TRAINING VIOLA–JONES DETECTORS FOR 3D OBJECTS BASED ON FULLY SYNTHETIC DATA FOR USE IN RESCUE MISSIONS WITH UAV

S.A. Usilin[1−3], *V.V. Arlazarov*[1−4], *N.S. Rokhlin*[5], *S.A. Rudyka*[5],
S.A. Matveev[5], *A.A. Zatsarinnyy*[2]

[1]Smart Engines Service LLC, Moscow, Russian Federation
[2]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russian Federation
[3]Moscow Institute of Physics and Technology, Moscow, Russian Federation
[4]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, Moscow, Russian Federation
[5]Baltic State Technical University "VOENMEH" named after D.F. Ustinov, St. Petersburg, Russian Federation
E-mails: usilin@smartengines.com, vva@smartengines.com, 5hark1@rambler.ru, rudika_sa@voenmeh.ru, matveev_sa@voenmeh.ru, azatsarinny@ipiran.ru

In this paper, the problem of training the Viola–Jones detector for 3D objects is considered on the example of an inflatable life raft PSN-10. The detector is trained on a fully synthetic training dataset. The paper discusses in detail the methods of modelling an inflatable life raft, water surface, various weather conditions. As a feature space, we use edge Haar-like features, which allow training the detector that is resistant to various lighting conditions. To increase the computational efficiency, the L1 norm is used to calculate the magnitude of the image gradient. The performance of the trained detector is estimated on real data obtained during the rescue operation of the trawler "Dalniy Vostok". The proposed method for training the Viola–Jones detectors can be successfully used as a component of hardware and software "assistants" of the UAV.

Keywords: machine learning; object detection; Viola–Jones; classification; 3D object; UAV; rescue mission.

Introduction

At the present time various types of unmanned aerial vehicles (UAVs) have been used for quite a long time [1]. Such complexes are used in the army, various law enforcement agencies, as well as emergency services. UAVs are widely used in search and rescue operations. Surveying the area using a UAV can significantly speed up the search for victims [2]. Actually, UAVs are used especially effectively in search and rescue operations in areas with extreme natural and climatic conditions (for example, in conditions of constant ice cover or drifting ice in the Arctic seas) [3].

In general, UAVs are mainly equipped with optoelectronic or thermal imaging detection equipment. Vision systems come to the fore to support search and rescue operations. Due to the long-range range and rather high levels of spatial and color resolution of modern linear and matrix optical radiation detectors, technical vision systems can serve as irreplaceable sources of information for automatic recognition, navigation, guidance, and information support for search and rescue operations [4, 5].

The problem of finding objects is one of the key tasks of such "assistants". Today machine learning is used as the main tool for solving such problems [1, 6, 7]. However, the use of machine learning methods to automate various UAV tasks faces a number of difficulties. First, almost all modern machine learning methods use a statistical approach to training detectors and classifiers. This means that it is necessary to prepare a fairly extensive training dataset to train classifiers and detectors. Fortunately, rescue operations do not take place very often in our world. As a result, it is not possible to collect a sufficient number of training samples. Moreover, the methods of "augmentation" (reproduction of real data) [8, 9] are also poorly applicable in this problem, since often real data are absent at all.

The second problem when developing "assistants" for UAVs is the limitations on computations. Obviously, computer vision tasks should be performed directly on board of the UAV, where in general weak but energy-efficient computers are installed [3]. Consequently, deep and heavy neural networks cannot be used to solve the object detection and recognition problem.

The solution to the above problems is training the object detector on fully synthesized data. This approach is actively used to solve various recognition problems. So, there are OCR technologies [10, 11] trained entirely on synthesized data [12, 13]. Moreover, there are methods to train deep neural networks to recognize 3D objects [14].

In addition to neural networks, other statistical methods for training detectors and classifiers are currently investigated for the applicability of synthetic training dataset. So, there are several works that consider the issue of training the Viola–Jones detector for 2D and 3D objects [15, 16]. However, it is arguable that due to the peculiarities of the Viola–Jones algorithm, it is not possible to train an effective object classifier [16].

In this paper, we consider the problem of training the Viola–Jones detector for a 3D object using fully synthetic data. The inflatable life raft PSN-10 is chosen as the object under study (see Fig. 1). Although the target infatable life raft is orange, we consider grayscale images only. This is due to the fact that in most cases UAVs are equipped with monochromatic cameras, since such cameras are more light-sensitive than color ones. The main focus of the paper is made on the method of generating a training dataset and choosing the feature space which allows to train an efficient Viola–Jones detector of the PSN-10 life raft even on the grayscale images.

