

# Through-Foliage Tracking with Airborne Optical Sectioning

Rakesh John Amala Arokia Nathan<sup>a</sup>, Indrajit Kurmi<sup>a</sup>, David C. Schedl<sup>b</sup> and Oliver Bimber<sup>a,\*</sup>

<sup>a</sup>Johannes Kepler University Linz, Altenberger Straße 69, Linz, 4040, Austria

<sup>b</sup>University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg Austria, 4232, Austria

## ARTICLE INFO

### Keywords:

synthetic aperture imaging  
anomaly detection  
occlusion removal  
tracking

## ABSTRACT

Detecting and tracking moving targets through foliage is difficult, and for many cases even impossible in regular aerial images and videos. We present an initial light-weight and drone-operated 1D camera array that supports parallel synthetic aperture aerial imaging. Our main finding is that color anomaly detection benefits significantly from image integration when compared to conventional single images or video frames (on average 97% vs. 42% in precision in our field experiments). We demonstrate, that these two contributions can lead to the detection and tracking of moving people through densely occluding forest.

## 1. Introduction

With Airborne Optical Sectioning (AOS, Kurmi, Schedl and Bimber (2018, 2019a,b, 2020, 2021b,a); Bimber, Kurmi and Schedl (2019); Schedl, Kurmi and Bimber (2020a,b, 2021)) we have introduced a wide synthetic aperture imaging technique that employs conventional drones to sample images within large areas from above forest. These images are computationally combined (registered to the ground and averaged) to integral images which suppress strong occlusion and make visible hidden targets. AOS relies on a statistical chance that a point on the forest ground is unoccluded by vegetation from multiple perspectives, as explained by the statistical probability model in Kurmi et al. (2019a). The integral images can be further analyzed to support, for instance, automated person classification with advanced deep neural-networks. In Schedl et al. (2020b), we have shown that integrating single images before classification rather than combining classification results of single images is significantly more effective when classifying partially occluded persons in aerial thermal images (92% vs. 25% average precision). In Schedl et al. (2021), we demonstrate a first fully autonomous drone for search and rescue based on AOS. The main advantages of AOS over alternatives, such as LiDAR or Synthetic Aperture Radar, are its real-time performance; its applicability to other wavelengths, such as far infrared for wildlife observations and search and rescue, or near infrared for agriculture and forestry applications; and its high spatial resolution.

Thus far, the sequential sampling nature of AOS, limits its applications to static targets only. Moving targets, such as walking people or running animals lead to motion blur in integral images that are nearly impossible to detect and track. Applying AOS to far infrared (thermal imaging), as in Schedl et al. (2020b, 2021), restricts it to cold environment

temperatures, while using it in the visible range (RGB imaging) as in Kurmi et al. (2018), often suffers from too little sunlight penetrating through dense vegetation.

