Image Mosaicing for UAV-based Pushbroom Hyperspectral Imaging

Ralph Aaron Aguinaldo[1,2], Mara Alain Maestro[1], Micherene Clauzette Lofamia[1], Andrew Rafael Bañas[1], and Maricor Soriano[1]
[1]National Institute of Physics, University of the Philippines Diliman,
Quezon City, Philippines
[2]Email: rasaguinaldo@gmail.com

KEY WORDS: pushbroom, image stitching, image mosaicing

ABSTRACT: Hyperspectral imaging has many applications in agricultural and environmental remote sensing. There are many metrics that can be derived from spectral data such as vegetation indices which can detect the presence of stress and disease. A push-broom imaging system captures via line-scanning, and is capable of a higher spectral resolution than a snapshot camera which can only operate on limited spectral band counts. Combining this with the low-cost and flexible deployment of unmanned aerial vehicles (UAV) can yield on-demand data of high spatial and spectral resolution. The challenge lies in mosaicing the push-broom images. Since each image line-scan has only one spatial axis, using area and feature based image mosaicing algorithms is ineffective. We designed a push-broom hyperspectral imaging system from commercial-off-the-shelf (COTS) components intended for UAV flight captures. This paper details the image mosaicing approach that utilizes the data from Global Positioning System (GPS) and inertial measurement unit (IMU) sensors. The algorithm is homography-based and uses the position and attitude of the UAV and camera to compute the relative positions of the target and image planes, and projects the line-scans in a single mosaic. Test flights were performed using a DJI Matrice 600 Pro to determine the accuracy of the mosaics in preserving known measurements and features.

1. INTRODUCTION

1.1 Hyperspectral imaging in agriculture

In hyperspectral imaging, the spectrum for each pixel in the image is obtained to find objects, identify materials, or detect processes (Chilton, 2013). Most optical payloads for remote sensing are designed to collect spectral signatures (Hagen, 2013). In plants, variations in reflectance are caused by differences in species, health and/or potential yield-information that is useful in applications such as environmental monitoring, resources management and disaster risk reduction (Ferwerda, 2005). The use of hyperspectral remote sensing has increased for monitoring the development and health of vegetation, particularly food crops (Reina, 2018).

1.2 Advantages of line-scan capture

Multispectral imaging systems often employ taking multiple 2D images over several spectral values. However, a line-scanning or push-broom imager builds an image by stacking a series of line (1D) images corresponding to a series of multiple shots, like a flatbed scanner. This allows the scanning imager to take advantage of slits and prisms to disperse the light captured over one sensor axis, allowing a finer spectral resolution (Shawt & Burke, 2003). In a hyperspectral pushbroom imager, each effective 2D capture is composed of a spatial axis and a spectral axis distributed to the pixel count of the corresponding sensor axis. Subsequent lines from the scan form a hyperspectral cube, a stack of spatial captures at fine spectral increments (Suomanlainen et al, 2014).

1.3 Advantages of unmanned aerial vehicles

Unmanned aerial vehicles (UAV), or drones, are aircrafts that require no onboard crew, and have varying degrees of autonomy. There has been a surge of advancements and demand for commercial drones in the last decade, primarily for recreational, scientific and industrial applications (Hu & Lanzon, 2018). In agriculture, drones are used extensively in advanced large-scale farming for quick monitoring of crop growth and conditions over large areas. Primarily used for spectral imaging, assessment of crop conditions leads to efficient farming methods, adjustments and yield (Reina, 2018). While it is limited in area coverage as compared to satellites, UAVs have the advantage of lower cost, on demand deployability, user-defined flight trajectories, flexible capture parameters as well as low-altitude, high resolution captures. This makes UAV imaging especially effective in disaster and risk assessment, emergency surveys and remote sensing in areas that need closer inspections (Sharma et al, 2020) .

