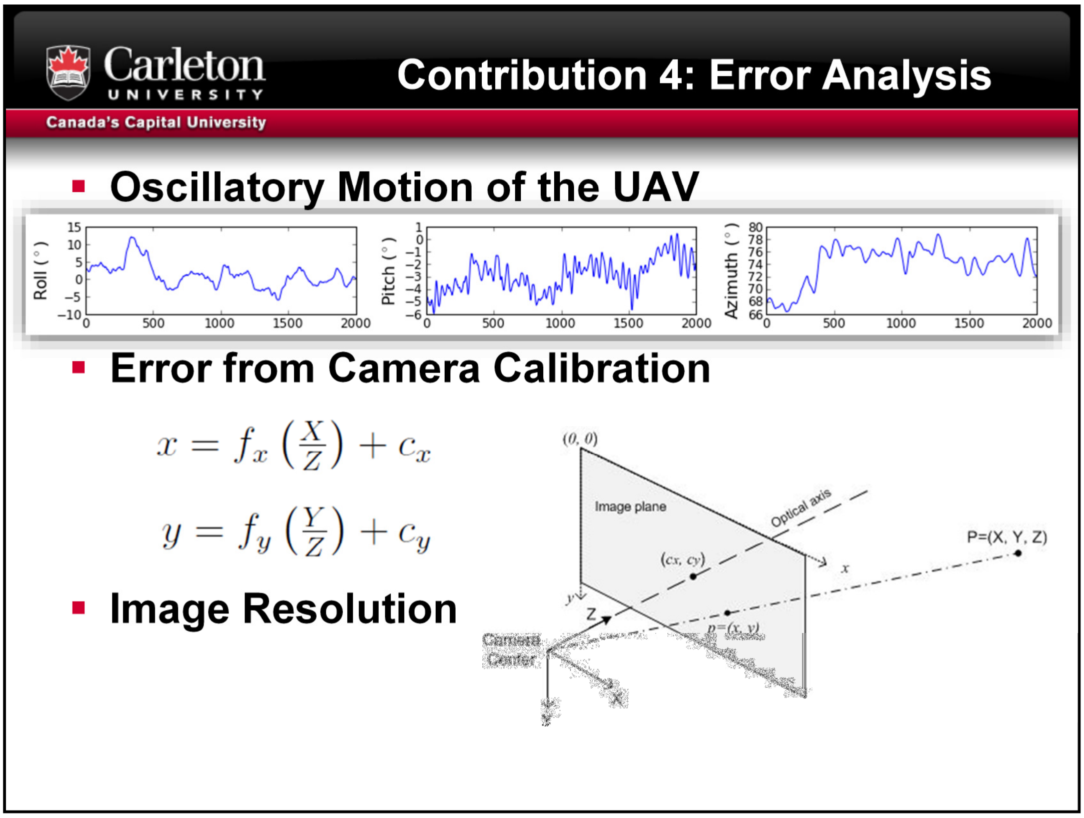
Carleton UNIVERSITY
Canada's Capital University

Contribution 3: Flight Result

- **This contribution is published in**

The 2012 IEEE International Instrumentation and Measurement Technology Conference

[7] F. Zhang, R. Goubran, and P. Straznicky, "Obstacle detection for low flying UAS using monocular camera," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference*, 2012, pp. 2133-2137.