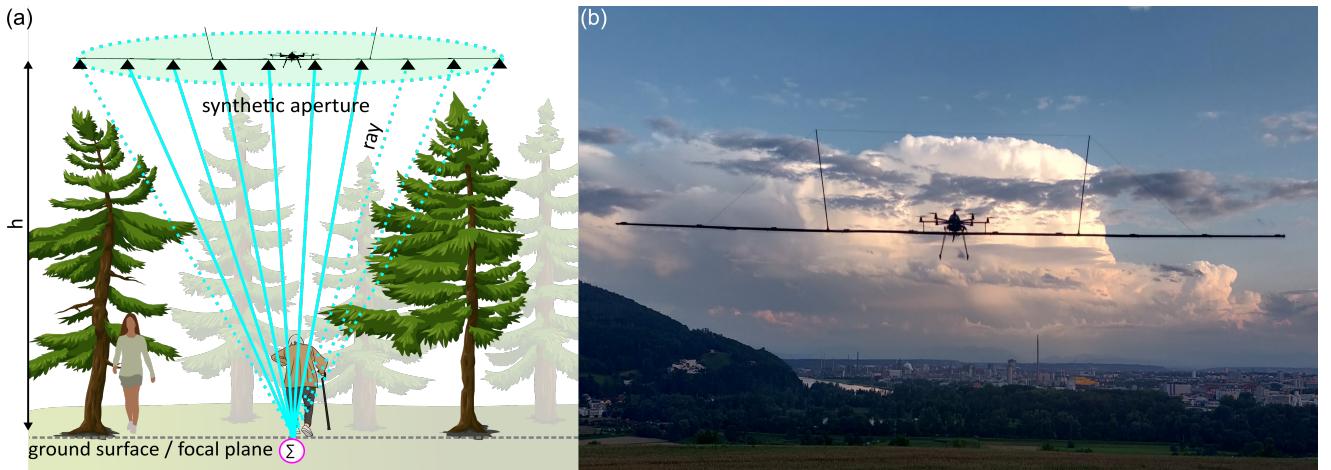
In this article we make two main contributions: First, we present an initial light-weight (<1kg) and drone-operated 1D camera array that supports parallel AOS sampling. While 1D and 2D camera arrays have already been used for implementing various visual effects (e.g., Vaish, Wilburn, Joshi and Levoy (2004); Vaish, Levoy, Szeliski, Zitnick and Kang (2006); Zhang, Jin and Dai (2018); Yang, Ma, Wang, Li, Yu and Zhang (2016); Joshi, Avidan, Matusik and Kriegman (2007); Pei, Zhang, Chen and Yang (2013); Yang, Zhang, Yu, Li, Ma, Tong, Yu and Ran (2014)), they have not been applied for aerial imaging (in particular with drones) because of their size and weight. Second, we show that color anomaly detection (e.g., Reed and Yu (1990); Ehret, Davy, Morel and Delbracio (2019)) benefits significantly from AOS integral images when compared to conventional single images (on average 97% vs. 42% in precision). Color anomaly detection is often used for automatized aerial image analysis in search and rescue applications (e.g., Morse, Thornton and Goodrich (2012); Agcayazi, Cawi, Jurgenson, Ghassemi and Cook (2016); Weldon and Hupy (2020)) because of its independence to environment temperature (in contrast to thermal imaging). However, it fails in presence of occlusion. We demonstrate, that these two contributions can lead to the detection and tracking of moving people through dense forest.

## 2. Integral Imaging with Airborne Camera Arrays

As illustrated in Fig. 1(a), our new payload captures multiple aerial images with a drone-operated 1D camera array. It samples the forest in parallel at flying altitude ( $h$ ) within the range of a synthetic aperture (SA) that equals the size of the camera array. This results in a structured 3D light-field formed by image pixels which are represented as light rays in a 3D volume, as discussed in Wetzstein, Ihrke, Lanman and Heidrich (2011); Wu, Masia, Jarabo, Zhang, Wang, Dai, Chai and Liu (2017). With known camera

\*Corresponding author

 rakesh.amala\_arokia\_nathan@jku.at (R.J.A.A. Nathan);  
 indrajit.kurmi@jku.at (I. Kurmi); david.schedl@fh-hagenberg.at (D.C. Schedl); oliver.bimber@jku.at (O. Bimber)  
ORCID(s): 0000-0001-5534-0182 (R.J.A.A. Nathan);  
0000-0001-7065-0509 (I. Kurmi); 0000-0002-7621-3526 (D.C. Schedl);  
0000-0001-9009-7827 (O. Bimber)



**Figure 1:** Airborne Optical Sectioning (AOS) captures multiple images within a synthetic aperture that, in this work, is defined by the size of the camera array (a). For occlusion removal, both –occluded and unoccluded rays (dashed and solid lines respectively)– are integrated for a common target point on the ground surface or an approximating focal plane (resulting in one pixel of the integral image). Our current prototype with ten RGB cameras attached on a 9m wide carbon fiber tube (b).

intrinsics, camera poses, and a representation of the terrain (either a digital elevation model if available, as in Schedl et al. (2021), or a focal plane approximation if not, as in Kurmi et al. (2020)), each ray's origin on the ground can be reconstructed. The occlusion-reduced integral image can be obtained by averaging all the rays that have the same origin. Depending on occlusion density, more or fewer rays of a surface point contain information of random occluders, while others contain the signal information of the target, as shown in Fig. 1(a). Therefore, integrating multiple rays (i.e., averaging their corresponding pixels) results in focus of the target and defocus of the occluders<sup>1</sup>. This increases the probability of detecting the target reliably under strong occlusion conditions Schedl et al. (2020b).

