ShadowSense: Unsupervised Domain Adaptation and Feature Fusion for Shadow-Agnostic Tree Crown Detection from RGB-Thermal Drone Imagery (Supplementary Material)

Rudraksh Kapil Seyed Mojtaba Marvasti-Zadeh Nadir Erbilgin* Nilanjan Ray*
University of Alberta, Canada
{rkapil, seyedmoj, erbilgin, nray1}@ualberta.ca
*Equal Contribution

In this supplementary material, we i) present an extended ablation study to support the design choices for the proposed ShadowSense model, ii) visualize additional qualitative results to compare our ShadowSense with the baseline RGB-only detector, and iii) provide additional information about the proposed dataset *RT-Trees*.

A. Extended Ablation Study

We performed comprehensive ablations to support the selection of hyperparameters. The quantitative results obtained with different values within the proposed method on the validation set are presented in Table 1. This configuration of the ShadowSense model represents the best-performing combination of hyperparameters. Each configuration's performance was assessed independently while setting all others to their best-performing values. Different distributions for the FPN alignment scales β and fusion scaling weight η were tested. In both cases, a descending order of values (largest to smallest FPN feature map) with medium variance was found to perform the best in terms of all three metrics considered. In general, assigning higher scales to the smaller feature maps (ascending order) performed worse than when using a descending order of scales, since smaller feature maps are of a lower resolution. The classic image masking procedure we use to generate binary masks relies on watershed segmentation [2, 3]. The performance of this mask generation process depends on the initial choice of thresholds for the defined BG/FG markers, and we found (20,100) to be the best choice among the alternatives. These alternatives had either less or more differences between the thresholds, which made it more difficult for the algorithm to correctly determine the marker for pixels with intensities in between. Thermal weight λ is used during inference to compute the weighted average of the FPN feature maps from the RGB and adapted thermal branches. Assigning a lower weight to thermal features than RGB features did not help enhance the identification of shadowed trees, though foreground performance is somewhat satisfactory. As the thermal weightage was increased, shadowed tree performance increased and eventually leveled off. Using a weight of 5 was found to be the best choice, which further increases the performance on foreground trees. NMS values around 0.10 were tested, according to the recommended default value in the baseline RGB-only detection model we considered [4], and 0.15 yielded the best performance.

B. Additional Qualitative Results

These are visualized in Fig. 1 for different scenarios with challenging illumination conditions, and demonstrate the advantage of the proposed method over the RGB-only baseline for shadowed trees in the background regions while maintaining the baseline's good performance on foreground regions.

C. RT-Trees Dataset

In this section, we describe the collection, pre-processing, and annotation procedures employed for the RT-Trees dataset. Furthermore, we showcase instances of challenging scenarios within our dataset to provide a broader perspective. A multi-camera DJI H20T sensor instrument was used to simultaneously acquire wide-angle RGB images of 4056×3040 pixels with an 82.9° display field of view (DFOV) and thermal images in the 8-14μm spectral band of 640×512 pixels with 40.6° DFOV. The RGB camera uses a 1/2.3C̈MOS (12 MP) sensor, while the thermal camera uses an uncooled VOx microbolometer sensor. The H20T also contains a zoom RGB camera but it is not used in this work. The sensor instrument was mounted to a Matrice 300 RTK drone and successive image pairs were captured with an 80% front and 75% side overlap in the thermal images via a fixed flight path.

