Computer Vision System for Landing Platform State Assessment Onboard of Unmanned Aerial Vehicle in Case of Input Visual Information Distortion

Iuliia Kim, Ilya Viksnin
ITMO University
Saint Petersburg, Russia
yulia1344@gmail.com, wixnin@mail.ru

Irina Kaisina
Kalashnikov Izhevsk State Technical University
Izhevsk, Russia
irinakayisina25@gmail.com

Vadim Kuznetsov
Unmanned Systems Group
Izhevsk, Russia
kvs@unmanned.ru

***Abstract*—The paper describes a computer vision system for organizing a safe landing of an unmanned aerial vehicle in conditions of potential distortions of the input video information. A sequence of methods for image preprocessing was proposed. Firstly, it is necessary to conduct a contrast enhancement using histogram equalization. After that, a Gaussian filter should be applied to remove an extra noise. Neural network YOLOv3-tiny was trained to recognize the state of the landing platform - open or closed. The achieved recognition accuracy on the test sample was 0.96. The algorithm was implemented in Jetson Nano and the achieved frame processing time is equal to 0.5 seconds.**

I. Introduction

At present, modern society tends to automate processes that can be associated for humans with routine, large amounts of data, and danger. This led to the development of Industry 4.0 characterized by the integration of the cyber-physical concept into different spheres of life. The main attributes of Industry 4.0 are represented below:

- autonomous robots;
- Internet of Things;
- blockchain technologies;
- bid data processing;
- quantum computing, etc.

In this paper, the focus is put on the development in the sphere of autonomous robots, particularly, unmanned aerial vehicles (UAVs). To provide a correct transport functionality, it is necessary to process the input information and make decisions for further activity based on the obtained results. The computer vision system is one of the essential UAV parts because it is responsible for processing video streams. The incorrect processing results will cause fatal consequences, as the UAV can misidentify obstacles, road signs, pedestrians, etc.

This paper describes a computer vision system elaborated for potential visual information distortions in case of adverse conditions (rain, snow, nighttime). The goal is to elaborate a system that will be able to assess for landing platform state classification and assessment of landing possibility.

A significant part of the autonomous transport work consists of visual information processing. Based on it, a vehicle can assess the environment, detect obstacles, monitor the territory for anomalies (fire, flood, etc.). Thus, it is vital to elaborate computer vision systems that provide a low error rate and minimize the risk of wrong information interpretation.

In this research the following materials and methods were used:

- the data collection (images) was conducted on the UAV board during the real flight;
- the annotations were made manually on the collected images;
- for the computer vision system assessment the quantitative metrics were used: accuracy, precision, recall, F1-score.

The structure of the paper is as follows: Section 1 describes the general aspects of the investigation, Section 2 contains the research task statement, Section 3 presents state of the art in the sphere of computer vision systems for UAVs, Section 4 describes data and elaborated computer vision system, Section 5 provides information about the experiment conducted for the assessment of the developed system, Section 6 summarizes the obtained results and states future research steps.

II. Research task statement

The research was conducted based on the data obtained from the Unmanned Systems Group. It provides a full production cycle of UAV complexes, starting from the composite, ending with the manufacture of high-capacity storage batteries, allowing the UAV to stay in the air for up to 6 hours. Also, it organizes UAV flights for different tasks, i.e. monitoring.

To provide an opportunity to use UAV in different locations without binding to the concrete station, an automated takeoff and landing platform is used - SuperBOX [1]. Its characteristics are enumerated in Table I.

The platform provides an opportunity to form a flight task and load the formed route into the control system. At the appointed time, the platform will deploy and prepare the UAV for departure with notification of the successful aviation work realization. At the end of the flight, the UAV is charged through the built-in autonomous battery charging system. The UAV model taken for this research is DJI Mavic 2Zoom with

TABLE I. SUPERBOX CHARACTERISTICS

Characteristics	Values
Dimensions (length * width * height)	3000 * 1600 * 1500 mm
Weight	300 kg
Power consumption	up to 1.5 kW
Dust and moisture protection class	IP43
Data transmission interfaces	Ethernet, KTP, DataLink
Precise point positioning	RTK, IR beacon, optical marker
Autonomous operation with no external power suppl	48 hours
UAV battery full charging time	3 hours

the integrated 4K camera. The video stream FPS is equal to 30.

