Air2Land: A ground-based vision dataset for UAV landing guidance

XXX[1], XXX[1]

***Abstract*— In this paper we present a new dataset, called Air2Land(A2L), for advancing the state-of-art in object detection and 6D pose estimation in the context of UAV auto landing research. It meets vision and robotics for ground-based vision guidance system having the multi-modal data obtained by different sensors, and pushes forward the development of computer vision and robotic algorithms targeted at visually assisted UAV landing. The dataset, contains sequential stereo images and other sensor data (UAV pose, PTU angles) simulated in various climate conditions and landing scenarios, can be employed for tasks including UAV tracking, keypoint localization, pose estimation and action recognition, etc. In addition to providing plentiful and scene-rich data, our dataset covers high risk scenarios that are hardly accessing in reality. The dataset is avaliable at https://zhengxch.github.io/Air2Land/**

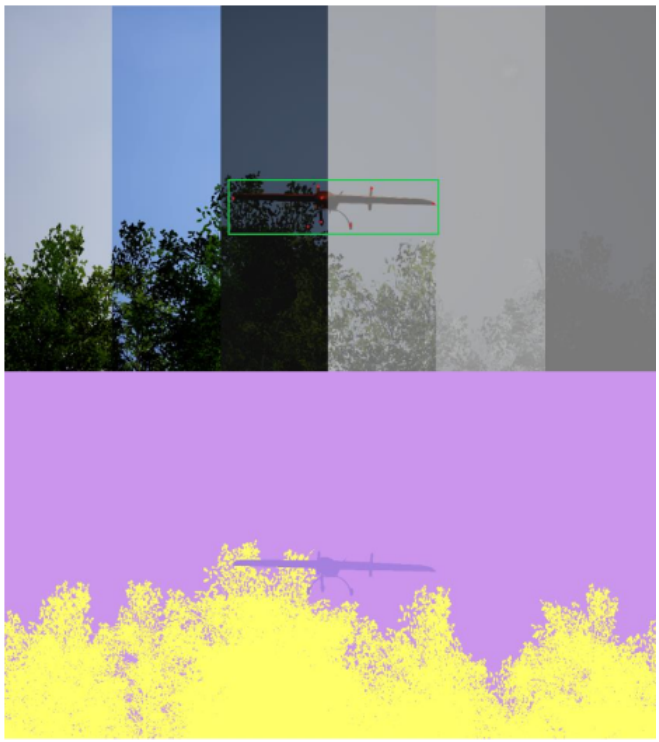

Fig. 1. In the Air2Land dataset, all pictures are labeled with object annotations, keypoint localization, instance segmentation, current location, altitude of UAV, and the rotation data from the PTU on the ground.

*This work was not supported by any organization

[1]School of Aeronautics and Astronautics, Sun Yat-Sen University, GuangZhou, China Email:xxx@mail.sysu.edu.cn

I. INTRODUCTION

In the past decades, unmanned aircraft vehicle (UAV) has been a hotspot in the field of robotic systems. Some of the research results have been emerged in military, industrial and civil applications. Since UAVs usually conduct repeated missions, auto landing and safe recovery are of great significance. To tackle this problem and guarantee the feasibility of UAV auto-landing within GPS-denied environment, several methods based on ground vision measuring system have been proposed [1]-[4]. However, the insufficient quality of data due to difficulties in implementing experiments and measuring error, have not only limited the verification of generation performance of the method but also hindered the introduction of data-driven solutions.

Observing the landing of fixed-wing UAV from the ground turntable camera can capture the landing progress of the UAV. Image processing and computer vision algorithms are then applied to these ground visual data to extract high-level information regarding the environment. Although there exist several famous public datasets (e.g., COCO[5], Pascal VOC[6]), the samples in these datasets consist of natural images that are mainly captured by handheld cameras. In the field of aerial image, existing datasets [7][8] mainly focus on the detection and surveillance of ground targets. Datasets that record the landing process of aerial drones based on ground vision equipment have not yet been proposed.