Thermal images were first upscaled through bilinear interpolation to 1500×1000, while RGB images were center-

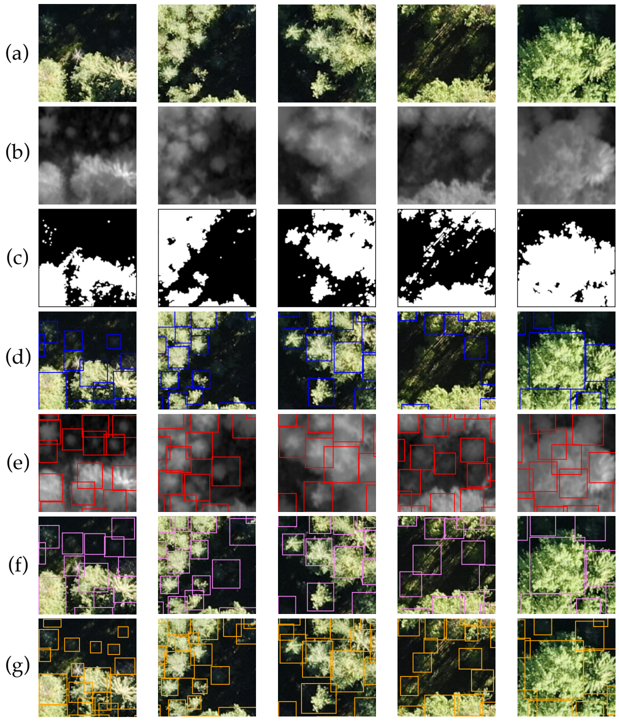


Figure 1. **Additional Detection Results**. Each column shows (a) RGB image, (b) Thermal image, (c) Generated mask; and predictions by (d) Baseline [4], (e) Our DAT-adapted thermal branch, (f) Proposed ShadowSense, and (e) Ground truth.

cropped to 1500×1000 pixels to discard edge distortions. This cropping size also ensures that both images in a pair display roughly the same amount of area. We then precisely co-registered each RGB-thermal pair using the normalized gradient fields-based workflow described in [1]. Fig. 7 shows an example pair of 1500×1000 RGB-thermal images from each of the 14 flights, highlighting the presence of varying illumination and climatic conditions within RT-Trees. Table 2 reports the flight start time, flight duration, sun elevation, sun azimuth, and air temperature at the time of each drone flight, along with details on the number of images captured and processed for training. The variation in weather and lighting conditions caused due to different sun positions once again highlights the challenge of RT-Trees. Similarly, Fig. 2 shows the image brightness (L) averaged across all pixels in LAB color space, distinguished by flight date. Images from flights later in the day (e.g., October 7) are typically darker than those taken closer to noon due to a lower sun position. The exception to this is November 24, where the significant presence of white snow cover is inflating the average brightness level. For evaluation purposes, we split the imaged area of one of the flights (August 30), reserving around 25% for the test set, 5% for the validation set, and using the rest for training, as illustrated in Fig. 3. Images from the same training area from all other flights were included in the training set, while images from the testing and validation areas were discarded, leading to roughly 70% of the total captured images being used for training. To allow for higher batch sizes during training, we first split each training area image (RGB and thermal) into six 500×500 patches, retaining the high degree of overlap between neighboring images. On the other hand, only the central 500×500 patch was considered for each image in the evaluation sets, and we sampled every third image in the capture sequence from these sets to eliminate overlap, resulting in 10 patches for validation and 63 for testing. Only non-overlapping images from a single flight date are included in the testing and validation sets to ensure that each tree only appears in one image in those sets so that detection performance is not overestimated. On the other hand, the inclusion of overlapping imagery from multiple dates helps promote diversity in the training set. This explains the seemingly large disparity in the number of training and validation/testing images.

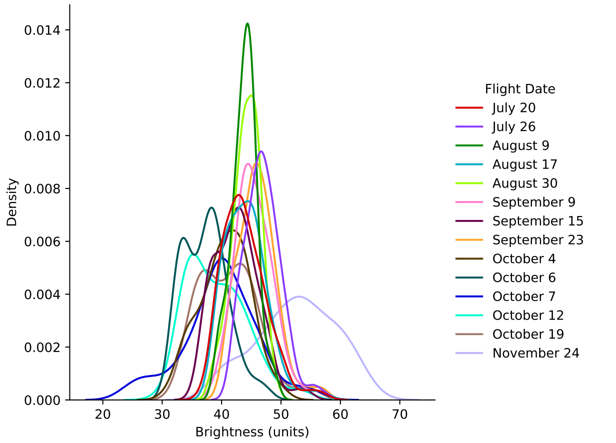


Figure 2. **Kernel Density Estimation Plot for Average Brightness** for images in each flight date, mapped first to LAB color space.