1. Inflatable Life Raft PSN-10

The inflatable life raft PSN-10 (Fig. 1) is a means of collective rescue of the crew and passengers of ships [17]. The raft has a capacity of 10 people. The raft ensures that the estimated number of people is kept afloat in sea conditions, protects them from the effects of bad weather and sudden temperature fluctuations. The raft carries food and water supplies, as well as life-saving essentials and signaling equipment to facilitate the search for people in distress. The life raft consists of the following main parts: an inflatable buoyancy chamber, inflatable racks to support the awning, inflatable bottom, and double awning with a heat-insulating air gap. The life raft PSN-10 is often used on Russian high-capacity cargo ships.

Вестник ЮУрГУ. Серия «**Математическое моделирование**
и программирование» (**Вестник ЮУрГУ ММП**). 2020. Т. 13, № 4. С. 94–106

95

Fig. 1. A sample of the inflatable life raft PSN-10

2. Viola–Jones Object Detection Method

The Viola–Jones object detection method was developed for real-time face detection in images [18, 19]. This method reduces the detection problem to the problem of binary classification at each image point, that is, for each rectangular image area taken with all kinds of shifts and scales, the hypothesis of the presence of the target object in the area is checked using a pre-trained classifier.

The Viola–Jones method uses rectangular Haar-like features [20], the value of which is calculated as the difference between the sums of the pixel's intensity of the image areas inside the adjacent rectangles. For efficient calculation of the value of Haar-like features, an integral image is used. In the literature, the integral image is also known as "summed-area table" [21], which for a grayscale image $f(y, x)$ with dimensions $M \times N$ is determined as follows:

$$I_f(y, x) = \sum_{i < y, j < x} f(i, j).$$

The Viola–Jones method associates with each feature a binary "weak" classifier $h(x) : X \rightarrow \{-1, +1\}$, which is usually presented as a decision tree with one branch:

$$h(x) = \{ \begin{array}{ll} +1, & \text{if } p \cdot f(x) < p \cdot \theta; \\ -1, & \text{otherwise}, \end{array}$$

where θ and p are the threshold value of the feature and the parity of the classifier, respectively. Using the AdaBoost machine learning method, a "strong" classifier is constructed as a linear superposition of the above "weak" classifiers:

$$S(x) = \left[\sum_{t=1}^{T} \alpha_t \cdot h_t(x) > 0 \right],$$

where $[\cdot]$ is an indicator function.

The high speed of the Viola–Jones method operation is ensured through the use of a cascade of "strong" classifiers, which allows localizing "empty" (object-free) image areas in

a small number of calculations:

$$Cascade\,(x) = \prod_{i=1}^{N} [S_i\,(x) > 0]\,.$$

The object detection in the image is performed by the constructed binary cascade classifier using the sliding window method [18, 19].

2.1. Edge Haar-Like Features

In order to construct an effective and resistant to different luminance object detector, one uses different kinds of edge features. In our work we use the original feature space described in the papers [22, 23]. These features are rectangular Haar-like features calculated over the magnitude of the image gradient.

Unlike the classical Haar-like features, such edge features are effective in generalizing objects containing a huge number of edges and robust to different luminance conditions.

Fig. 2 presents a sample of PSN-10 liferaft image: both the source grayscale image and the magnitude of the image gradient computed using the L1 norm.
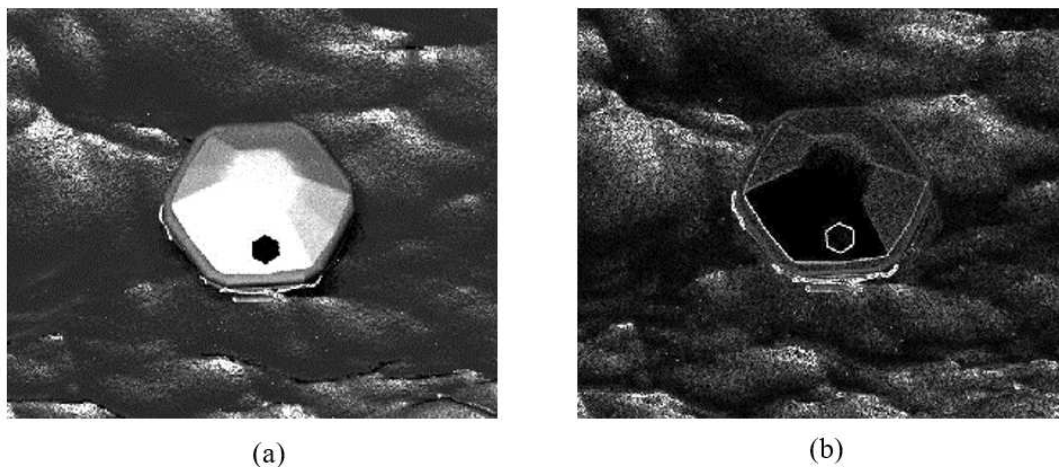