Figure 1(b) illustrates our current prototype. The drone basis system is a MikroKopter OktoXL 6S12 octocopter. The custom built light-weight camera array is based on a truss design which can handle the forces and vibrations due to wind by safely distributing them over the entire structure. Details on the design and construction of the camera array are provided in the supplementary material. It carries ten light weight DVR pin-hole cameras (12g each), attached equidistant (1m) to each other on a 9m long detachable and hollow carbon fibre tube (700g) which is segmented into detachable sections of varying lengths and with a gradual reduction in diameter in each section from 2.5cm at the drone centre to 1.5cm at the outermost section. The cameras are aligned in such a way that their optical axes are parallel and pointing downwards. They record images at a resolution of 1600X1200 pixels and videos at a resolution of 1280X720 and 30fps to individual SD cards. All cameras receive power from two central 7.2V Ni-MH batteries and are synchronously triggered from the drone's flight controller

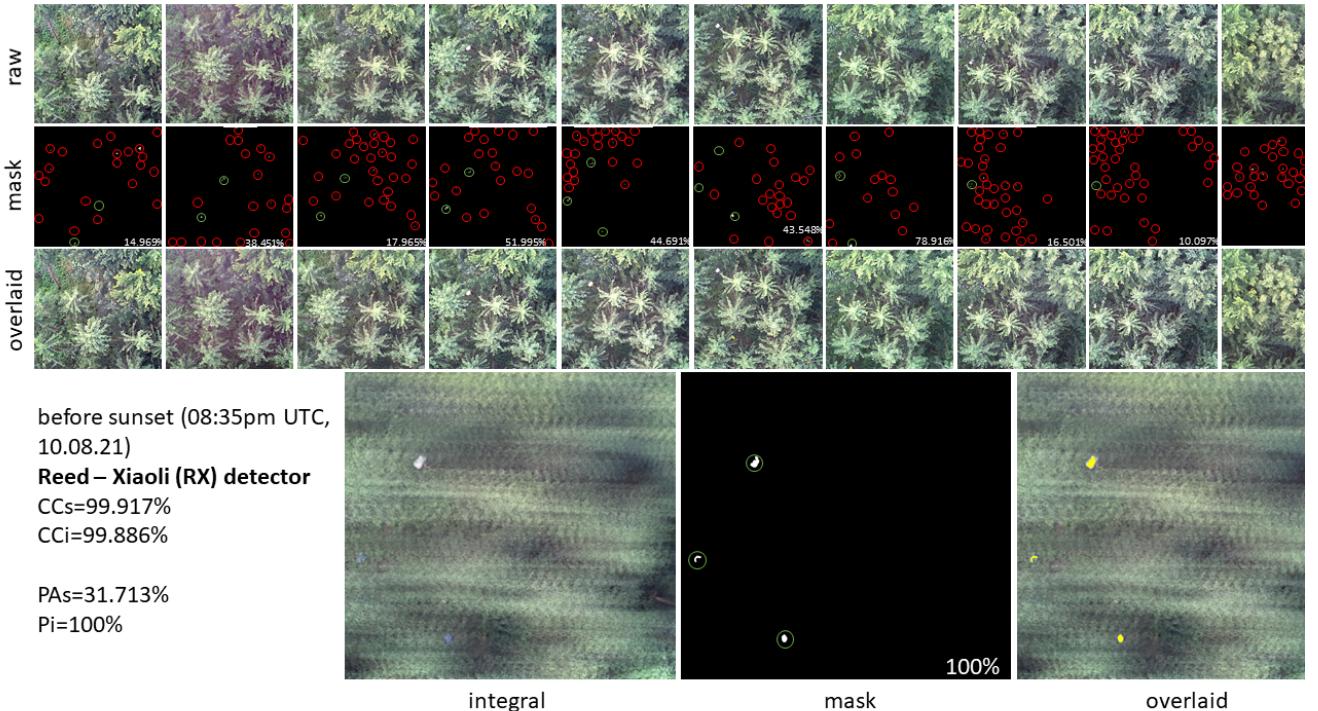
through a flat-band cable bus. To ensure stable flights with our payload, the drone's *PID* controller had to be reconfigured. *PID* stands for Proportional-Integral-Derivative which is a part of the flight control software that continuously reads the information provided by the drone's sensors and calculates how fast the motors must spin in order to retain the desired rotation speed for a stable flight. The *P* controller changes the motor power proportional to the angle of inclination and the *I* controller changes motor power continuously depending on the deflection angle and the time while the *D* controller responds to any rapid changes in the sensor data. The *PID* parameters were tuned to an *I*-dominant state (Gyro *P*: 100, Gyro *I*: 255, Gyro *D*: 10, all in a 0-255 range) for increased motor power to instantly compensate imbalances caused by external forces (such as wind) and to avoid oscillation of camera array's long lever. Supplementary videos 1 and 2 show test flights with default and with tuned *PID* parameters.