For the above problems, we build a new dataset called Air2Land which contains of diverse multi-modal data obtained by a hardware on the loop simulation system. Our main motivation is to provide a domain-specific dataset for machine learning tasks such as UAV detection/tracking, keypoint localization, pose estimation and action recognition in vision-assisted UAV landing research. These tasks can be targeted individually or simultaneously in a multi-task set up. With Air2Land, we provide not only sufficient data for verifying a measuring solution, but also a basis for training and benchmarking algorithms.

The Air2Land dataset combines vision and robotics for UAV autonomous landing having the multi-modal data from both airborne and ground sensors. Consisting of more than 76k pictures for UAV landing, the dataset covers changes in flight conditions, weather, seasons and other factors during the drone landing process. The pictures mainly are recorded at two virtual airports restored in full proportion according to the real scene through a simulator supported by the virtual engine UE4.

In this work, we emphasize the differences between ground-to-air perspective images and natural images in the

TABLE I

COMPARISON BETWEEN OUR AIR2LAND DATASET AND SIMILAR DATASETS USABLE FOR OBJECT DETECTION AND POSE ESTIMATION RELATED TO UAV. THE SYMBOL / DENOTES THAT THE INDICATOR IS NOT AVAILABLE FOR THE DATASET.

Dataset	**Air2Land**	AU-AIR	UAV123	Mid-Air	Pascal3D+	YCB-Video	LINEMOD
Number of frames	**76k**	3k	110k	119k	30k	134k	18k
Resolution	**1280×720**	1920×1080	1280×720	1024×1024	500×375	640×480	640×480
Data type	**synthetic**	real	synthetic	synthetic	real	synthetic	real
Environment	**outdoor**	outdoor	outdoor	outdoor	outdoor	indoor	indoor
Perspective	**Ground-to-air**	Air-to-ground	Air-to-ground	Air-to-ground	Various	Close shot	Close shot
Climate variations	**yes**	no	no	yes	no	no	no
Extreme lighting	**yes**	no	no	no	yes	no	no
Stereo	**yes**	no	no	yes	no	no	no
Bounding box coordinates	**yes**	yes	no	/	no	yes	no
Keypoints	**yes**	no	no	/	yes	no	no
6d pose	**yes**	no	no	/	yes	yes	yes
Camera pose	**yes**	yes	no	yes	no	no	no

context of object detection and pose estimation tasks. To this end, we compare images samples and object instance between Air2Land dataset and other dataset for similar object or dataset. To generate data, we constructed a hardware-in-the-loop semi-physical simulation system, which gives the synthetic image a motion characteristic closer to the real flight scene. In our experiment, we train and evaluate one keypoints localization algorithm (stacked hourglass network) on our dataset. We explore the use of deep learning algorithms for keypoints localization in the particular case of weather and light changes.

A. Related Work

It is a challenging task to photograph a UAV in flight from the perspective of a ground-based camera and to solve the UAV's position attitude. Park, Faessler et al. [1] have achieved the solving of the UAV's position attitude through the geometric relationship between diodes and photogrammetry by attaching LED lights to key features on the UAV and photographing them using ground-based sensors. By setting up colored marker points on the body of the aircraft as extracted key feature points, Altug, Oh et al. [2]process the sequential images captured by the camera to calculate the UAV's position attitude. These studies tend to use only a limited number of images, and there is currently no dataset proposed specifically for this task.

Most of the current datasets associated with UAVs are from onboard cameras to observe the external environment or objects. AU-AIR [8][7] enables UAVs to fly at different altitudes, photograph a variety of objects on the ground and provids flight states such as speed, altitude, attitude angle and other flight states, thus forming a multimodal dataset. Mueller et al. [7] present UAV 123, a dataset for low-altitude UAV target tracking, and the authors compare the performance of a variety of current trackers on this dataset. Although the tracked object contains a drone in flight, it only provides information on the image coordinates of the object drone. Fonder et al. [9] propose a synthetic dataset (Mid-Air) for low altitude drone flights in unstructured environments (e.g., forest, country). It includes multi-modal data regarding the flight (e.g., visual, GPS, IMU data) without any annotations for visual data.