Next, image pairs in the testing and validation sets were annotated with bounding boxes in a two-step manner. First, visible tree crowns were delineated from the RGB image through careful inspection. Then, the corresponding registered thermal image was used to identify shadowed tree crowns that had been missed in the first step – these new boxes were marked as *“difficult”*. In total, 447 out of the 3611 tree crowns in the testing set were marked as difficult. We provide additional statistics about RT-Trees using the annotations for the testing set. In general, the difficult boxes were fewer in number than non-difficult (visible) ones (see Fig. 4) and were of a smaller area. This is because shadowed trees are typically shorter and smaller than their neighbors (Fig. 5). Although the annotated bounding boxes were primarily square (1:1 linear relation in Fig. 6), a considerable number of rectangular boxes are present in the testing set due to the presence of different species with non-circular crowns, partial tree crowns at the edges of images, and overlapping canopies in the densely forested region, another challenge posed by the proposed dataset.

References

[1] Rudraksh Kapil, Guillermo Castilla, Seyed Mojtaba Marvasti-Zadeh, Devin Goodsman, Nadir Erbilgin, and Nilanjan Ray. Orthomosaicking thermal drone images of forests via simultaneously acquired RGB images. *Remote Sensing*, 15(10):2653, 2023. 2

[2] Pierre J. Soille and Marc M. Ansoult. Automated basin delineation from digital elevation models using mathematical morphology. *Signal Processing*, 20(2):171–182, 1990. 1

[3] Richard Szeliski. *Computer Vision Algorithms and Applications 2nd Edition*. Springer London, 2021. 1

[4] Ben G. Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11):1309, June 2019. 1, 2

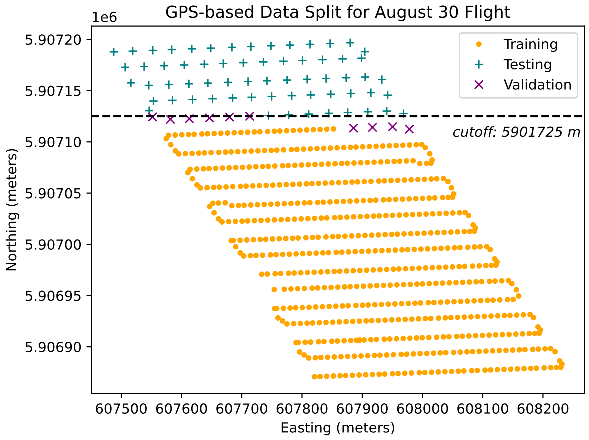


Figure 3. **GPS-based Data Split**. Each point represents the location where a drone image was taken. The assigned cutoff line separates the testing area from the training (and validation) area.

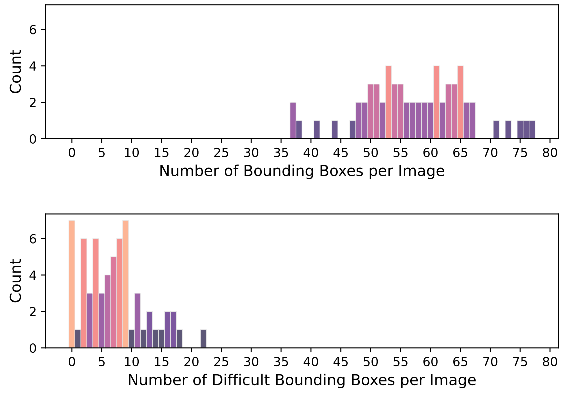


Figure 4. **Distribution of Bounding Boxes per Image** for all boxes (top) and only difficult boxes (bottom) in the testing set.