All image processing and computations (including precise pose estimation of the cameras using the photogrammetry package COLMAP (Schönberger, Zheng, Pollefeys and Frahm (2016); Schönberger and Frahm (2016))), camera intrinsic calibration using OpenCV's pinhole camera model, image rectification and integral image computation as explained in Kurmi et al. (2018); Schedl et al. (2020b) are carried out offline.

### 3. AOS Enhanced Color Anomaly Detection

Color anomaly detection finds pixels or clusters of pixels in an image with significant color differences in comparison to their neighbours. Most color anomaly detection techniques are applied to hyper-spectral or multi-spectral images (Manolakis, Marden, Shaw et al. (2003)), with the Reed-Xiaoli (RX) anomaly detector being standard and most widely used (Smetek and Bauer (2007)). The RX anomaly

<sup>1</sup>AOS source code, data, and publications are available at <https://github.com/JKU-ICG/AOS/>.



**Figure 2:** Color anomaly detection under bright light conditions (before sunset). Ten single images recorded by camera array (top three rows): raw RGB images, RX anomaly mask, mask overlaid over raw image. Red and green circles indicate false and true positives. Numbers are precision values in %. The corresponding integral image (raw, mask, and overlaid mask) is shown at the bottom row. Visible are three targets: two moving people and one static light source (largest cluster) for spatial reference.

detector (Reed and Yu (1990)) characterizes the image background in terms of a covariance matrix and calculates the RX score based on the Mahalanobis distance between a test pixel  $r$  and the background as follows:

$$\alpha_{RXD}(r) = (r - \delta)^T K^{-1} (r - \delta), \quad (1)$$

where  $r$  is the spectral vector of the pixel under test,  $\delta$  is the spectral mean vector of the background and  $K$  is the covariance matrix.