However, the UAV can be used in unfavorable conditions because of bad weather, nighttime, extremely cold/hot seasons, etc. In these cases, there is a risk that the platform can be damaged and the UAV will not be able to land safely. That is why it is proposed to integrate a computer vision system into the UAV so that it could assess the state of the landing platform and based on the analysis results, could make decisions about landing actions. The landing platform is represented in Fig. 1

Fig. 1. Landing platform

The research task is the elaboration of a computer vision system for UAV that can function in case of potential input visual information distortions and provide a recognition accuracy not less than 0.85.

III. STATE OF THE ART

There were many investigations dedicated to vision-based landing. They provide effective algorithms for landing marker detection and tracking.

One of the ways for detecting the object and its contour is binarization. There are different binarization variations, and they are described in [2].

The paper [3] describes UAV landing using computer vision on a moving vehicle. To build a computer vision system onboard the UAV, a Cameleon 3 USB camera and a fisheye lens are used to create a wide panoramic or hemispherical image. The UAV has a Jetson TK1 to process visual information. To organize the landing on a moving vehicle, a color marker is traced by computer vision methods.

The marker is detected in the following sequence:

1) noise removal through a Gaussian filter;
2) image conversion from RGB to HSV;
3) binarization;
4) noise removal using a morphological filter;
5) determination of the position of the detected object.

In the work [4], the authors also propose a system for landing UAVs. The algorithm is similar to the previous one:

1) image conversion from RGB to HSV;
2) binarization;
3) noise removal;
4) determination of the center of the image;
5) determination of the center of the detected object.

UAV landing process was investigated in [5]. The authors of this paper developed a marker for a landing platform based on the AprilTag system. The landing process is organized by combining the results of marker detection (through pixel-by-pixel readout and analysis) and the results of the UAV position relative to the marker (using the RPP method - Robust Pose from Planar target). Extended Kalman Filter helps to combine data from various sensors. Thus, the landing platform is positioned and tracked.

The landing platform marker detection by pattern matching is presented in [6]. A dataset of marker key points is generated using the Speeded up Robust Features (SURF) method. The output video stream frame is matched against examples from the generated dataset. The search for the closest similarity to descriptors from a dataset is performed using FLANN (Fast Library for Approximate Nearest Neighbors).

In [7], the following algorithm for landing platform recognition is proposed:

1) image conversion from RGB to HSV;
2) binarization;
3) detection of geometric shapes that make up the landing platform;
4) determination of the probability of reliable recognition of the detected platform based on comparative partial analysis;
5) the platform is considered classified if the probability of reliable recognition is greater than or equal to 0.6.

UAVs often operate in open spaces. Since the camera is not protected from external influences, there is a risk of distortion of the input visual information due to natural reasons: rain, fog, nighttime, snow. That is why the researchers work on image preprocessing to minimize the distortion negative effect to recognition accuracy.

In [8], a method was proposed for preprocessing images onboard a drone to eliminate blur in the frame. When the frame is taken, the measure of the image blur is calculated and corrected accordingly.

In paper [9], to improve the quality of the analyzed images for monitoring agriculture using UAVs, frames converted to grayscale were subjected to the process of leveling the intensity histogram.

In [10], to reduce image noise, a median filter is used, due to which pattern recognition is faster and with fewer errors.

In this paper, it is planned to elaborate a sequence of methods to minimize image distortions and provide a high recognition accuracy.

IV. Computer vision system for landing platform state assessment

A. Data description

In this research, it is planned to elaborate a computer vision system that will determine the landing platform state: closed or open. The platform is considered open if 4 markers can be visible. A dataset of 25229 images was collected (2688x1512 pixels): 9516 images of the closed platform and 15713 images of the open one. In the Fig. 2 and Fig. 3 the closed and open platforms are represented, respectively.