(a)   (b)

Fig. 2. A sample of PSN-10 image: (a) the source grayscale image, (b) the magnitude of the image gradient

3. Data Synthetization

To obtain 2D images, it is necessary to simulate the PSN-10 raft [17], simulate the water surface, place all objects on the scene, and create animation for the objects. In order to construct a correct 3D model, we found enough information about the life raft characteristics and a set of images. Fig. 3(a) presents a sample photo of the real PSN-10 life raft.

As a set of requirements to images, we take to be grayscale images with resolution of 1920x1080 px and contain only one item of the life raft PSN-10. The weather condition are considered to be variable, and the excitement of the water should not exceed 2 points.

All these requirements were considered in the data synthesis process described below.

Fig. 3. The life raft PSN-10: (a) the source image of the raft, (b) the 3D model of PSN-10

3.1. PSN-10 Modelling

Based on the found photos, a 3D model of the raft was constructed (see Fig. 3(b)). In order to construct the 3D model, we used the Blender software [24]. To simplify the modelling process, we decided to divide the raft into the following components: sides, awning, rope, bottom, small pillow.

After modelling, textures (images that reproduce the visual properties of any surfaces or objects) and shaders (programs for the graphics card processor that are used in 3D graphics to determine the final parameters of an object or image) must be applied to all parts of the raft.

The model is constructed with a low polygonal mesh, since the images are obtained from a long distance to the object and the angularity of some parts of the raft is not visually noticeable, but this approach allows to draw the object in the final image quicker.

3.2. Water Surface Modelling

The Blender software allows to create a water surface by applying the "Ocean" modifier to the plane. The settings allow to make the water surface of the desired size, adjust the depth, set the wave size, wave direction, wind strength, adjust the foam generation, etc.

After adjusting water surface parameters, we applied shaders to make the water appear more realistic. Empirically, we selected a combination of shaders that gives an image close to real water.

3.3. Scene Modelling

The stage of modelling the scene consists of placing objects, setting up animation, lights, camera movement, etc. Since the problem is to detect a drifting raft, the PSN-10 model moves along the wave due to the movement of water. However, it is necessary to adjust the interaction of the raft and the water surface, creating the effect of rolling the raft on the waves. The camera moves along a randomly created curve, and the object tracking function allows to keep the raft in focus.