1.3 Challenges in line-scan stitching

Most remote sensing data are arranged and represented spatially via image mosaicing or stitching. Image mosaicing is the combination of two or more images in a single image space. Often, this would include image transformations that account for differences in capture parameters to project the images in such a way that features and details are preserved and aligned properly (Ghosh & Kaabouch, 2016.). Widely used algorithms include the SIFT and SURF algorithms. Scale-Invariant Feature Transforms (SIFT) -based mosaicing looks for image features that are invariant to scale, noise and illumination to find points at which to join images together (Lowe, 1999). Speeded-Up Robust Features (SURF)-based is like SIFT, but with reduced computational complexity to speed up performance (Bay et al, 2006) **.** While said algorithms are effective on conventional 2D images, they cannot handle a line-scan input, as these algorithms rely on finding and matching features to align images.

1.4 Proposed solution

This study aims to develop an image stitching algorithm designed to handle the 1-dimensional spatial output of aerial hyperspectral imaging. While there are often no matchable features in a line of pixels, UAVs are equipped with sensors such as the Global Positioning System (GPS) to determine its lateral position and altitude, and inertial measurement units (IMU) to determine its orientation. If camera parameters and capture conditions are known, it is possible to estimate the image plane coordinates of any point on a target area by building a camera model.

2. STITCHING ALGORITHM

2.1 Image coordinate estimation via homography matrix

The algorithm is a homography-based approach that uses the pinhole camera model to project real world coordinates to image coordinate systems. Any real world object coordinate (X_o, Y_o, Z_o) can be mapped to an arbitrary space coordinate (x_a, y_a, z_a) using

$$\begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix} = C_{int} C_{ext} \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{bmatrix} \tag{1}$$

where C_{int} is the intrinsic camera matrix C_{ext} and extrinsic camera matrix, respectively (Hartley and Zisserman, 2003). The arbitrary space coordinate can be mapped to a two-dimensional image plane via a scale factor $1/z_a$ as

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \frac{1}{z_a} \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix}. \tag{2}$$

The intrinsic camera matrix is

(3)

and is defined by the constant parameters that are intrinsic to the imaging system, namely the focal length f, the sensor size (x_s, y_s), and the width and the height of the captured image (w,h). The optical center (o_x, o_y) of the camera is at (w/2, h/2).

The extrinsic camera matrix is

$$C_{ext} = \begin{bmatrix} R & T \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{4}$$

and consists of external parameters that define the capture geometry. C_{ext} is composed of a rotation matrix R and a translation matrix T. The translation matrix defines the position of the camera in the real world coordinate system,

and can be calculated from the GPS records of the UAV. Meanwhile, the rotation matrix is the orientation of the optical axis relative to the real world coordinate system, and can be calculated from the IMU unit of the UAV-system and camera gimbal. The rotation matrix can be broken down into its components from (left to right) as

[redacted] (5)

for roll(□, along x-axis), pitch (□, along y-axis) and yaw (□) , the rotations along the x, y and z, respectively.

The product of intrinsic and extrinsic camera matrices is the homography matrix *H*, defined as

$$\begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix} = H \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{x_s} & 0 & \frac{w}{2} \\ 0 & \frac{f}{y_s} & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_o \\ Y_o \\ Z_o = 0 \\ 1 \end{bmatrix}. \quad (6)$$

Note that when we are looking for a planar surface in the real world, to compute the homography, we set $Z_o = 0$, thus removing the contribution of the third column of C_{ext}. *H* will then map a real-world object point in the arbitrary space. The coordinates in the pixel space, (x_i, y_i), can then be calculated by applying the scale factor using

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \frac{1}{z_a} \begin{bmatrix} \frac{f}{x_s} & 0 & \frac{w}{2} \\ 0 & \frac{f}{y_s} & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ r_{31} & r_{32} & t_z \end{bmatrix} \begin{bmatrix} X_o \\ Y_o \\ 1 \end{bmatrix}. \quad (7)$$