For the task of positioning and pose estimation of rigid body targets, most of the available datasets are related to the use of robotic grasping of objects on the platform. For example, YCB-Video [10] and LINEMOD [11] take close shots of objects from the grasping range and give the 6d positional orientation of the target object to inform the robotic grasping. The objects in these datasets are mainly daily life objects that are not moving and the imaging process is static and stable. The size of the objects varies little in scale throughout the image.

Table 1 summarizes some common datasets and compares them to our Air 2Land dataset (see first column), according to the amount of data, the acquisition conditions, and types of data. Table 1 also highlights that the largest and most complete datasets were mainly designed for ground target tracking or static object pose estimation.

B. Contribution

Looking also at the comparison of existing datasets in Table I, the followings are the main contributions of this work:

- To the best of our knowledge, the Air2Land dataset is the first ground-to-air perspective dataset for UAV detection and pose estimation. The dataset includes flight data (i.e., time, GPS, altitude) in addition to visual data and objects and keypoints annotations. We emphasize the differences between object in terrestrial camera and others.
- Considering the practical applicability, we form a baseline training and testing keypoint localization algorithm with the Air2Land dataset. We studied and explored the effects of color and light variations on localization results.

Fig. 2. Elaboration of Air2Land's data generation scheme. Unlike simple picture synthesis, the real-world simulator is driven by flight control data generated by the hardware-in-the-loop system, and kinematic characteristics are passed from the controller to the drone model, then to the scene simulator, and finally naturally retained within the output sequence of images.

II. The terrestrial stereo guidance system for UAV safe landing

A lot of mature research work on guiding the take-off and landing of UAVs through ground-based systems have been implemented [12]. Setting the guidance unit on the ground has the advantages of obtaining rich computing resources and supporting multiple aircraft, which can fully copy with the loss of GPS signal during the landing phase. At present, most relevant researches choose laser or microwave radar as the detection method. Although accuracy will not be lost, it has the disadvantage of insufficient robustness due to active detection and reliance on microwave signals for judgment. To this end, a ground stereo guidance system for UAV safe landing has been constructed and updated for several times in our previous work [13]-[15]. In our guidance system, two PTUs (pan-tilt unit) are symmetrically located on both sides of the run way with each camera fixed on. The computing unit processes the images captured by cameras during UAV landing and estimates the its spatial coordinate and posture. The motion parameters are wirelessly feedback to the on-board autopilot to assist its safe landing.

Although we have built a physical experiment environment and carried out many experiments in the early stage, fixed-wing drones have certain risks in landing phase and it is difficult to design some rare or even dangerous scenes, which are likely to appear and fatal in reality. Therefore, we virtualize airports and drone, and use the hardware-in-the-loop semi-physical simulation environment as a platform for us to carry out repetitive experiments, collect data and verify algorithms.

As shown in Figure 2, the Pixhawk autopilot runs the actual airborne flight control code, which is responsible for the control of the drone during the loading and landing of the mission route, the monitoring and feedback of the drone flight status, the analysis and execution of operating instructions, etc. QGroundControl is a ground station for UAV flight protection, mainly responsible for flight mission planning, digital map display, data link transmission, and comprehensive data analysis. ROS connects various components to realize the transmission of messages on different topics, and the 3D simulator performs simulation and display of the scene.

As the core of scene simulation and data generation, the 3D simulator realizes real-time rendering based on Unreal Engine[16] rendering pipeline, in combination with Airsim, a plugin developed by Shah et al. [17] for this engine. Despite being incompletely physically consistent, the scene simulator has reached a level of visual likelihood and realism which competes with simple ray-tracing algorithms while offering the benefits of reduced render time compared to their physically accurate alternatives.