The images should be given under different weather conditions, therefore, rain, snow and fog were added to the scene. The raindrops and snowflakes were modelled separately.
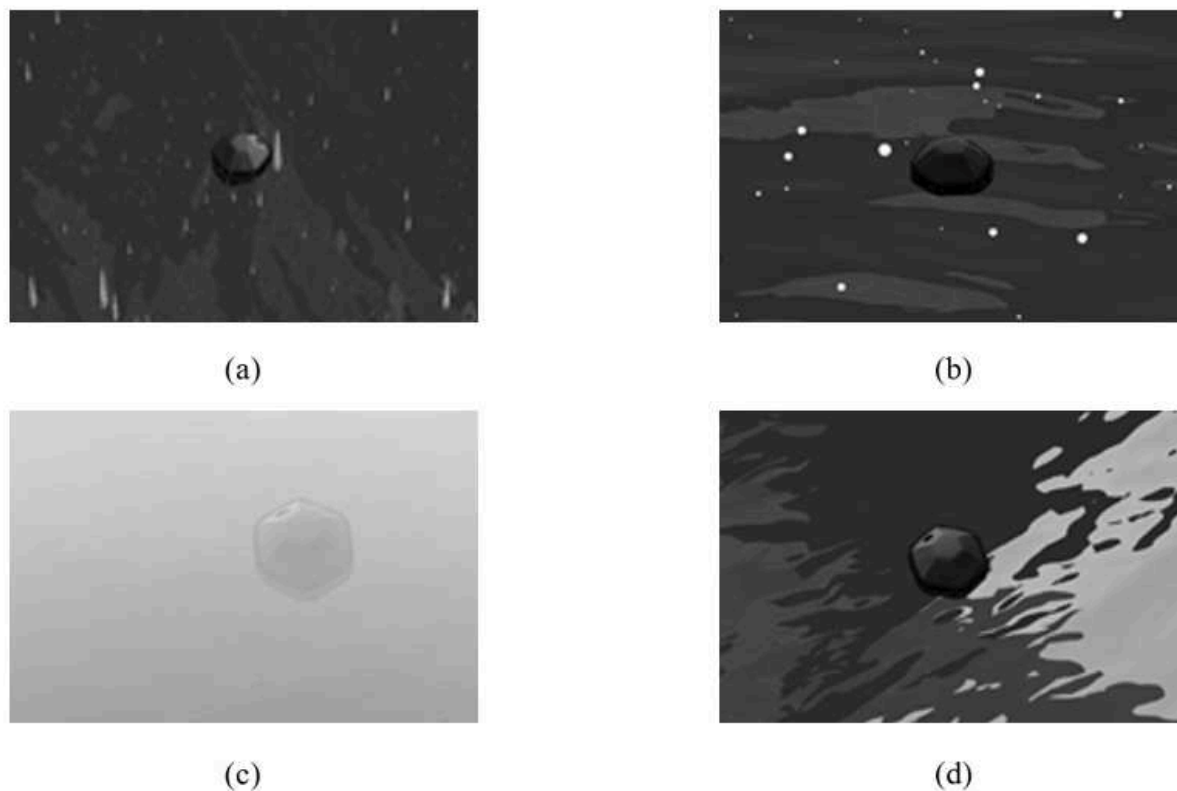
Fig. 4. Different weather conditions of modelled scenes: (a) rain, (b) snow, (c) frog, and (d) fair weather

The Blender software also has built-in toolkit to add fog to a scene. Fig. 4 presents same samples of the modelled scenes.

4. Experimental Results

In accordance with described above technique we prepared 3003 synthetic images. This dataset was separated into two parts: the first one contained 600 images and was used to train the life raft detector while the remaining part (2403 images) was used to evaluate the quality of the trained detector.

The training dataset was used to generate both positive and negative samples. The negative samples are produced by cutting "empty" sub-windows (image regions without life rafts) from the source images. We used Edge Haar-like Features as described above.

All in all, we trained the cascade classifier, which consists of 11 levels and contains 55 features. The full structure of the trained cascade classifier is presented in Table 1.

Table 1

The structure of the trained cascade
classifier of the life raft PSN-10

Level No.	1	2	3	4	5	6	7	8	9	10	11
Weak Classifier Count	2	2	3	5	4	6	7	5	5	9	7

The remaining part of synthetic dataset contained 2403 images and was used to evaluate the quality of the trained detector. Using specified dataset, we calculated the following measures: true positive (actual positives that are correctly identified), false positive (actual negatives that are classified as positives) and false negative (actual positives that are not detected with the trained detector). Based on specified statistical measures precision (positive predictive value), recall (true positive rate) and F-measure were calculated. All values are presented in Table 2.

Table 2

The quality of trained life raft detector evaluated on the synthetic data

True Positive	False Positive	False Negative	Precision	Recall	F-measure
2374	69	29	0,97176	0,98793	0,97978

In our problem, it is very important to detect all life rafts. Consequently, false negative errors are much more serious than false positive errors. That is why the value of recall measure is more important in our problem.

In order to estimate the applicability of the trained detector to real cases, we found a few images from the real rescue operation. The Russian-flagged fishing trawler Dalniy Vostok sank on 1 April 2015, off Russia's Kamchatka Peninsula in the Sea of Okhotsk. Half of the crew was rescued thanks to the fact that they successfully evacuated on time on inflatable life rafts PSN-10. There is a set of videos of this rescue operation on the Internet in free access.

Therefore, based on the video, we prepared a small dataset contained 161 images. It is clear that such a small dataset cannot be used for fully-featured evaluation of the detector but can be useful to estimate the applicability of the trained detector to real data. Fig. 5 demonstrates how the trained detector works on real data.