2.2 Algorithm Flowchart

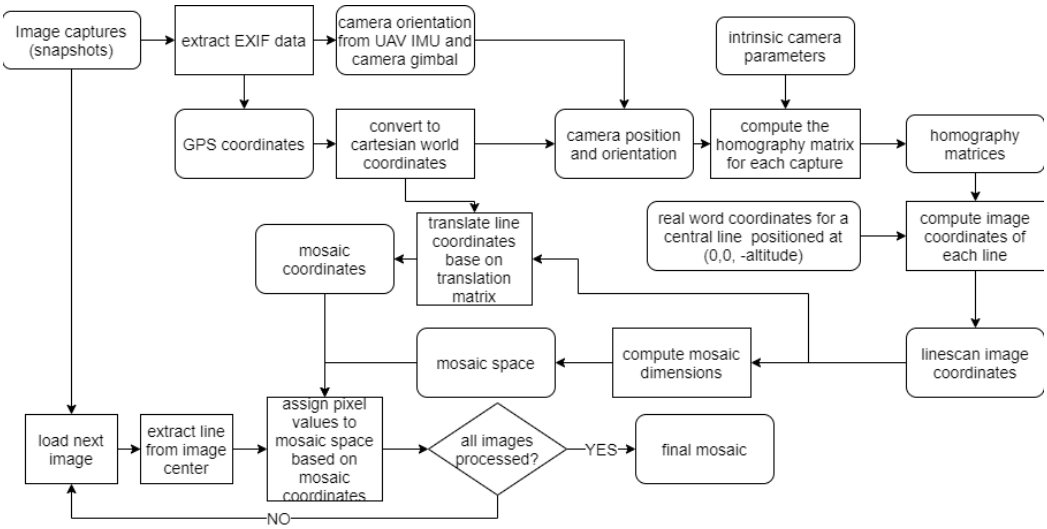


Figure 1. Algorithm flowchart for line-scan image mosaicing

Figure 1. shows a flowchart for the algorithm. During the software development that happened in parallel with the imaging system design and assembly, actual 2D aerial snapshots were used, with the central line perpendicular to scan direction isolated to simulate the 1D output of a line-scan hyperspectral imager. The next section discusses the unmanned aerial vehicles

3. DATA GATHERING PARAMETERS

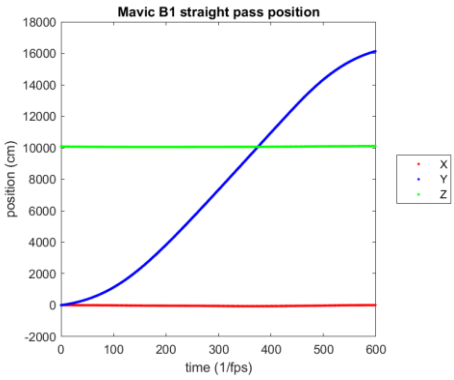
There are two set-ups used in testing the stitching algorithms. Table 1 details the UAV, camera and flight parameters in the aerial captures. Set-up 1 uses a drone that does not record attitude data, and was used exclusively for near-stable orientation flights, where only the translations are significantly changed. Set-up 2 was used to investigate how well rotational input and direction changes are incorporated.

.

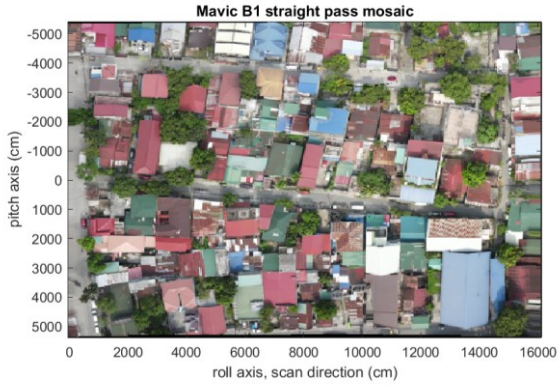
Parameters		Set-up 1	Set-up 2
UAV		Mavic Air 2	DJI Matrice 600 Pro
imaging system	camera	built-in	Zenmuse X3
	pixel size	4.434 microns	1.55 microns
	focal length	8.0 mm	3.6 mm
	capture modes	video (60 FPS)	snapshot
	image size	1920x1080	4000x3000
position sensors	GPS sensor	Present (SRT)	Present (EXIF)
	GPS format	WGS84	WGS84 (EXIF)
orientation sensors	IMU	N/A (present but cannot be extracted)	Present (EXIF)
	camera gimbal	N/A (present but cannot be extracted)	Present (EXIF)
Flight parameters	Locations	Residential Area, Quezon City	University Area, Quezon City
	Motion type	straight flight, sinusoidal altitude, sinusoidal horizontal	lawnmower
	Capture height	100 meters	30-50 meters
	Flight speed	10-15 m/s	2-5 m/s (auto 98% front overlap)