Different from completely use virtual synthesis software to generate images, our simulation system includes the physical controller and professional dynamics simulation software into the loop, which retains the kinematic characteristics of the simulation data to the greatest extent. This kinematic feature enables us to make full use of the motion correlation of picture frames to accomplish some tasks, such as the position and pose estimation of sequence images.

III. Object detection and Pose estimation in terrestrial image vs others

Vision-based object detection and pose estimation are very common tasks, the availability of large amounts of data and computing power enables deep neural networks to achieve state of art results in part of them. Compared with detection, it's obvious that pose estimation can accurately represent the distance and orientation information of an object. Generally, pose estimation tasks are divided into two categories, one of which is to use airborne or vehicle-mounted cameras to capture a structured environment during movement, and estimate one's own attitude through the feature correlation of adjacent frame images. This task is often called visual odometry in

SLAM (Simultaneous localization and mapping). The other is to use the camera to take a close-up shot of the rigid object in the desktop or cabinet to determine its orientation to facilitate positioning and grasping operations. This task is more common in robotic manipulation, usually called object 6d pose estimation.

Ground-to-air perspective images have different characteristics from images appearing in the above two types of tasks due to having an unique view. First of all, from the perspective of the ground camera, the landing of UAV is a dynamic process from the distant sky to the near ground. During the imaging process, the scale of the drone in the image and the proportion of the target area in the entire image vary greatly. These changes are more significant when compared to images of objects placed on a flat surface taken at close range, which basically maintains a fixed shooting distance and target area ratio. In the ground-based camera field of vision, the UAV has grown from small to large, from overlapping several starting points to covering most of the image, whereas the proportion of objects in an object 6d pose estimation task usually does not vary that much. In particular, when the drone is close to the ground camera, the turntable will also rotate at a large angle to track the drone, resulting in rapid switching of the background. The dramatic diversification in scale and background pose a great challenge to target area positioning and feature-based pose estimation. In addition, in the task of manipulator grasping, the camera often does not move or rotate but is fixed in a suitable imaging area with fixed angle.

Secondly, when comparing with visual odometry, although the two ideas of obtaining the relative pose of the camera and the photographed object by solving PNP problems are similar, these two issues are not simply an inverted relationship between subject and object. The visual odometry focuses on the static part of the image, usually the background, and moving objects will not be used as reference for camera's self-positioning, while the terrestrial camera follows the foreground part closely, and the background information is often ignored. This difference means that when we try to use some end-to-end approach for pose calculation, we cannot use the information of the entire picture as the input of neural network like visual odometry.

Finally, since the inspected UAV has no texture, only characteristic edge and anchor point information can be selected as image features for pose estimation. These features are easily lost or obscured when the UAV attitude changes significantly. Besides, the left-right symmetry of the drone is likely to cause ambiguity in the semantics of key features or confusion in target orientation.

IV. Air2Land – the multi-modal UAV dataset

After the details related to the simulation environment and challenges addressed by the task, we now propose a synthetic dataset for UAV landing guidance. Each sample contains information including the image captured by the terrestrial camera, object and keypoint annotations, and the corresponding UAV attitude as well as PTU rotation angles.

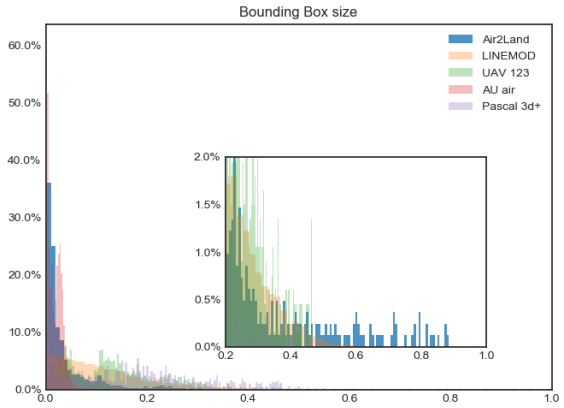

Fig. 3. Comparison of bounding box sizes for observations from different datasets. The drone's landing process from far to near makes its bounding box cover a wide range of proportions of the full image, with some samples spread over intervals with proportions greater than or equal to 0.5.