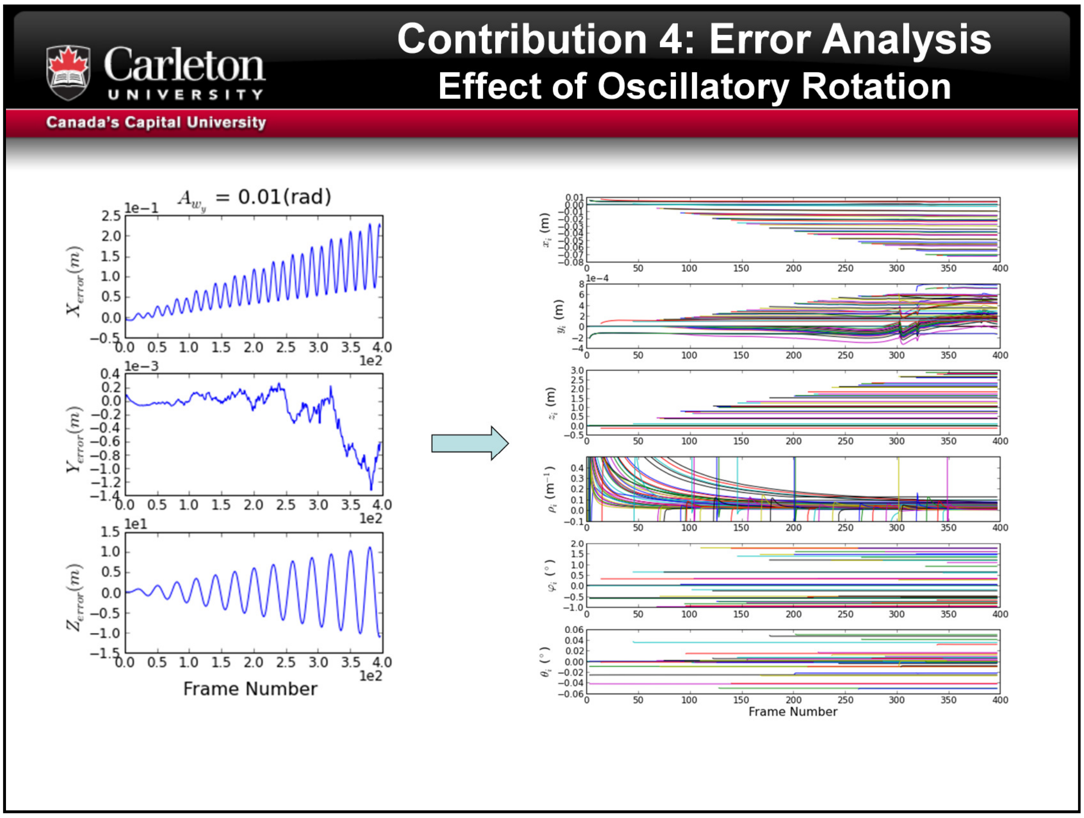


The 4th contribution is that I further analyzed the performance of the algorithm under a number of scenario.

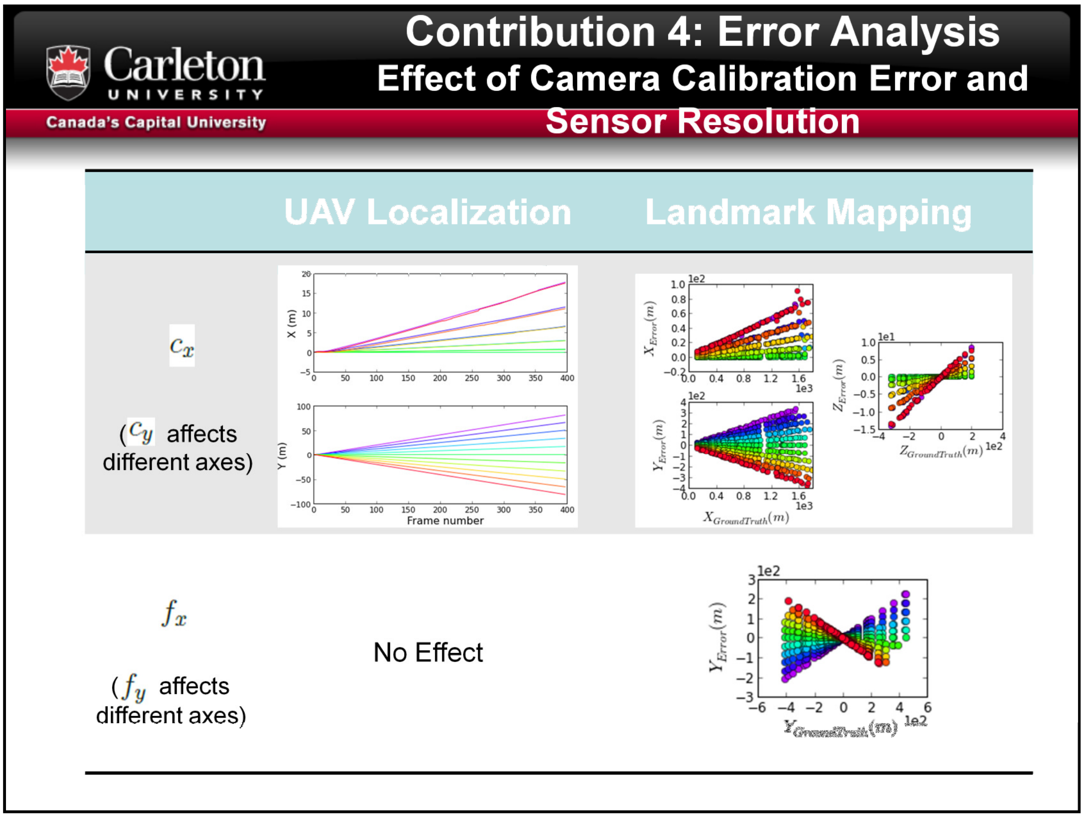
1st scenario is oscillatory motion of the UAV. The test flight data showed that UAV is under a lot of oscillatory motion, especially the rotations. Here is a piece of rotation recording of the SUAS. We can see that the Y and Z axis has lots of oscillatory rotations.