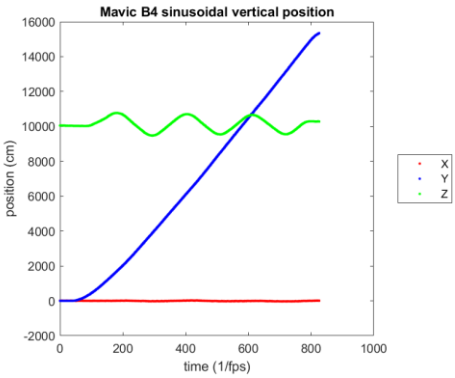
4. RESULTS AND DISCUSSION

4.1 Image mosaics

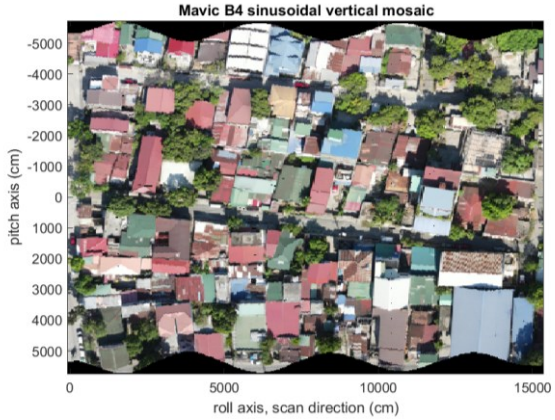

Mavic B1 straight pass position
position (cm)
time (1/fps)
X
Y
Z

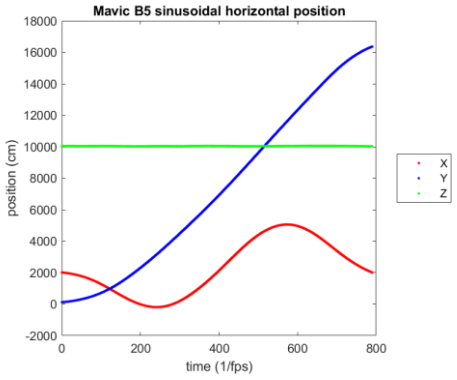
a-1) a-2)


Mavic B1 straight pass mosaic
pitch axis (cm)
roll axis, scan direction (cm)

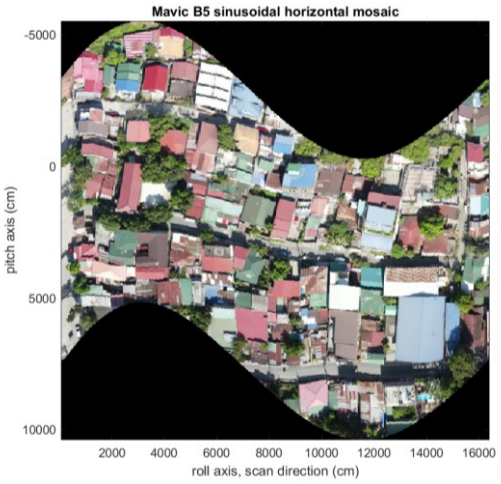

Mavic B4 sinusoidal vertical position
position (cm)
time (1/fps)
X
Y
Z

b-1) b-2)


Mavic B4 sinusoidal vertical mosaic
pitch axis (cm)
roll axis, scan direction (cm)