All images are recorded at a resolution of 1280×720 at a frame rate of 25 frame per second (fps), while the sensor data corresponding to each frame is saved to a specified tag file.

A. UAV Platform

We adopted "Pioneer", produced by VIGA Tech company, as our platform, since it has sufficient wingspan to be detected at a distance from the lens, reserving enough time and spatial intervals for us to record and analyze the data. In addition, its feature points are more dispersed and distinguishable. We provide the CAD model of the UAV and the coordinates of several anchor points that we consider to be important, which are indispensable in the methods of pose estimation based on contour matching or 2D-3D correspondence point matching.

B. Dataset Collection

We conducted a variety of flight tests and collected data at two virtual airports that were 1:1 replicas of the real airports in Changsha, Hunan, and Xuchang, Henan, where we conducted our out-of-field flight tests. By setting up flight missions in the QGC control station and introducing dynamic interference to the aircraft in Xplane, we conducted a large number of virtual flight tests based on the semi-physical simulation system mentioned above, tried many experiments that were difficult to carry out in reality.

We carefully designed the route and set up appropriate dynamic interference. Flight tests such as normal flight, large-angle rotation under crosswind interference, go-around, and instability crash were carried out at two airports, with all critical data recorded. To enrich visual changes, we traversed the time and weather conditions for each flight trial through interface provided by the Airsim plugin, covering the early morning, midday, and evening hours and weather conditions such as rain, snow, haze, and wind. Using a pre-specified communication protocol, we repeated the trials offline by

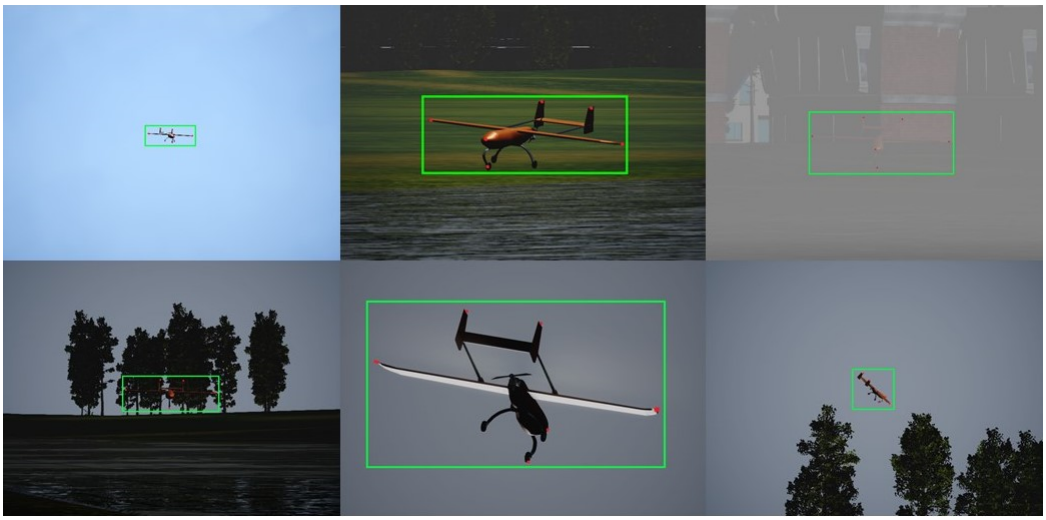

Fig. 4. The UAV present a very diverse range of perspectives and scale variations in the field of view of ground-based cameras, and the labeling results for some typical cases are shown in the figure.

varying only the weather and time factors to ensure that the conditions were identical except for the visual imaging factors, which laid the foundation for our later data analysis.

Each flight test corresponds to 48 light and weather changes, taking into account all possible external conditions and posing a considerable challenge to the detection algorithm.

C. Annotation

Generating data synthetically helps to record ground-truth data which is more reliable or even impossible to capture with real sensors, thereby ensuring the accuracy of annotation.