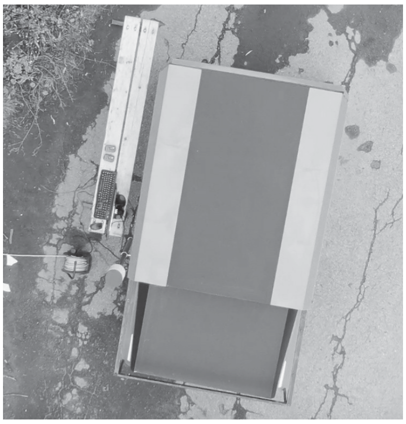

Fig. 2. Closed landing platform

Fig. 3. Open landing platform

The pictures were taken in the following conditions: sunny weather, snow, nighttime. Based on the collected images, it is planned to augment the existing dataset and train a neural network.

TABLE II. Comparison of contrast enhancement methods

Method	Work time (sec)	Average pixel saturation
Gamma-correction	8.79	52.48
Histogram equalization	0.02	61.46

B. Computer vision system

1) Image preprocessing: The common image distortions are enumerated below:

- uneven light distribution;
- noise;
- blur.

The landing platform has colored markers that stand out from the image background. That is why it is necessary to highlight them. It is possible via contrast enhancement. To realize it, two methods were tested:

- histogram equalization [11];
- gamma correction [12].

For the method assessment, two metrics were used:

- time;
- average pixel saturation (this metric was taken from HSV image format).

The comparison results are represented in Table II.

The method of histogram equalization showed an acceptable speed for processing the video stream in real-time, and also significantly increased the contrast, which made it possible to distinguish the color markers of the landing platform from the general background. That is why it was chosen for the given research task.

The saturation channel histograms before and after the contrast enhancement are represented in Fig. 4 and Fig. 5, respectively.

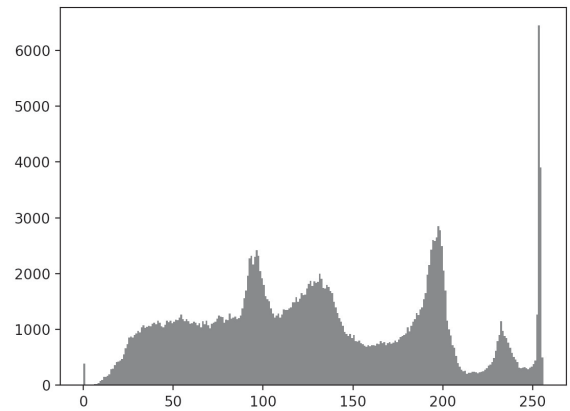

Fig. 4. Saturation histogram before equalization

The method histogram equalization has a negative consequence - it can cause additional noise because the details become more highlighted. That is why after contrast enhancement, it is necessary to remove extra noise. For this purpose,

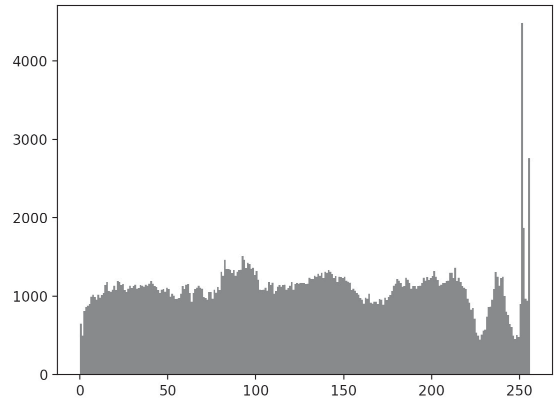

Fig. 5. Saturation histogram after equalization

TABLE III. COMPARISON OF DENOISING FILTERS

Method	Work time (sec)	PSNR	SSIM
Median filter	0.0014	27.98	0.962
Bilateral filter	0.3721	30.11	0.981
Gaussian blur	0.0007	27.2	0.96

spatial filters can be used. In this research, three filters were analyzed:

- median filter [13];
- bilateral filter [14];
- Gaussian blur [15].

The median filter goes through the image as a window (kernel). Intensity values within the filter window are sorted in ascending or descending order, and the value in the middle of the ordered list goes to the filter output.