Fig. 5. Evaluation of the trained life raft detector on the real data (rescue operation of the Russian-flagged fishing trawler Dalniy Vostok in 2015)

Table 3 shows the quality of the trained detector on the real data. In general, the quality on the real data is less than the quality on the synthetic data. This fact can be

explained by significantly different weather conditions. The excitement of the water is big enough in the real dataset, while it was not exceed 2 points in our synthetic data.

<div align="right">**Table 3**</div>

<div align="center">The quality of trained life raft detector evaluated on the real data
(rescue operation of the Russian-flagged fishing trawler Dalniy Vostok in 2015)</div>

True Positive	False Positive	False Negative	Precision	Recall	F-measure
143	48	18	0,74869	0,88820	0,81250

Conclusion

This paper considers the problem of training a detector like Viola–Jones one for 3D objects on a fully synthetic training dataset. The PSN-10 inflatable liferaft was chosen as the object under study. To obtain training images, we simulated the PSN-10 life raft, simulated the watersurface, placed all objects on the scene, and created animation for the objects. In accordance with described technique we prepared 2403 synthetic training images.

In order to construct an effective and resistant to different luminance object detector we used original edge features described in the papers [22, 23]. To increase the computational efficiency, the L1 norm was used to calculate the magnitude of the image gradient.

To estimate the applicability of the trained detector to real cases we found a few images from the real rescue operation of the Russian-flagged fishing trawler Dalniy Vostok sank on 1 April 2015. It was shown that Viola–Jones detector trained on fully synthetic dataset sufficiently works on real data.

References

1. Hongyang Yu, Guorong Li, Weigang Zhang, Qingming Huang, Dawei Du, Qi Tian, Nicu Sebe. The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline. *International Journal of Computer Vision*, 2020, vol. 128, no. 5, pp. 1141–1159.

2. Dumin D., Dinh T.D., Pham V.D., Kirichek R. Application of Installed Systems of GSM-Device Detection on UAVs for Searching Victim in Result of Emergency Situations. *Information Technologies and Telecommunications*, 2018, vol. 6, no. 2, pp. 62–69.

3. Matveev S.A., Rudyka S.A., Petrov Yu.V., Zhdanov A.S. Onboard Complex of Information Support of Search and Rescue Operations in Arctic. *Issues of Radio Electronics*, 2019, no. 6, pp. 30–37. (in Russian) DOI: 10.21778/2218-5453-2019-6-30-37

4. Garmash V.N., Korobochkin D.M., Matveev S.A., Petrov Yu.V., Rudyka S.A., Sukhov T.M. Complexing Information from Different Sources in the On-Board Systems Search and Rescue Operations. *Issues of Radio Electronics*, 2018, no. 7, pp. 30–37. (in Russian) DOI: 10.21778/2218-5453-2018-7-139-146

5. Matveev S.A., Bizov A.N., Bistrov S.Yu., Garmash V.N., Isenko S.I., Korobochkin D.M., Petrov Yu.V., Rudika S.A., Strahov S.Yu., Sircev A.N. Helicopter System that Provide Information Support for Safety of Flights and Conduct Search and Rescue Operations. *Bulletin of the Kyrgyz-Russian Slavic University*, 2018, vol. 18, no. 12, pp. 60–64.

6. Leira F.S., Johansen T.A., Fossen T.I. Automatic Detection, Classification and Tracking of Objects in the Ocean Surface from UAVs Using a Thermal Camera. *IEEE Aerospace Conference*, Big Sky, USA, 2015, pp. 1–10.

7. Du Dawei, Yunkai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, Qi Tuan. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. *Proceedings of the European Conference on Computer Vision*, 2018, pp. 370–386.

8. Matalov D.P., Usilin S.A., Arlazarov V.V. Single-Sample Augmentation Framework for Training Viola–Jones Classifiers. *12th International Conference on Machine Vision*, Munich, Germany, 2020, no. 11433, pp. 1–9. DOI: 10.1117/12.2559435

9. Emelyanov S.O., Ivanova A.A., Shvets E.A., Nikolaev D.P. Methods of Training Data Augmentation in the Task of Image Classification. *Sensory Systems*, 2018, vol. 32, no. 3, pp. 236–245. DOI: 10.1134/S0235009218030058

10. Arlazarov V.V., Slavin O.A., Uskov A.V., Janiszewski I.M. Modelling the Flow of Character Recognition Results in Video Stream. *Bulletin of the South Ural State University: Mathematical Modelling, Programming and Computer Software*, 2018, vol. 11, no. 2, pp. 14–28. DOI: 10.14529/mmp180202