The 3D Reality Simulator follows a strict coordinate transformation and projection transformation relationship when performing scene simulation, and this transformation relationship can be used as the basis for our batch automatic annotation of data. We make full use of the principles of spatial coordinate transformation and OpenGL perspective projection to write a script program for automatic projection annotation of the simulated images, which eliminates the possible errors caused by manual annotation. The annotation results for some typical images are shown in Figure 4.

V. Evaluation and Analysis

The construction of a theoretical model for target pose estimation based on the geometric constraints of the target, with clear physical connotations and strong interpretability, is a common solution to the problem of target spatial position pose estimation in computer vision research. The most widely used pose solution method is to apply the pinhole camera model to solve the correspondence between 2D image points and 3D spatial points as a PnP problem. Fully considering the influence of the strength of the feature, visibility, envelope and number of anchor points, we selected the left wingtip, left tail tip, front gear, nose tip, right wingtip, and the right tail tip of the drone as feature anchor points.

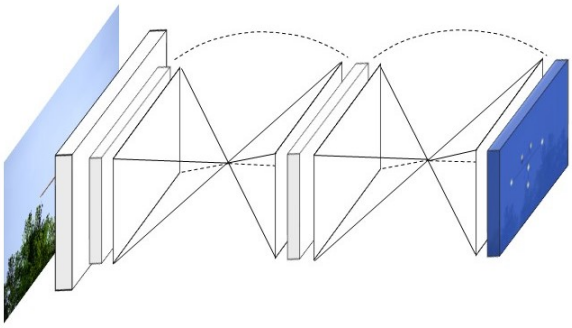

Fig. 5. The keypoint localization neural network model consists of two stacked hourglass network modules.

To analyze the robustness of methods to visual changes, we use images generated under sunny, cloudy, light rain, moderate rain, light snow, moderate snow as the training set, and the testing set is composed of images under heavy snow, heavy rain, and foggy weather, both of which cover the early morning, midday, and evening time periods. The number ratio of the training set and testing set is about 7:3. During the evaluation, we consider real-time performance rather than achieving a state-of-art accuracy for the sake of the applicability. We choose Stacked hourglass network[18], which is widely used in human pose estimation, as our baseline method. In order to reduce the parameters to obtain a lighter model, we set the number of hourglass modules of the network to 2, with each hourglass module has one residual block.

To compare localization performances, we use percentage of correct key-points (PCK) from [19] that is a prominent metric in human pose estimation. It measures accuracy of the localization of body joints. Detected joint is considered correct if the distance between the predicted and the true joint is within a certain threshold (threshold varies).Drawing on this metric, we choose the bounding box of the drone's nose

TABLE II

COMPARISON BETWEEN OUR AIR2LAND DATASET AND SIMILAR DATASETS USABLE FOR OBJECT DETECTION AND POSE ESTIMATION RELATED TO UAV. THE SYMBOL / DENOTES THAT THE INDICATOR IS NOT AVAILABLE FOR THE DATASET.

weather	time	LW	LT	FG	N	RW	RT	mean	fps
foggy	8:00	62.02	70.91	68.60	72.76	67.32	79.32	70.16	29.98
	12:00	63.63	72.76	77.57	73.87	75.33	73.97	72.86	31.61
	17:00	60.29	77.13	72.41	73.05	68.20	75.38	71.08	30.07
Heavy rain	8:00	63.62	79.42	77.41	75.99	73.08	78.95	74.75	30.04
	12:00	68.90	78.82	79.92	77.98	78.92	78.99	77.26	31.40
	17:00	60. 63	72.77	74.39	79.24	69.36	76.39	74.43	27.24
Heavy snow	8:00	61.34	68.09	69.57	70.87	60.86	70.54	**66.88**	21.67
	12:00	62.62	70.01	72.04	74.90	60.64	72.28	**68.75**	24.94
	17:00	64.29	63.06	66.30	70.38	65.32	69.77	**66.52**	23.82
mean		**62.75**	72.55	73.13	74.34	**68.78**	75.07	71.41	27.86

as a benchmark, and set the correct threshold of prediction to 50% of the length of the rectangular box, which we also call PCKh@0.5.