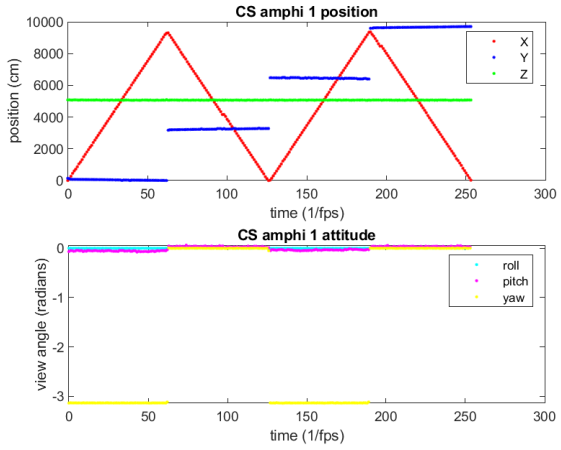


c-1)



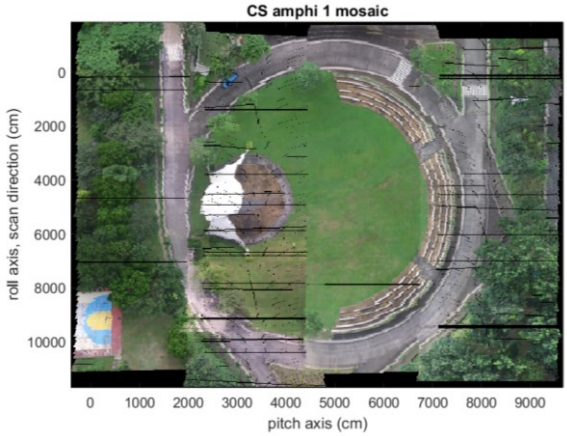
c-2)

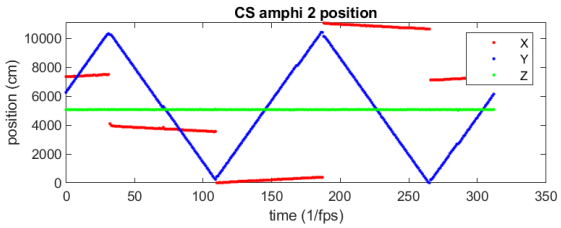
Figure 2. Mosaics from Set-up 1 (residential area) from Mavic Air 2. Left images are XYZ positions over time. Right images are corresponding mosaics . a) Straight pass, b) Sinusoidal vertical, c)Sinusoidal horizontal.

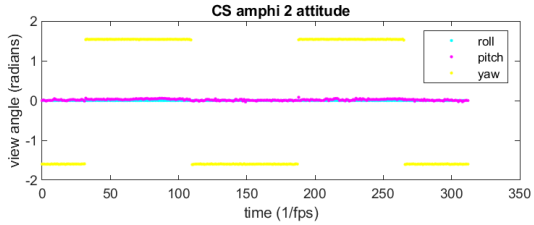


a-1)

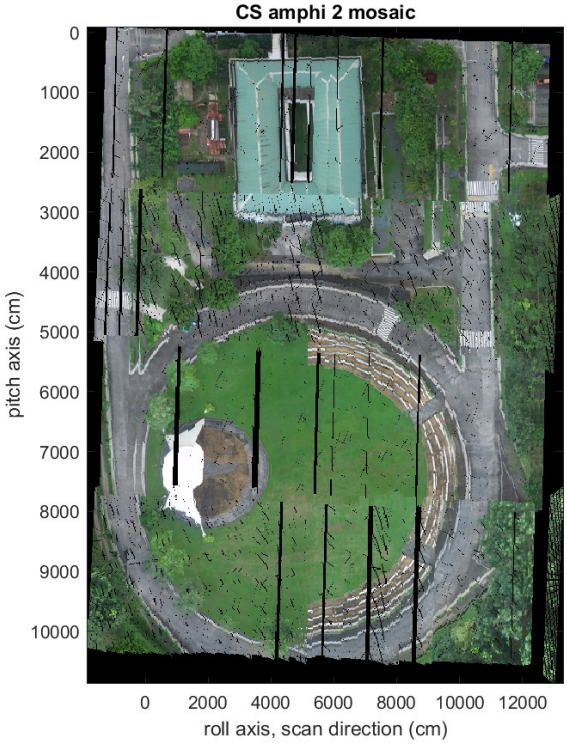
a-2)