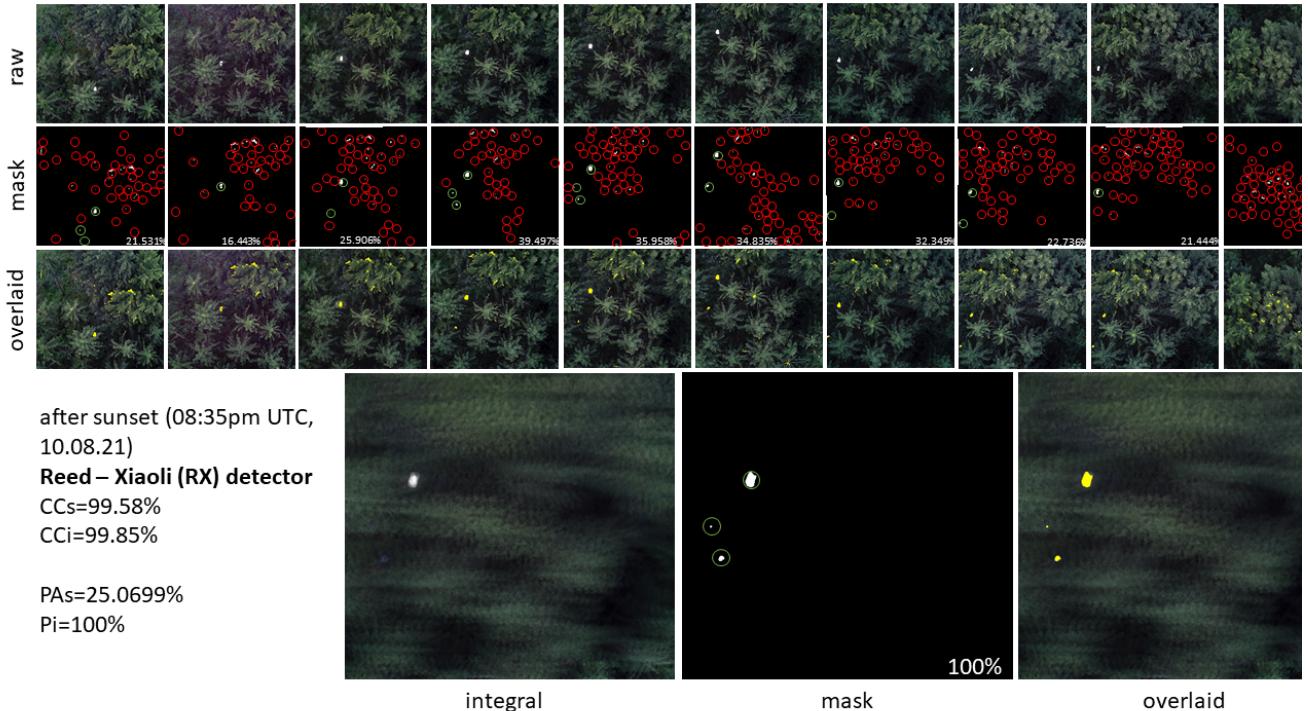
The RX scores are thresholded based on its cumulative probability distribution function. The threshold represents the RX score confidence beyond which the pixel under test is considered as an anomaly. The RX score confidence is set to maximize the detection of true positive pixels (i.e., maximum recall of the targets with a minimum one pixel) and minimize the detection of false positive pixels to attain the maximum pixel based precision value.

Figures 2 and 3 present visual results of the RX anomaly detector applied to a dataset captured with our drone prototype over dense mixed forest before and after sunset (performed in compliance with the legal flight regulations, before sunset+30min). We either apply the RX detector to the ten single images captured by the camera array individually, or to the integral images that result from registering and averaging the same ten images. For both cases, optimal thresholds (i.e., one average threshold ( $CC_s$ ) for all single images and one threshold ( $CC_i$ ) for the integral image) are found as

explained above (i.e., by maximizing true positive pixels and recall of the targets while minimizing false positive pixels). For all results, we use precision (ratio of true positive and true + false positive pixels in percent) as a quality metric. Here,  $PAs$  is the precision average over all ten single images, and  $Pi$  is the precision value of the corresponding integral image. Note, that the number of ground truth pixels (i.e., the number of target pixels under occlusion free conditions) are unknown. Note also, that the precision value is 0% in case of no true positives (i.e., target is completely occluded or not within the field of view of the camera) but in the presence false positives (i.e., wrong detections). For the case of no false positives but in the presence of true positives, the precision value is 100%.

As demonstrated in Figs. 2 and 3, color anomaly detection benefits significantly from image integration. While strong occlusion causes many false positives but little true positives in single images, false positives are almost entirely removed and true positives are significantly increased in the integral images.

We found that the mean values  $\delta$  do not significantly differ between single and integral images, since in both cases the occlusion background is dominant. The color of target pixels ( $r$ ) shifts slightly towards  $\delta$  in the integral images (due to averaging with the occluding background). The covariance values in  $K$ , however, are decreased significantly by a factor of 3 - 5 for integral images when compared to



**Figure 3:** Color anomaly detection under dark light conditions (after sunset). Ten single images recorded by camera array (top three rows): raw RGB images, RX anomaly mask, mask overlaid over raw image. Red and green circles indicate false and true positives. Numbers are precision values in %. The corresponding integral image (raw, mask, and overlaid mask) is shown at the bottom row. Visible are three targets: two moving people and one static light source (largest cluster) for spatial reference.

the covariance values of single images. This is the dominant reason for higher RX scores in the integral images, since since the RX detector multiplies with the inverse of  $K$  (Eqn. 1).