11. Bulatov K.B. A Method to Reduce Errors of String Recognition Based on Combination of Several Recognition Results with Per-Character Alternatives. *Bulletin of the South Ural State University: Mathematical Modelling, Programming and Computer Software*, 2019, vol. 12, no. 3, pp. 74–88. DOI: 10.14529/mmp190307

12. Chernyshova Y.S., Sheshkus A.V., Arlazarov V.V. Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images. *IEEE Access*, 2020, no. 8, pp. 32587–32600. DOI: 10.1109/ACCESS.2020.2974051

13. Gayer A.V., Chernyshova Y.S., Sheshkus A.V. Artificial Training Data Generation for the Task of Character Recognition of Fields of Russian Passport. *Sensory Systems*, 2018, vol. 32, no. 3, pp. 230–235. DOI: 10.1134/S023500921803006X

14. Danielczuk M., Matl M., Gupta S., Li A., Lee A., Mahler J., Goldberg K. Segmenting Unknown 3D Objects from Real Depth Images using Mask R-CNN Trained on Synthetic Data. *International Conference on Robotics and Automation*, 2019, pp. 7283–7290. DOI: 10.1109/ICRA.2019.8793744

15. Akimov A.V., Sirota A.A. Synthetic Data Generation Models and Algorithms for Training Image Recognition Algorithms Using the Viola–Jones Framework. *Computer Optics*, 2016, vol. 40, no. 6, pp. 911–918.

16. Mogelmose A., Trivedi M.M., Moeslund T.B. Learning to Detect Traffic Signs: Comparative Evaluation of Synthetic and Real-World Datasets. *Proceedings of the 21st International Conference on Pattern Recognition*, 2012, pp. 3452–3455.

17. Afanasyev I.I., Laptev V.N., Pirogov V.P. Analysis of the Rescue Assets Range of the Russian Navy. *Scientific Bulletin of the Volsk Military Institute of Material Support: Military Scientific Journal*, 2015, no. 2, p. 150–154.

18. Viola P., Jones M. Rapid Object Detection Using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition*, 2001, no. 1, pp. 511–518.

19. Viola P., Jones M. Robust Real-Time Object Detection. *International Journal of Computer Vision*, 2001, no. 4, pp. 34–47.

20. Papageorgiou C.P., Oren M., Poggio T. A General Framework for Object Detection. *Sixth International Conference Computer Vision*, 1998, vol. 6, no. 1, pp. 555–562.

21. Lewis J.P. Fast Template Matching. *Proceedings Vision Interface*, 1995, pp. 120–123.

22. Kotov A.A., Usilin S.A., Gladilin S.A., Nikolaev D.P. Construction of Robust Features for Detection and Classification of Objects without Characteristic Brightness Contrasts. *Journal of Information Technologies and Computing Systems*, 2014, no. 1, pp. 53–60.

23. Matalov D.P., Usilin S.A., Arlazarov V.V. Modification of the Viola–Jones Approach for the Detection of the Government Seal Stamp of the Russian Federation. *Eleventh International Conference on Machine Vision*, 2019, pp. 11041. DOI: 10.1117/12.2522793

24. *Home of the Blender Project – Free and Open 3D Creation Software.* 2020. Available at: https://www.blender.org/

ОБУЧЕНИЕ ДЕТЕКТОРОВ ВИОЛЫ – ДЖОНСА ДЛЯ 3D-ОБЪЕКТОВ НА ОСНОВЕ ПОЛНОСТЬЮ СИНТЕТИЧЕСКИХ ДАННЫХ ДЛЯ ИСПОЛЬЗОВАНИЯ В СПАСАТЕЛЬНЫХ МИССИЯХ С БПЛА

С.А. Усилин[1−3], *В.В. Арлазаров*[1−4], *Н.С. Рохлин*[5], *С.А. Рудыка*[5], *С.А. Матвеев*[5], *А.А. Зацаринный*[2]

[1]ООО «Смарт Энджинс Сервис», г. Москва, Российская Федерация
[2]Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Российская Федерация
[3]Московский физико-технический институт, г. Москва, Российская Федерация
[4]Институт проблем передачи информации имени А.А. Харкевича РАН, г. Москва, Российская Федерация
[5]Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, г. Санкт-Петербург, Российская Федерация