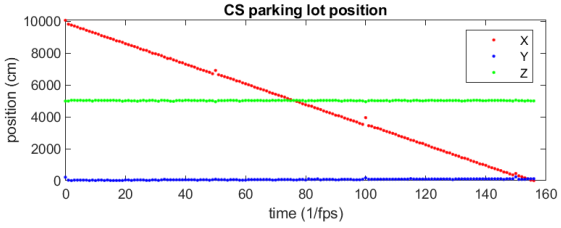


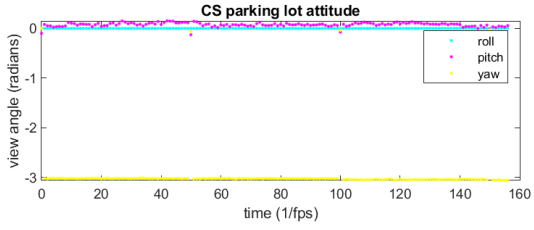

CS amphi 2 position
position (cm)
time (1/fps)
X
Y
Z


CS amphi 2 attitude
view angle (radians)
time (1/fps)
roll
pitch
yaw

b-1) b-2)


CS amphi 2 mosaic
pitch axis (cm)
roll axis, scan direction (cm)


CS parking lot position
position (cm)
time (1/fps)
X
Y
Z


CS parking lot attitude
view angle (radians)
time (1/fps)
roll
pitch
yaw

c-1) c-2)

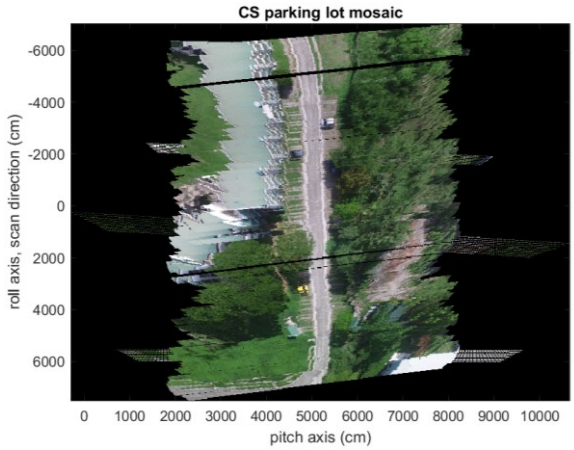


Figure 3.Mosaics from Set-up 2 (university campus area) from DJI Matrice 600 Pro. Left images are XYZ positions over time. Right images are corresponding mosaics . a-b) Lawnmower capture, c) Straight pass.

Figure 2 shows the results of the residential area capture. Since there is no rotation data, the capute is limited to single camera orientation assumptions (zero roll, zero pitch). The yaw is estimated from the first and last positions in the GPS. Figure 2a shows a standard single pass. Figure 2b shows a path with sinusoidal altitude. changes in altitude, effectively image distance, changes the swath of the capture, and can be seen by the sinusoidal increase and decrease in the boundaries of the mosaic. Figure 2c shows a sidewinding flight path. To fill in gaps to detect alignment, 4 pixel lines were taken per image.

Figure 3 shows the results of the university area capture. With rotational data available, changes in direction were attempted. Figure 3a-b shows lawnmower paths. Figure 3c shows a straight path. To fill in gaps to detect alignment, 10 pixel lines were taken per image.

For all captures, minimum possible speed (within safe battery life limits) was set to ensure that gaps are minimized. The stitched images were generated from source images scaled down to 20% for Set-up 1, and 10% for Set-up 2. This is done to speed up processing and analysis. Camera matrices were scaled accordingly.