The bilateral filter replaces the intensity of each pixel with a weighted average of the intensities from adjacent pixels. It works as in (1).

$$I^{filtered} = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(|I(x_i) - I(x)|) g_s(x_i - x) \quad (1)$$

where $I^{filtered}$ is filtered image, I - raw image, x - pixel coordinates, W - kernel with center in x, f_r - range kernel for smoothing intensity differences, g_s - range kernel for smoothing coordinate differences.

The Gaussian blur also uses a kernel and removes the noise according to (2).

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (2)$$

The filter comparison was conducted with three metrics:

- work time;
- peak signal-to-noise ratio (PSNR) [16];
- structural similarity index measure (SSIM) [17].

The parameter comparison for the denoising filters is represented in Table III.

TABLE IV. TRAINING PARAMETERS

	YOLOv3-tiny	SSD Mobilenet	Faster R-CNN Resnet
Momentum	0.9	0.89	0.9
Decay	0.0005	0.003	0.004
Learning rate	0.001	0.0004	0.0004
Batch size	32	6	2

TABLE V. JETSON NANO TECHNICAL CHARACTERISTICS

Characteristics	Values
GPU	128-core Maxwell
CPU	Quad-core ARM A57 @ 1.43 GHz
Memory	0

It can be noted that the Gaussian filter has the least work time, while the values of the other metrics are approximately the same. The preference was given to the Gaussian filter.

2) Neural network: After performing the preprocessing operations, the step of recognition should be fulfilled. For this step, it was decided to train a neural network due to its flexibility and opportunity to find object patterns with their geometrical schemes and characteristic details.

For the computer vision tasks, convolutional neural network are used. The image dataset was divided into two sets: 80% train and 20% test. Then the augmentation process was conducted with the following parameters:

- rotation - 359°;
- saturation fluctuations - 255;
- hue fluctuations - 179.

For the training process, three neural network architectures were tested:

- YOLOv3-tiny;
- SSD Mobilenet;
- Faster R-CNN Resnet.

Training parameters are represented in Table IV.

After the training process, it is necessary to perform the model evaluation.

V. COMPUTER VISION SYSTEM EVALUATION

A. Experiment setup

It is planned to perform the recognition process on the UAV board. Thus, a computer with high performance and low weight should be integrated. Currently, it was decided to use Jetson Nano, so the neural network evaluation was performed on this device.

Technical characteristics are enumerated in Table V.

For the model assessment, four basic parameters were used:

- True Positive (TP) - cases when the closed platform is classified as closed;
- False Positive (FP) - cases when the open platform is classified as closed;
- True Negative (TN) - cases when the open platform is classified as open;

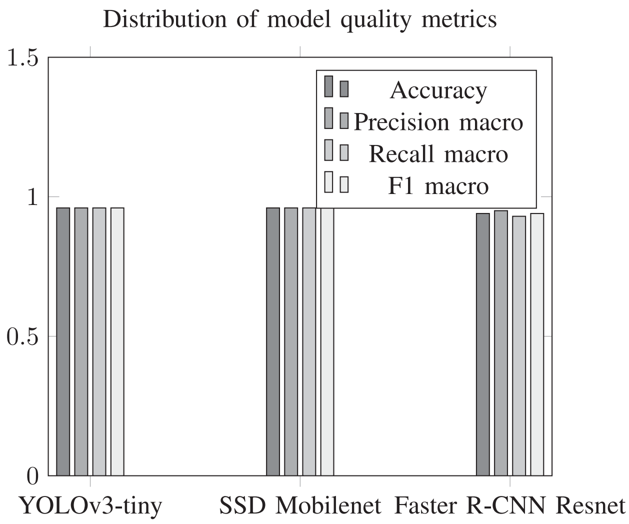

Fig. 6. Distribution of metrics by neural network architectures

- False Negative (FN) - cases when the closed platform is classified as open.

Based on these parameters, the following metrics are calculated:

- Accuracy according to (3);

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

- Precision according to (4);

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- Recall, as in (5);

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- F1-score, as in (6).

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{6}$$

B. Result assessment

The result distribution of accuracy, precision, recall, and f1-score is represented in Fig 6.

It is possible to see that all the models showed similar results. The accuracy values are acceptable for the research goal. In order to take into account more detailed information for model assessment, confusion matrix were built for test sets. The confusion matrix for Yolov3-tiny, Mobilenet, and Resnet are in Fig. 7, Fig. 8 and Fig. 9, respectively.