Table 1 presents quantitative results of twenty data sets captured before (SS-) and after (SS+) sunset. Here,  $P_s$  are the precision values of single images ( $C0-C9$ ),  $PAs$  the precision average over all ten images in each set, and  $P_i$  the precision value of the corresponding integral image. On average, we achieve an improvement from 42% to 97% in precision when AOS is used in combination with color anomaly detection. We did not find a significant difference of this improvement when separating the low light SS+ sets (48% vs. 97%) and bright light SS- sets (37% vs. 97%). All data sets are available (Nathan, Kurmi, Schedl and Bimber (2021)).

#### 4. Tracking Occluded Targets

By making use of the camera array's video recording functionality and by extending the AOS image integration process to compute integral videos allows the application of a multi-object tracker to the RX anomaly mask time series at recording speed (30fps in our case). For tracking, we apply a blob detector to the binary RX anomaly masks and a Kalman filter to predict the tracked blob locations in each frame.

As illustrated in Fig. 4, object tracking in integral videos performs significantly better than in regular video recordings. While many false positive targets are detected, the true positive targets are not detected or are not consistently tracked throughout regular (single) video frames. However, few (none, in the example shown in Fig. 4) false positive targets are detected and all true positive targets are mostly consistently tracked throughout the integral video frames. Supplementary videos 3 and 4 show this (two persons tracked) and an additional example (one person tracked).

The reason for the improved tracking quality is clearly the enhanced color anomaly detection in the integral video frames.

#### 5. Conclusion and Future Work

While detecting and tracking moving targets through foliage is difficult (and for many cases even impossible) in regular aerial images or videos, it becomes practically feasible with image integration – which is the core principle of Airborne Optical Sectioning. We have already shown that classification significantly benefits from image integration (Schedl et al. (2020b)). In this work we demonstrate that the same holds true for color anomaly detection. This finding together with the implementation of an initial drone operated camera array for parallel synthetic aperture aerial imaging allows presenting first results of tracking moving people

**Table 1**

Comparison of color anomaly detection result between single images and integral images before and after sunset.

flight	set no.	Ps										PAs	Pi
		C0	C1	C2	C3	C4	C5	C6	C7	C8	C9		
SS+	1	9.3	18.1	8.0	9.7	14.1	18.5	5.7	10.9	5.7	5.7	10.6	98.3
SS-	2	0	37.2	58.6	59.2	68.1	45.9	85.2	86.7	33.6	0	47.4	79.9
SS-	3	0	35.7	64.8	47.4	49.2	68.7	29.0	44.0	0	0	33.9	100
SS+	4	84.5	81.8	67.2	83.5	94.9	89.3	94.1	84.9	72.2	0	75.2	98.8
SS+	5	21.5	16.4	25.9	39.5	36.0	34.8	32.3	22.7	21.4	0	25.1	100
SS+	6	34.9	98.0	96.5	95.9	64.3	84.7	64.7	96.6	100	92.2	82.8	100
SS+	7	30.8	66.8	92.9	100	91.3	99.6	100	100	72.2	0	75.3	85.2
SS+	8	75.4	100	78.0	94.1	99.0	100	79.1	100	95.1	35.4	85.6	100
SS+	9	47.1	41.7	26.6	40.1	50.8	45.4	48.1	48.7	0	6.7	35.5	95.2
SS-	10	1.1	28.0	41.9	55.0	80.0	0	0	0	0	0	20.6	100
SS-	11	0	2.9	16.5	37.0	56.4	31.2	100	57.4	35.5	0	33.7	100
SS-	12	15.0	38.5	18.0	52.0	44.7	43.5	78.9	16.5	10.1	0	31.7	100
SS-	13	25.7	26.8	33.8	54.6	58.5	39.0	11.4	3.0	6.2	0	25.9	99.2
SS-	14	15.4	47.6	40.0	49.0	75.4	39.3	6.6	0	0	0	27.3	98.5
SS-	15	7.9	72.6	76.8	91.3	94.6	60.5	11.5	0	35.1	22.8	47.3	93.1
SS-	16	57.7	42.3	59.0	92.8	99.6	68.6	32.4	15.4	18.7	0	48.7	100
SS-	17	0	60.5	90.4	97.2	61.2	24.2	85.0	42.7	1.8	0	46.3	100
SS+	18	20.4	25.6	24.9	25.4	15.0	16.2	23.7	14.1	2.0	14.4	18.2	91.4
SS+	19	0.1	64.3	38.1	50.5	30.5	50.5	67.7	73.0	30.6	0	40.5	100
SS+	20	9.5	38.8	17.8	6.0	24.4	68.0	81.4	26.8	0	0	27.2	100
												41.9	97.0