4.1 Feature alignment, orthogonality and comparisons to known lengths

An obvious performance metric of a stitching algorithm is how well it can preserve details of the target area. The images in Figure 2 and Figure 3 were inspected in terms of alignment of straight lines, and orthogonality of perpendicular lines. The presence of roof areas provided an abundance of such corners in the residential area, while the seats and road paints provided for the university campus captures.

For the residential data, orthogonality can deviate by 1 pixel, or 27.7 cm over long measurements. This is due to the inherent skew added by the yaw estimation . The dimensions of visible edges were also compared to Google Earth captures, with a maximum difference of 2 pixels, or 55.4 cm. Note that such variations are expected in scaled down resolution, i.e. 1 pixel should be 5.54 cm in an ideal 1 line per frame mosaic. Also the assumption of the camera model is that the target points are all coplanar, therefore the presence of buildings and trees are significant deviations in proportion to capture heights, which were measured from take-off altitude.

There are many rotation and positional errors in the university campus data. While sizes and overall dimensions were cleared, there are many detail misalignments. Figures 3a-1and 3b-1 show noisy rotational data over time. There are significant misalignments visible between passes of the lawnmower. Figure 3c-1 also has noisy rotational components, resulting in a heavy skew that affects orthogonality, and periodic strip yaw exaggerations.

4.2 Errors and Assessments

The residential data set was performed in a relatively stable weather condition, and thus camera orientation variations are very minimal. It is evident in the alignment and image quality that the mosaicing sufficiently recreated the scene, even with variations in horizontal and vertical translations. However, upon closer inspection of Figures 2a-2 and 2b-2, the terminal ends are slightly warped . This can be accounted for by the flight mechanism of the Mavic Air 2. The UAV has a forward pitch when accelerating, and backwards pitch when decelerating. This causes greater apparent motion in between frames even when there is smaller displacement. Similarly, Figure 2c-2 has prominent warps in regions of changes in lateral direction. The UAV has to roll left and right to generate the horizontal net-force. These

warps can be solved by using a UAV that can output its rotational data.

Figure 3a-2 and 3b-2 showed that the algorithm can handle 180 degree turns if given rotation data. However, the built-in GPS cannot handle sharp changes in position, such as changing passes in lawnmower captures. This reflects in feature misalignments.

Another visual artifact is the presence of sharp discontinuities in adjacent pixel color values in Figures 3a-2 and 3b-2. Note the differences in color of the cement road and grass between passes. This is attributed to changing cloud cover during capture, as the entire capture lasted about 20 mins, with the UAV moving at approximately 2 m/s. At this speed, the UAV is also more prone to occasional gusts of wind, which were present during the capture.

The noisy rotation data of the campus data set features overshoots, making the otherwise straight edges appear jagged. This indicates poor sensor response from the gimbal and IMU. The effects are very prominent in Figure 3c, where there was a possible calibration error, adding a yaw offset value. Notice in Figure 3c-1, at every 50 frames, there is a spike in the X position and yaw value, resulting in bad relative placement of corresponding strips in Figure 3c-2.

As mentioned, the mosaics are compensated for the lower frame rate of the data set. Figure 4 shows the effect of using only one line with insufficient frame rate for a given flight speed. Visible blank gaps appear for the spatial coordinates where there is no corresponding frame directly under. Also noticeable are the curved lines of NaN values, artifacts that are caused by digitization and incorrect rotation values.

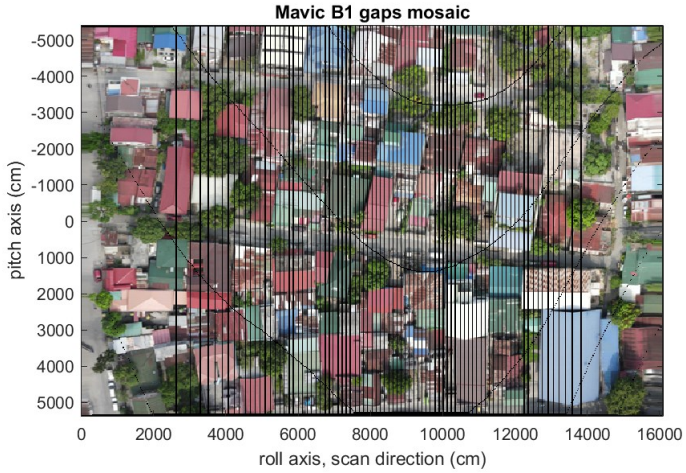


Figure 4. Residential data at 1 line per frame.