We use the epoch of 50, batch of 16 and RMS optimizer with learning rate = 2.5e-4. Images are resized to 512 to fit the network input. The training is stopped when the performance of the model on the validation dataset starts to degrade. After completing the training, we tested the model using the image set corresponding aforementioned extreme weather condition, and the results are given in the Table II.

We observe that the PCK values of left wing tip and right wing tip are significantly lower than that of other anchor points, which may happen due to the wingtip anchor points varies considerably in appearance characteristics from different viewpoint and is located at the edge of the region of interest. When considering the generalization ability of the model under different weather conditions, it can be seen that the model generalizes best in heavy rain with a PCK value of about 75, while the model generalizes worst in heavy snow, and the test results in foggy weather are in the middle of the two. As shown in the error detection example in Figure 4, the visibility of the images generated in heavy snow and hazy weather is reduced very significantly, and locating the coordinates of the anchor points from these images is very challenging. For the time metric, the model's average inference speed over the three test subsets is about 27FPS on 1080Ti, and if consider the spatial pose estimation link that has not yet been added, it is clear that increased speed is necessary for practical applications.

VI. OVERALL LIMITATIONS

Despite all our efforts, there is still much room left for improving in Air2Land.

The first one is that our dataset only contains two simulated airports and single observing object. Although we have tired our best to enrich the samples by simulating weather and light variations, we have not addressed this problem at all. In addition, due to the unique nature of the terrestrial vision-guided UAV landing mission, the UAV's angle variation is always within a limited range, and our observables tend to present similar perspectives in camera, with a unipolar sample distribution. This means that the algorithm's ability to generalize to other perspectives is untested.

The second limitation is that the simulator we use does not include situation in which the drone appears truncated in the image or breaks out of the camera's field of view. Both can have implications for algorithms based on sequential images for posture tracking. Our dataset should nonetheless be useful to get a reliable performance score for simple scenarios and therefore to get a first overview on the potential of any tested method.

VII. CONCLUSION

We introduce a new synthetic dataset, named Air2Land, featuring 76k frames of UAV landing process recorded by terrestrial camera with varied illumination and climate conditions. The content of our dataset contains multiple synchronization modalities that provide data for posture estimation tasks (e.g., UAV pose solving) as well as purely computer vision tasks (e.g., target detection or semantic point localization). Our original intent was to provide a dataset that could be used for algorithm testing by our research colleagues in the field of UAV landing state monitoring based on external observational methods. While partially achieves this effect, the size and content of this dataset make it also useful for training and testing generic machine learning algorithms.

We explain in detail the process of generating the simulation data, thereby highlighting the kinematic laws inherent in the images. In addition, we analyze ground-based and other view camera imaging results and highlight their implications for the positional estimation task. Moreover, since we consider real-time performance and applicability in real-world scenarios, we test the generalization capability of the keypoint location algorithm on our dataset.

REFERENCES

[1] Park, S., Won, D. H., Kang, M. S., Kim, T. J., Lee, H. G., and Kwon, S. J. (2005, August). Ric (robust internal-loop compensator) based flight control of a quad-rotor type uav. In 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3542-3547). IEEE.