В работе рассматривается задача обучения детектора Виолы – Джонса для 3D объектов на примере надувного спасательного плота ПСН-10. Обучение детектора выполняется на полностью синтетическом обучающем наборе. В работе подробно рассматриваются способы моделирования надувного спасательного плота, водной поверхности, различных погодных условий. В качестве признакового пространства используются граничные признаки, позволяющие обучить детектор, устойчивый к различным условиям освещения. Для повышения вычислительной эффективности при вычислении значения градиента использовалась норма L1. Эффективность обученного детектора оценена в том числе на реальных данных, полученных в процессе спасательной операции траулера «Дальний Восток». Предложенный в работе способ обучения детекторов Виолы – Джонса может быть успешно использован в качестве составляющего элемента программно-аппаратных «ассистентов» БПЛА.

Ключевые слова: машинное обучение; поиск объектов; Виола – Джонс; классификация; 3D-объекты; БПЛА; спасательная миссия.

Литература

1. Hongyang Yu. The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline / Hongyang Yu, Guorong Li, Weigang Zhang, Qingming Huang, Dawei Du, Qi Tian, Nicu Sebe // International Journal of Computer Vision. – 2020. – V. 128, № 5. – P. 1141–1159.

2. Dumin, D. Application of Installed Systems of GSM-Device Detection on UAVs for Searching Victim in Result of Emergency Situations / D. Dumin, T.D. Dinh, V.D. Pham, R. Kirichek // Information Technologies and Telecommunications. – 2018. – V. 6, № 2. – P. 62–69.

3. Матвеев, С.А. Бортовой комплекс информационной поддержки проведения поисково-спасательных операций в условиях Арктики / С.А. Матвеев, С.А. Рудыка, Ю.В. Петров, А.С. Жданов // Вопросы радиоэлектроники. – 2019. –№ 6. – С. 30–37.

4. Гармаш, В.Н. Комплексирование информации от разнородных источников в бортовых комплексах обеспечения поисково-спасательных операций / В.Н. Гармаш, Д.М. Коробочкин, С.А. Матвеев, Ю.В. Петров, С.А. Рудыка, Т.М. Сухов // Вопросы радиоэлектроники. – 2018. – №. 7. – С. 30–37.

5. Матвеев, С.А. Вертолетный комплекс информационной поддержки безопасности полетов и проведения поисково-спасательных операций / С.А. Матвеев, А.Н. Бызов, С.Ю. Быстров, В.Н. Гармаш, С.И. Исенков, Д.М. Коробочкин, Ю.В. Петров, С.А. Рудыка, С.Ю. Страхов, А.Н. Сырцев // Вестник Кыргызско-российского славянского университета. – 2018. – Т. 18, № 12. – С. 60–64.

6. Leira, F.S. Automatic Detection, Classification and Tracking of Objects in the Ocean Surface from UAVs Using a Thermal Camera / F.S. Leira, T.A. Johansen, T.I. Fossen // IEEE Aerospace Conference, Big Sky, USA. – 2015. – P. 1–10.

7. Dawei Du. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking / Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, Qi Tian // Proceedings of the European Conference on Computer Vision. – 2018. – P. 370–386.

8. Matalov, D.P. Single-Sample Augmentation Framework for Training Viola – Jones Classifiers / D.P. Matalov, S.A. Usilin, V.V. Arlazarov // 12th International Conference on Machine Vision, Munich, Germany. – 2020. – № 11433. – P. 1–9.

9. Емельянов, С.О. Методы аугментации обучающих выборок в задачах классификации изображений / С.О. Емельянов, А.А. Иванова, Е.А. Швец, Д.П. Николаев // Сенсорные системы. – 2018. – Т. 32, № 3. – С. 236–245.

10. Арлазаров, В.В. Моделирование потока результатов распознавания символов в видео-последовательностях / В.В. Арлазаров, О.А. Славин, А.В. Усков, И.М. Янишевский // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2018. – Т. 11, № 2. – С. 14–28.

11. Булатов, К.Б. Метод уменьшения числа ошибок распознавания строки, основанный на комбинировании множества результатов распознавания с использованием альтернатив символов / К.Б. Булатов // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2019. – Т. 12, № 3. – С. 74–88.

12. Chernyshova, Y.S. Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images / Y.S. Chernyshova, A.V. Sheshkus, V.V. Arlazarov // IEEE Access. – 2020. – № 8. – P. 32587–32600.

13. Gayer, A.V. Artificial Training Data Generation for the Task of Character Recognition of Fields of Russian Passport / A.V. Gayer, Y.S. Chernyshova, A.V. Sheshkus // Sensory Systems. – 2018. – V. 32, № 3. – P. 230–235.