2nd scenario is error in camera calibration. It was found that when feeding the calibration program with different amount and different views of the input images, the calibration result is different. Therefore, the calibration results do have error in them, and it would be good to know how much impact they have on the accuracy of the algorithm.

3rd scenario is the sensor resolution. Since there are lots of choice for image sensor at various resolution. This analysis will help us deciding how much resolution we need for the targeted distance.



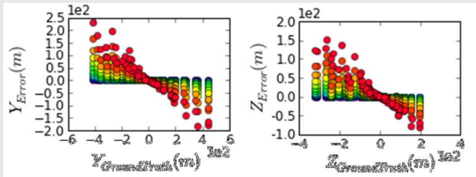
The analysis on oscillatory motion confirmed that CC-EKF-SLAM is very sensitive to the oscillatory rotation. This slides shows result of UAV experiencing rotation around Y. The estimated UAV position are showing oscillatory and diverging error on both X and Z. The analysis also showed that as a result of this self localization error, the landmark mapping have offset error for landmark not initialized on the first frame



The next two slides summarized the effect of camera calibration error. For error in Cx, or Cy, which is the coordinate where optical axis intercept the image plane, it affects both UAV localization and landmark mapping, and the error can be modelled by 1st order polynomial function. For the scaling factor fx,and fy, it doesn't affect UAV localization, only the landmark mapping. The error can also be modelled by 1st order polynomial function

Carleton UNIVERSITY
Canada's Capital University

Contribution 4: Error Analysis
Effect of Camera Calibration Error and Sensor Resolution

	UAV Localization	Landmarks Mapping
Lens Distortion	Diverging Error	
Image Resolution	1080 x 1440 or higher	1080 x 1440 or higher

For lens distortion, it caused diverging error in UAV localization, as well as error in landmark mapping. The further the landmark locate from the optical center ON IMAGR PLANE, the more error it suffer. At last for sensor resolution, in order to achieve a good accuracy for distant object, sensor resolution should be 1080x1440 or higher.

Carleton UNIVERSITY
Canada's Capital University

Conclusion

- Contribution 1:

  Aerial Video and Data Collected through Test Flight

- Contribution 2:

  CC-EKF-SLAM Implemented

- Contribution 3:

  Aerial Data Processed by CC-EKF-SLAM

- Contribution 4:

  Error Analysis

Carleton UNIVERSITY
Canada's Capital University

Recommendations for Future Work

- Add lens distortion model
- Increase sensor resolution
- Add landmark quality checking and filtering function
- Research on map joining algorithm
- Increase accuracy by syncing to GPS
- Investigate on the sensitivity problem to oscillatory rotation

Carleton UNIVERSITY

Canada's Capital University

Acknowledgement

- Rafik Goubran, Paul Straznicky
- Sander Geophysics Ltd.
- Ontario Centers of Excellence and NSERC
- My Family