For the test set, SSD Mobilenet showed a higher tendency to False Negative values (the closed platform was classified as open), which is critical because there is a risk for UAV to land improperly. Faster R-CNN Resnet apart from YOLOv3-tiny provides a higher False Positive rate (the open platform was classified as closed). This error can cause the waste of resources while the UAV will assess the platform with other

Predicted labels

Actual labels		closed	open
	closed	3021	64
	open	129	1753

Fig. 7. Confusion matrix for YOLOv3-tiny

Predicted labels

Actual labels		closed	open
	closed	3000	85
	open	110	1772

Fig. 8. Confusion matrix for SSD Mobilenet

Predicted labels

Actual labels		closed	open
	closed	3024	61
	open	224	1658

Fig. 9. Confusion matrix for Faster R-CNN Resnet

TABLE VI. YOLOv3-TINY RECOGNITION QUALITY FOR DIFFERENT SHOOTING CONDITIONS

Metric	Sunny weather	Snow	Snow + night time
Accuracy	0.94	0.98	0.98
Precision	0.94	0.98	0.98
Recall	0.95	0.98	0.98
F1-score	0.94	0.98	0.98

sensors or find a reserved platform. Thus, it was decided to use YOLOv3-tiny for the recognition stage.

The YOLOv3-tiny recognition quality assessment was conducted with the images made in different weather and time conditions. The test set was divided into three subsets: images of the platform in sunny weather, the platform with snow background, the platform with snow background at night. Table VI illustrates the results for these subsets.

The lowest values were shown in a subset of images with sunny weather. The reason is that the excessive illumination makes the platform blend into the background.

In case of video stream unavailability or extremely bad visibility conditions, the UAV can switch over to spatial sensors (IMU) and remote manual control.

In Jetson Nano, one frame is processed in 0.5 seconds. Thus, 2 FPS can be provided which is considered acceptable for the current task. The recognition results are represented in Fig. 10 and Fig. 11 for closed and open platform, respectively. Then, a recognition module application structure was formed and implemented in Jetson Nano. It is showed in Fig. 12.

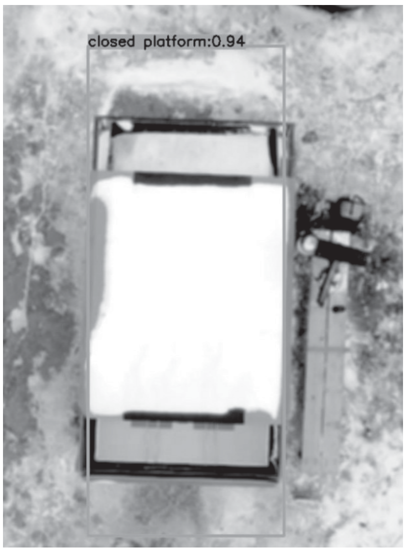

Fig. 10. Classified closed landing platform

Further, it is planned to switch from Jetson Nano to Jetson Xavier NX, integrate it into UAV, and test the elaborated computer vision system in a natural flight.

Fig. 11. Classified open landing platform

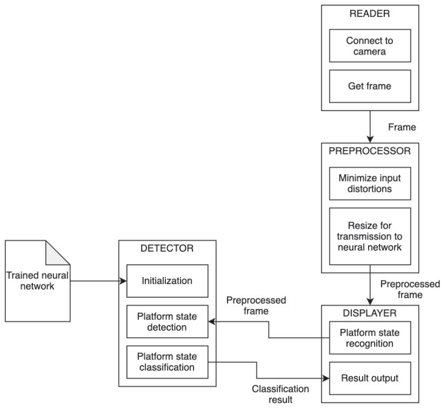

Fig. 12. Computer vision software module structure

VI. CONCLUSION

In this paper, the aspects of computer vision usage in UAV were investigated. The potential input image distortions were taken into account, and the authors elaborated the computer vision system which includes several stages from preprocessing to classification result. The developed system was tested on Jetson Nano and the recognition accuracy achieved the following results:

- accuracy - 0.96;
- precision - 0.96;
- recall - 0.96;
- F1 - 0.96;
- FPS - 2.