**Figure 4:** Motion based multi-object tracking. Subset of single video frames (two top rows) and corresponding subset of integral frames (two bottom rows): raw RGB frames, RX anomaly mask and tracking labels overlaid. Visible are three targets: two moving people (labels 1 and 3) and one static light source (label 2) for spatial reference.

through dense forest. Besides people, other targets (e.g. vehicles or animals) can be detected and tracked in the same way. This might impact many application domains, such as search and rescue, surveillance, border control, wildlife observation, and others.

The utilized RX color anomaly detector and the applied combination of blob detection and Kalman filter for tracking are only implementation examples. They can be replaced by

more advanced techniques, but we believe that our main finding (i.e., anomaly detection and tracking benefit significantly from image integration) will still apply.

Color anomaly detection is clearly limited to detectable target color. In our experiments, targets were colored in white, black, blue, and red. Greenish color would have most likely not been detected. A combination of color (RGB), thermal (IR), and time (motion itself) channels for anomaly

detection might result in further improvements. This has to be investigated in future. Furthermore, the implications of parallel-sequential sampling strategies and other sampling devices, such as re-configurable drone swarms instead of camera arrays with a fixed sampling pattern, have to be explored.

## Acknowledgements

The camera array structure was designed and constructed by JKU's Institute of Structural Lightweight Design. This research was funded by the Austrian Science Fund (FWF) under grant number P 32185-NBL, and by the State of Upper Austria and the Austrian Federal Ministry of Education, Science and Research via the LIT – Linz Institute of Technology under grant number LIT-2019-8-SEE-114.

## CRedit authorship contribution statement

**Rakesh John Amala Arokia Nathan:** Software, Data curation, Writing - Original draft preparation. **Indrajit Kurmi:** Software, Data curation, Writing - Original draft preparation. **David C. Schedl:** Software, Data curation. **Oliver Bimber:** Conceptualization of this study, Methodology, Writing - Original draft preparation.