5. CONCLUSIONS

The stitching algorithm was assessed as successful in creating an image mosaic of line-scan captures. Generally, feature alignments and measurements were preserved if the flights meet the stability requirements. The algorithm, however, is sensitive to the quality of the positional sensor inputs, (GPS and IMU), and the temporal resolution of the data.

Poor quality rotational data can lead to significant misalignments and wrong reprojections, and in the case of this study, ended up being more detrimental due to overshoots from otherwise stable flights. Likewise GPS quality can lead to similar problems, more prominently in terms of alignment, especially in multiple pass flights. Possible ways to remedy the errors in future studies include improvements in sensor quality, signal processing such as noise reduction, and better environmental conditions.

The temporal resolution of the data is important in ensuring the presence of sufficient image points. The data presented in the paper compensated by taking more than one line in each image to check for local alignments. Having low framerates relative to the UAV flight speed can lead to gaps in the mosaic space.

Note that resolution values are expected to decrease once the actual hyperspectral imager is used for the aerial capture, as the presence of a slit can lead to blurring due to the convolution of the slit width with the imaging system. Also, the exposure time of the capture should be high enough to get enough light, but low enough to avoid significant blurring.

6. ACKNOWLEDGEMENTS

This study was done under the Sustained Support for Local Space Technology & Applications Mastery, Innovation and Advancement (STAMINA4SPACE) Program- Project 1: Optical Payload Technology, In-depth Knowledge Acquisition, and Localization (OPTIKAL) of the Department of Science and Technology. The researchers also acknowledge the contribution of drone operators, Engr. Lorenz Toleos and Engr. Marck Lorenz Taguse who performed the aerial captures for this study.

7. REFERENCES

Bay, H., et al, 2006. A multi-sensor robotic platform for ground mapping and estimation beyond the visible spectrum. European Conference on Computer Vision, pp. 404-417.

Chilton, A., 2013. The Working Principle and Key Applications of Infrared Sensors, Retrieved July 7, 2020, from https://www.azosensors.com/

Ferwerda, J.G., 2005. Charting the quality of forage: measuring and mapping the variation of chemical components in foliage with hyperspectral remote sensing. ITC Dissertation, Wageningen University, Wageningen, Netherlands.

Ghosh, D. and Kaabouch, N., 2016. A survey on image mosaicing techniques. Journal of Visual Communication and Image Representation, 34, pp. 1-11.

Hagen, N., 2013. Review of snapshot spectral imaging technologies. Optical Engineering, 52 (9), pp. 47-56.

Hartley, J. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, pp. 87-127.

Hu, J. and Lanzon, A., 2018. An innovative tri-rotor drone and associated distributed aerial drone swarm control. Robotics and Autonomous Systems, 103, pp. 162-174.

Lowe, D., 1999. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision, 2, pp. 1150-1157.

Reina, G., 2018. A multi-sensor robotic platform for ground mapping and estimation beyond the visible spectrum. Precision Agriculture, 20 (2), pp. 423-444.

Shawt, G.A. and Burke, H. K., 2003. Spectral imaging for remote sensing. Lincoln Lab, 14, pp.3-28.

Sharma, A., et al, 2020. Communication and networking technologies for UAVs: A survey. Journal of Network and Computer Applications, 168, pp. 102739.

Suomanlainen, J., et al, 2014. A lightweight hyperspectral mapping and photogrammetric processing chain for unmanned aerial vehicles. Remote Sensing, 6, pp. 11013-11030.