The current version of the computer vision system is intended to assess the platform state without UAV positioning relative to the platform. Prospectively, Jetson Xavier NX will replace Jetson Nano, it will be integrated into UAV and the elaborated system will be tested in the conditions of the real flight. For this research step, DJI Mavic was used, and then, the algorithm will be implemented in the custom UAV - Supercam

X6M2 [18]. After the native experiments for the elaborated computer vision, the vision-based UAV positioning system relative to the platform is planned to be developed.

REFERENCES

[1] "Superbox atlp," (Date last accessed 01-March-2021). [Online]. Available: https://en.supercam.aero/catalog/superbox-atlp

[2] B. Shumilov, Y. Gerasimova, and A. Makarov, "On binarization of images at the pavement defects recognition," in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. IEEE, 2018, pp. 107–110.

[3] H. Lee, S. Jung, and D. H. Shim, "Vision-based uav landing on the moving vehicle," in *2016 International conference on unmanned aircraft systems (ICUAS)*. IEEE, 2016, pp. 1–7.

[4] M. Rabah, A. Rohan, M. Talha, K.-H. Nam, and S. H. Kim, "Autonomous vision-based target detection and safe landing for uav," *International Journal of Control, Automation and Systems*, vol. 16, no. 6, pp. 3013–3025, 2018.

[5] O. Araar, N. Aouf, and I. Vitanov, "Vision based autonomous landing of multirotor uav on moving platform," *Journal of Intelligent & Robotic Systems*, vol. 85, no. 2, pp. 369–384, 2017.

[6] V. Sudevan, A. Shukla, and H. Karki, "Vision based autonomous landing of an unmanned aerial vehicle on a stationary target," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2017, pp. 362–367.

[7] J. Garcia-Pulido, G. Pajares, S. Dormido, and J. M. de la Cruz, "Recognition of a landing platform for unmanned aerial vehicles by using computer vision-based techniques," *Expert Systems with Applications*, vol. 76, pp. 152–165, 2017.

[8] S. Bang, H. Kim, and H. Kim, "Uav-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching," *Automation in construction*, vol. 84, pp. 70–80, 2017.

[9] Y. Li, M. Qian, P. Liu, Q. Cai, X. Li, J. Guo, H. Yan, F. Yu, K. Yuan, J. Yu *et al.*, "The recognition of rice images by uav based on capsule network," *Cluster Computing*, vol. 22, no. 4, pp. 9515–9524, 2019.

[10] J. L. Sanchez-Lopez, J. Pestana, S. Saripalli, and P. Campoy, "An approach toward visual autonomous ship board landing of a vtol uav," *Journal of Intelligent & Robotic Systems*, vol. 74, no. 1, pp. 113–127, 2014.

[11] M. Kaur, J. Kaur, and J. Kaur, "Survey of contrast enhancement techniques based on histogram equalization," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 7, 2011.

[12] S. A. Amiri and H. Hassanpour, "A preprocessing approach for image analysis using gamma correction," *International Journal of Computer Applications*, vol. 38, no. 12, pp. 38–46, 2012.

[13] A. Boyat and B. K. Joshi, "Image denoising using wavelet transform and median filtering," in *2013 Nirma University International Conference on Engineering (NUiCONE)*. IEEE, 2013, pp. 1–6.

[14] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, *Bilateral filtering: Theory and applications*. Now Publishers Inc, 2009.

[15] B. S. Kumar, "Image denoising based on gaussian/bilateral filter and its method noise thresholding," *Signal, Image and Video Processing*, vol. 7, no. 6, pp. 1159–1172, 2013.

[16] T. Veldhuizen, "Measures of image quality," *CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision*, 2010.

[17] P. Ndajah, H. Kikuchi, M. Yukawa, H. Watanabe, and S. Muramatsu, "Ssim image quality metric for denoised images," in *Proc. 3rd WSEAS Int. Conf. on Visualization, Imaging and Simulation*, 2010, pp. 53–58.

[18] "Supercam x6m2," (Date last accessed 10-March-2021). [Online]. Available: https://supercam.aero/catalog/supercam-x6m2