14. Danielczuk, M. Segmenting Unknown 3D Objects from Real Depth Images using Mask R-CNN Trained on Synthetic Data / M. Danielczuk, M. Matl, S. Gupta, A. Li, A. Lee, J. Mahler, K. Goldberg // International Conference on Robotics and Automation. – 2019. – P. 7283–7290.

15. Akimov, A.V. Synthetic Data Generation Models and Algorithms for Training Image Recognition Algorithms Using the Viola – Jones Framework / A.V. Akimov, A.A. Sirota // Computer Optics. – 2016. – V. 40, № 6. – P. 911–918.

16. Mogelmose, A. Learning to Detect Traffic Signs: Comparative Evaluation of Synthetic and Real-World Datasets / A. Mogelmose, M.M. Trivedi, T.B. Moeslund // Proceedings of the 21st International Conference on Pattern Recognition. – 2012. – P. 3452–3455.

17. Afanasyev, I.I. Analysis of the Rescue Assets Range of the Russian Navy / I.I. Afanasyev, V.N. Laptev, V.P. Pirogov // Scientific Bulletin of the Volsk Military Institute of Material Support: Military Scientific Journal. – 2015. – № 2. – P. 150–154.

18. Viola, P. Rapid Object Detection Using a Boosted Cascade of Simple Features / P. Viola, M. Jones // Computer Vision and Pattern Recognition. – 2001. – № 1. – P. 511–518.

19. Viola, P. Robust Real-Time Object Detection / P. Viola, M. Jones // International Journal of Computer Vision. – 2001. – № 4. – P. 34–47.

20. Papageorgiou, C.P. A General Framework for Object Detection / C.P. Papageorgiou, M. Oren, T. Poggio // Sixth International Conference Computer Vision. – 1998. – V. 6, № 1. – P. 555–562.

21. Lewis, J.P. Fast Template Matching / J.P. Lewis // Proceedings Vision Interface. – 1995. – P. 120–123.

22. Котов, А.А. Построение устойчивых признаков детекции и классификации объектов, не обладающих характерными яркостными контрастами / А.А. Котов, С.А. Усилин, С.А. Гладилин, Д.П. Николаев // Информационные технологии и вычислительные системы. – 2014. – №. 1. – С. 53–60.

23. Matalov, D.P. Modification of the Viola – Jones Approach for the Detection of the Government Seal Stamp of the Russian Federation / D.P. Matalov, S.A. Usilin, V.V. Arlazarov // Eleventh International Conference on Machine Vision. – 2019. – P. 11041.

24. Home of the Blender Project – Free and Open 3D Creation Software. – 2020. – URL: https://www.blender.org/

Сергей Александрович Усилин, кандидат технических наук, исполнительный директор, ООО «Смарт Энджинс Сервис» (г. Москва, Российская Федерация); старший научный сотрудник, Федеральный исследовательский центр «Информатика и управление» РАН (г. Москва, Российская Федерация); преподаватель, Московский физико-технический институт (г. Москва, Российская Федерация), usilin@smartengines.com.

Владимир Викторович Арлазаров, кандидат технических наук, генеральный директор, ООО «Смарт Энджинс Сервис» (г. Москва, Российская Федерация); заведующий отделом, Федеральный исследовательский центр «Информатика и управление» РАН (г. Москва, Российская Федерация); и.о. ведущего научного сотрудника, Институт проблем передачи информации имени А.А. Харкевича РАН (г. Москва, Российская Федерация); преподаватель, Московский физико-технический институт (г. Москва, Российская Федерация), vva@smartengines.com.

Николай Сергеевич Рохлин, инженер, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (г. Санкт-Петербург, Российская Федерация), 5hark1@rambler.ru.

Станислав Анатольевич Рудыка, начальник научно-исследовательской части, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (г. Санкт-Петербург, Российская Федерация), rudika_sa@voenmeh.ru.

Станислав Алексеевич Матвеев, кандидат технических наук, проректор по научной работе и инновационному развитию, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (г. Санкт-Петербург, Российская Федерация), matveev_sa@voenmeh.ru.

Александр Алексеевич Зацаринный, доктор технических наук, профессор, заместитель директора, Федеральный исследовательский центр «Информатика и управление» РАН (г. Москва, Российская Федерация), azatsarinny@ipiran.ru.