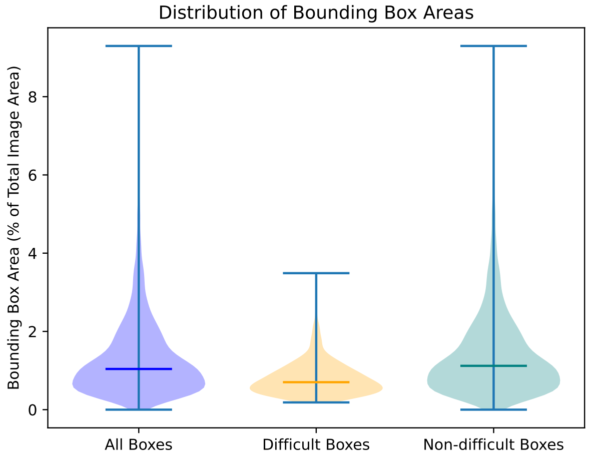


Figure 5. **Distribution of Bounding Boxes Areas** for all boxes, difficult boxes only, and non-difficult boxes (i.e., visible in RGB image) in the testing set.

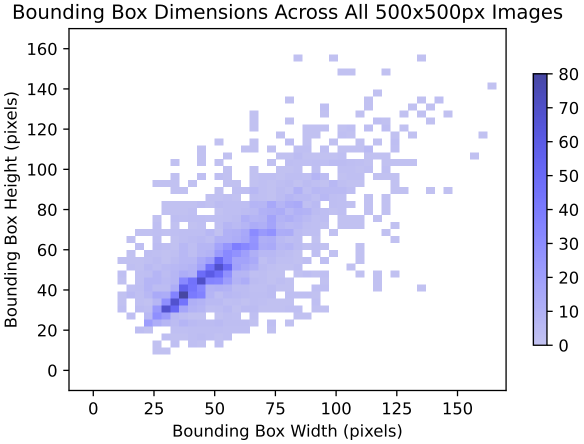


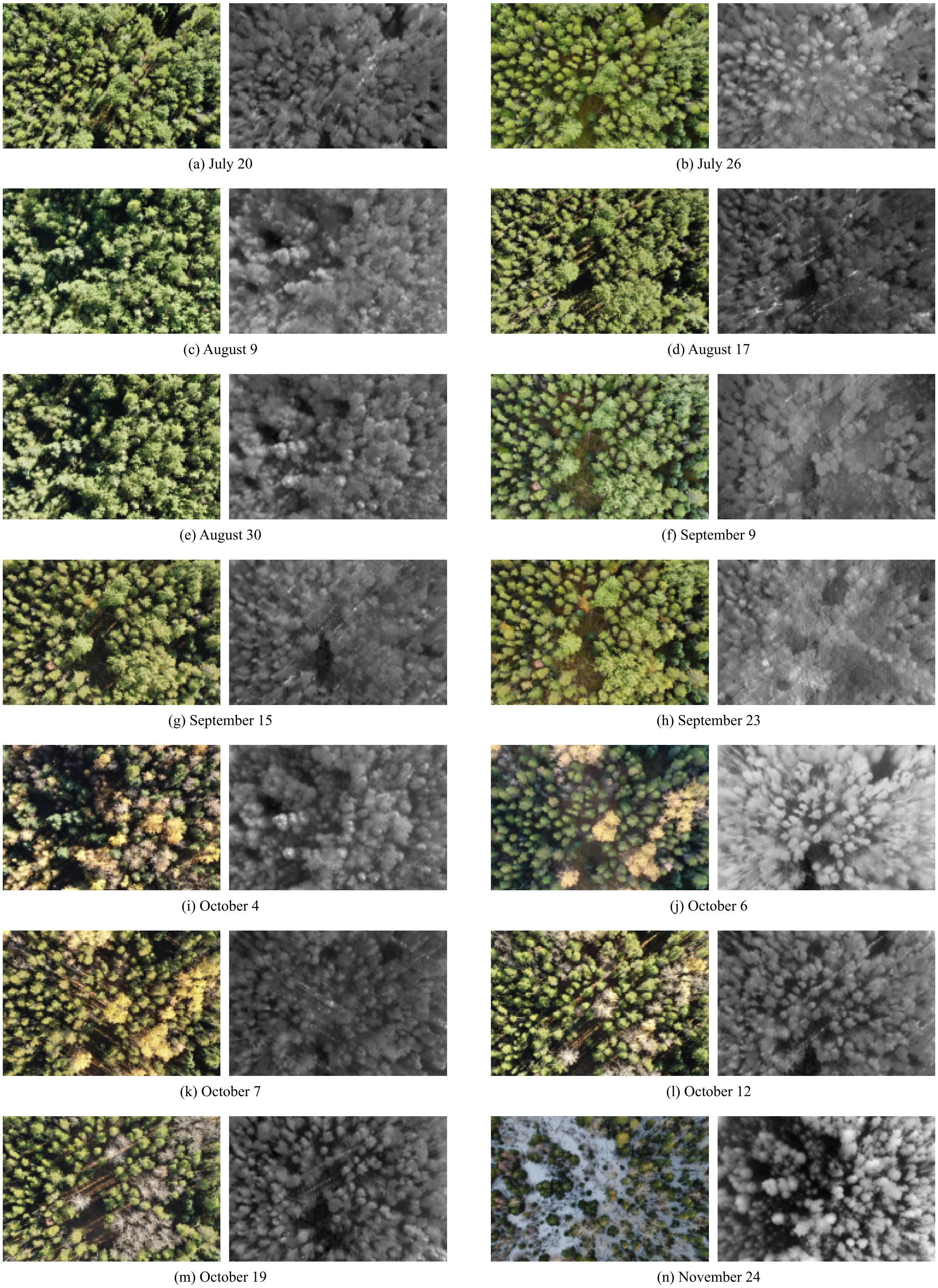
Figure 6. **Distribution of Bounding Boxes Dimensions** for all boxes in the testing set.

Table 1. **Extended Ablation Study** for different hyperparameter settings in the proposed method based on AP50 and AP100 metrics, trained without annotation on the RT-trees training set. Results on the RT-Trees validation set are reported. While changing one hyperparameter, all others are set to their best-performing values as described in the implementation details for the proposed ShadowSense configuration (also emboldened here).