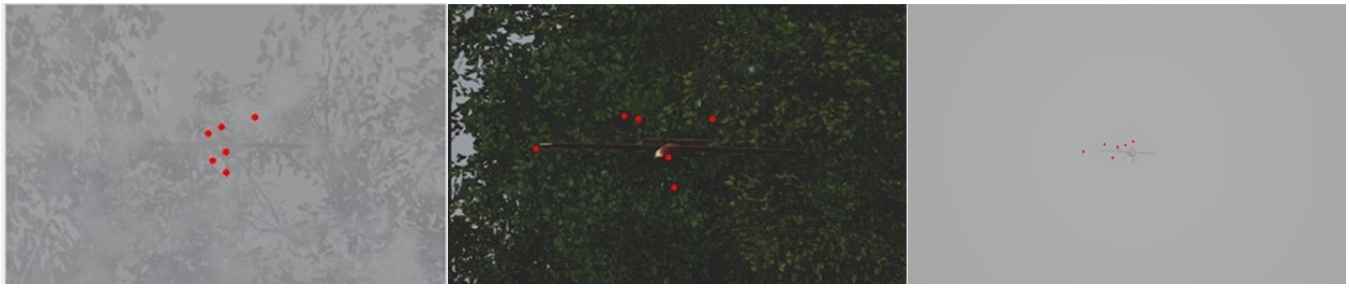

Fig. 6. In low visibility weather conditions such as heavy snow and haze, the keypoint location algorithm is prone to false detection, background interference and longer distances also have a negative impact.

[2] Faessler, M., Mueggler, E., Schwabe, K., and Scaramuzza, D. (2014, May). A monocular pose estimation system based on infrared leds. In 2014 IEEE international conference on robotics and automation (ICRA) (pp. 907-913). IEEE.

[3] Altuğ, E., Ostrowski, J. P., and Taylor, C. J. (2005). Control of a quadrotor helicopter using dual camera visual feedback. The International Journal of Robotics Research, 24(5), 329-341.

[4] Oh, H., Won, D. Y., Huh, S. S., Shim, D. H., Tahk, M. J., and Tsourdos, A. (2011). Indoor UAV control using multi-camera visual feedback. Journal of Intelligent and Robotic Systems, 61(1-4), 57-84.

[5] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... and Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

[6] Everingham, M. , Gool, L. V. , Williams, C. K. I. , and Zisserman, W. A. . (2010). The pascal visual object classes (voc) challenge. International Journal of Computer Vision.

[7] Matthias Mueller, Neil Smith, and Bernard Ghanem. (2016). A benchmark and simulator for uav tracking. Far East Journal of Mathematical ences, 2(2), 445-461.

[8] Bozcan, I. , and Kayacan, E. . (2020). Au-air: a multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance.

[9] Fonder, M. , and Droogenbroeck, M. V. . (2019). Mid-Air: A Multi-Modal Dataset for Extremely Low Altitude Drone Flights. IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE.

[10] Xiang, Y., Schmidt, T., Narayanan, V., and Fox, D. (2017). Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. arXiv preprint arXiv:1711.00199.

[11] Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., and Navab, N. (2012, November). Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In Asian conference on computer vision (pp. 548-562). Springer, Berlin, Heidelberg.

[12] UCARS-V2: UAV Common Automatic Recovery System-Version 2 ForShipboard Operations.

[13] Tang, D., Hu, T., Shen, L., Zhang, D., and Zhou, D. (2015, April). Chan-Vese model based binocular visual object extraction for UAV autonomous take-off and landing. In 2015 5th International Conference on Information Science and Technology (ICIST) (pp. 67-73). IEEE.

[14] Kong, W., Zhang, D., Wang, X., Xian, Z., and Zhang, J. (2013, November). Autonomous landing of an UAV with a ground-based actuated infrared stereo vision system. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2963-2970). IEEE.
[15] Zhang, Y., Shen, L., Cong, Y., Zhou, D., and Zhang, D. (2013, December). Ground-based visual guidance in autonomous UAV landing. In Sixth International Conference on Machine Vision (ICMV 2013) (Vol. 9067, p. 90671W). International Society for Optics and Photonics.
[16] Epic Games. Unreal Engine web site. https://www.unrealengine.com.
[17] Shah, S. , Dey, D. , Lovett, C. , and Kapoor, A. . (2017). Airsim: high-fidelity visual and physical simulation for autonomous vehicles.
[18] Yang, J., Liu, Q., and Zhang, K. (2017). Stacked hourglass network for robust facial landmark localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 79-87).
[19] Yang, Y., and Ramanan, D. (2012). Articulated human detection with flexible mixtures of parts. IEEE transactions on pattern analysis and machine intelligence, 35(12), 2878-2890.