## References

- Agcayazi, M.T., Cawi, E., Jurgenson, A., Ghassemi, P., Cook, G., 2016. Resquad: Toward a semi-autonomous wilderness search and rescue unmanned aerial system, in: 2016 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 898–904.
- Bimber, O., Kurmi, I., Schedl, D.C., 2019. Synthetic aperture imaging with drones. *IEEE computer graphics and applications* 39, 8–15.
- Ehret, T., Davy, A., Morel, J.M., Delbracio, M., 2019. Image anomalies: A review and synthesis of detection methods. *Journal of Mathematical Imaging and Vision* 61, 710–743.
- Joshi, N., Avidan, S., Matusik, W., Kriegman, D.J., 2007. Synthetic aperture tracking: Tracking through occlusions, in: 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8.
- Kurmi, I., Schedl, D.C., Bimber, O., 2018. Airborne optical sectioning. *Journal of Imaging* 4, 102.
- Kurmi, I., Schedl, D.C., Bimber, O., 2019a. A statistical view on synthetic aperture imaging for occlusion removal. *IEEE Sensors Journal* 19, 9374–9383.
- Kurmi, I., Schedl, D.C., Bimber, O., 2019b. Thermal airborne optical sectioning. *Remote Sensing* 11, 1668.
- Kurmi, I., Schedl, D.C., Bimber, O., 2020. Fast automatic visibility optimization for thermal synthetic aperture visualization. *IEEE Geoscience and Remote Sensing Letters* 18, 836–840.
- Kurmi, I., Schedl, D.C., Bimber, O., 2021a. Combined person classification with airborne optical sectioning. *arXiv preprint arXiv:2106.10077*.
- Kurmi, I., Schedl, D.C., Bimber, O., 2021b. Pose error reduction for focus enhancement in thermal synthetic aperture visualization. *IEEE Geoscience and Remote Sensing Letters*.
- Manolakis, D., Marden, D., Shaw, G.A., et al., 2003. Hyperspectral image processing for automatic target detection applications. *Lincoln laboratory journal* 14, 79–116.
- Morse, B.S., Thornton, D., Goodrich, M.A., 2012. Color anomaly detection and suggestion for wilderness search and rescue, in: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 455–462.
- [dataset] Nathan, R.J.A.A., Kurmi, I., Schedl, D.C., Bimber, O., 2021. Through-foliage tracking with airborne optical sectioning. Zenodo, v1, <https://doi.org/10.5281/zenodo.5680949>.
- Pei, Z., Zhang, Y., Chen, X., Yang, Y.H., 2013. Synthetic aperture imaging using pixel labeling via energy minimization. *Pattern Recognition* 46, 174–187.
- Reed, I.S., Yu, X., 1990. Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution. *IEEE transactions on acoustics, speech, and signal processing* 38, 1760–1770.
- Schedl, D.C., Kurmi, I., Bimber, O., 2020a. Airborne optical sectioning for nesting observation. *Nature Scientific Reports* 10, 1–7.
- Schedl, D.C., Kurmi, I., Bimber, O., 2020b. Search and rescue with airborne optical sectioning. *Nature Machine Intelligence* 2, 783–790.
- Schedl, D.C., Kurmi, I., Bimber, O., 2021. An autonomous drone for search and rescue in forests using airborne optical sectioning. *Science Robotics* 6, eabg1188.
- Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M., 2016. Pixelwise view selection for unstructured multi-view stereo, in: European Conference on Computer Vision (ECCV).
- Schönberger, J.L., Frahm, J., 2016. Structure-from-motion revisited, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4104–4113.
- Smetek, T.E., Bauer, K.W., 2007. Finding hyperspectral anomalies using multivariate outlier detection, in: 2007 IEEE Aerospace Conference, IEEE. pp. 1–24.
- Vaish, V., Levoy, M., Szeliski, R., Zitnick, C.L., Kang, S.B., 2006. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE. pp. 2331–2338.
- Vaish, V., Wilburn, B., Joshi, N., Levoy, M., 2004. Using plane + parallax for calibrating dense camera arrays, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., pp. I–I.
- Weldon, W.T., Hupy, J., 2020. Investigating methods for integrating unmanned aerial systems in search and rescue operations. *Drones* 4.
- Wetzstein, G., Ihrke, I., Lanman, D., Heidrich, W., 2011. Computational plenoptic imaging, in: Computer Graphics Forum, Wiley Online Library. pp. 2397–2426.
- Wu, G., Masia, B., Jarabo, A., Zhang, Y., Wang, L., Dai, Q., Chai, T., Liu, Y., 2017. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing* 11, 926–954.
- Yang, T., Ma, W., Wang, S., Li, J., Yu, J., Zhang, Y., 2016. Kinect based real-time synthetic aperture imaging through occlusion. *Multimedia Tools and Applications* 75, 6925–6943.
- Yang, T., Zhang, Y., Yu, J., Li, J., Ma, W., Tong, X., Yu, R., Ran, L., 2014. All-in-focus synthetic aperture imaging, in: Computer Vision – ECCV 2014, Springer International Publishing, Cham. pp. 1–15.
- Zhang, H., Jin, X., Dai, Q., 2018. Synthetic aperture based on plenoptic camera for seeing through occlusions, in: Pacific Rim Conference on Multimedia, Springer. pp. 158–167.