Hyperparameter	**Value(s)**	**All Trees**		**Shadowed Trees**
		% AP50 (↑)	**% AR100** (↑)	**% Identified** (↑)
FPN Alignment Scales β In order of largest (lowest) to smallest (highest) FPN level.	[1.0, 1.0, 1.0, 1.0, 1.0] Identical Weighting	54.02	24.73	19.20
	[1.0, 1.0, 0.75, 0.5, 0.25] Descending w/ Low Variance	54.20	24.97	15.20
	[1.0, 1.0, 0.5, 0.05, 0.01] Descending w/ Med. Variance	**55.48**	**25.86**	**20.80**
	[1.0, 0.5, 0.1, 0.01, 0.005] Descending w/ High Variance	53.20	24.55	20.00
	[0.25, 0.5, 0.75, 1.0, 1.0] Ascending w/ Low Variance	50.96	24.70	12.80
	[0.01, 0.05, 0.5, 1.0, 1.0] Ascending w/ Med. Variance	50.21	24.61	13.60
	[0.005, 0.01, 0.1, 0.5, 1.0] Ascending w/ High Variance	48.49	23.41	8.00
Fusion Scaling Weights η In order of largest (lowest) to smallest (highest) FPN level.	[1.0, 1.0, 1.0, 1.0, 1.0] Identical Weighting	54.62	24.28	19.60
	[1.0, 1.0, 0.8, 0.6, 0.4] Descending w/ Low Variance	54.84	24.88	**20.80**
	[1.0, 1.0, 0.5, 0.2, 0.2] Descending w/ Med. Variance	**55.48**	**25.86**	**20.80**
	[1.0, 0.5, 0.2, 0.05, 0.01] Descending w/ High Variance	51.71	23.00	20.00
	[0.4, 0.6, 0.8, 1.0, 1.0] Ascending w/ Low Variance	52.09	22.75	17.60
	[0.2, 0.2, 0.5, 1.0, 1.0] Ascending w/ Med. Variance	50.86	22.26	16.00
	[0.01, 0.05, 0.2, 0.5, 1.0] Ascending w/ High Variance	48.81	21.07	10.40
Intensity Thresholds Min. and max. markers used in mask generation.	[30, 75] Low number of initially unmarked intensities	49.86	21.74	8.00
	[20,100] Medium number of initially unmarked intensities	**55.48**	**25.86**	**20.80**
	[10,125] High number of initially unmarked intensities	51.67	22.23	18.40
Thermal Weight λ_T Used in weighted fusion during inference.	0.5 Higher weighting for RGB features	50.57	22.65	13.80
	1.0 Identical weighting for RGB and thermal features	50.31	22.03	16.00
	2.5 |	52.61	23.08	18.40
	5.0 ↓	**55.48**	**25.86**	**20.80**
	7.5 Higher weighting for thermal features	52.01	22.86	**20.80**
Non-max Suppression NMS threshold used in training and inference.	0.05 Less overlap in filtered predictions	49.15	22.25	17.20
	0.10 |	50.50	23.30	16.00
	0.15 ↓	**55.48**	**25.86**	**20.80**
	0.20 More overlap in filtered predictions	50.60	23.97	17.20

Table 2. **RT-Trees Dataset Information by Flight Date**. All dates are from the year 2022. Information about the flight, lighting and weather conditions, and number of images is listed. Approximately 70% of the raw image pairs captured for a given date are sampled for the training set based on GPS location (see Fig. 3), and then divided into six 500×500 patches. From the August 30 data, 63 images are taken for testing and 10 for validation, hence the total number of image pairs in RT-Trees is 49879.

Flight Date	**Time of First Capture (24h)**	**Flight Duration (min)**	**Sun Elevation** (°)	**Sun Azimuth** (°)	**Air Temperature** ($^\circ C$)	**Raw Image Pairs Captured**	**500×500 Patches in Training Set**
July 20	11:04	27	44.77	120.15	20.3	827	3582
July 26	10:18	28	37.51	109.32	20.8	828	3588
August 9	10:16	28	34.45	111.71	19.8	820	3552
August 17	12:15	33	46.12	147.29	24.5	825	3570
August 30	11:21	31	37.26	134.21	25.4	814	3516
September 9	11:40	32	36.08	142.39	14.0	824	3570
September 15	11:00	27	30.20	133.15	17.5	825	3582
September 23	11:14	28	29.12	139.04	14.7	820	3552
October 4	11:13	31	25.43	141.56	16.8	820	3558
October 6	15:16	27	27.22	210.10	9.8	808	3498
October 7	19:02	27	0.26	261.26	2.8	819	3552
October 12	10:53	27	20.92	138.40	11.5	826	3576
October 19	11:35	27	22.29	150.20	12.4	821	3558
November 24	16:10	28	2.21	230.46	4.7	819	3552
Total						**11496**	**49806**



(a) July 20

(b) July 26

(c) August 9

(d) August 17

(e) August 30

(f) September 9

(g) September 15

(h) September 23

(i) October 4

(j) October 6

(k) October 7

(l) October 12

(m) October 19

(n) November 24

Figure 7. **Example of Drone-collected Image Pairs** for each flight date after performing co-registration.