Real-Time Embedded Computer Vision on UAVs
UAVision2018 Workshop Summary

Kristof Van Beeck[1(✉)], Tinne Tuytelaars[2], Davide Scarramuza[3], and Toon Goedemé[1]

[1] EAVISE, Campus De Nayer, KU Leuven, Sint-Katelijne-Waver, Belgium
kristof.vanbeeck@kuleuven.be
[2] PSI, KU Leuven, Leuven, Belgium
[3] Robotics and Perception Group, ETH Zürich, Zürich, Switzerland

Abstract. In this paper we present an overview of the contributed work presented at the UAVision2018 ECCV workshop. This workshop focused on real-time image processing on-board of Unmanned Aerial Vehicles (UAVs). For such applications the computational complexity of state-of-the-art computer vision algorithms often conflicts with the need for real-time operation and the extreme resource limitations of the hardware. Apart from a summary of the accepted workshop papers, this work also aims to identify common challenges and concerns which were addressed by multiple authors during the workshop, and their proposed solutions.

Keywords: Computer vision · Real-time · UAVs · Embedded hardware · Deep learning · GPUs · Hardware optimizations

1 Introduction

This paper contains a summary of the material presented at the 2nd International Workshop on Computer Vision for UAVs (UAVision 2018). This workshop took place in conjunction with ECCV2018, Munich, Germany on Saturday the 8th of September 2018. Apart from a brief summarization of each paper, we also identified a number of common concerns, challenges and possible proposed solutions that several authors addressed during the workshop.

This workshop focused on state-of-the-art real-time image processing on-board of Unmanned Aerial Vehicles. Indeed, cameras make ideal sensors for drones as they are lightweight, power-efficient and an enormously rich source of information about the environment in numerous applications. Although lots of information can be derived from camera images using the newest computer vision algorithms, the use of them on-board of UAVs poses unique challenges. Their computational complexity often conflicts with the need for real-time operation and the extreme resource limitations of the platform. Of course, developers have the choice to run their image processing on-board or on a remote processing device, although the latter requires a wireless link with high bandwidth,

L. Leal-Taixé and S. Roth (Eds.): ECCV 2018 Workshops, LNCS 11130, pp. 3–10, 2019.
https://doi.org/10.1007/978-3-030-11012-3_1

minimal latency and ultra-reliable connection. Indeed, truly autonomous drones should not have to rely on a wireless datalink, thus on-board real-time processing is a necessity. However, because of the limitations of UAVs (lightweight processing devices, limited on-board computational power, limited electrical power on-board), extreme algorithmic optimization and deployment on state-of-the-art embedded hardware (such as embedded GPUs) is the only solution. In this workshop we focused on enabling embedded processing in drones, making efficient use of specific embedded hardware and highly optimizing computer vision algorithms towards real-time applications.

The remainder of this paper is structured as follows. Section 2 gives an overview and short summary of each presented paper at our workshop. In Sect. 3 we discuss the challenges that were identified by multiple authors during the workshop and their proposed solutions. Finally, we conclude this work in Sect. 4.

2 Contributed Papers

In total nine papers were accepted for publication at the UAVision2018 workshop. The first four papers listed below were accepted as full oral presentation (i.e. 20 min), whereas the five consecutive papers were accepted as short oral presentation (i.e. 15 min). Below we list and summarize each paper using the paper abstracts.

2.1 Teaching UAVs to Race: End-to-End Regression of Agile Controls in Simulation [7]

Automating the navigation of unmanned aerial vehicles (UAVs) in diverse scenarios has gained much attention in recent years. However, teaching UAVs to fly in challenging environments remains an unsolved problem, mainly due to the lack of training data. In this paper [7], the authors trained a deep neural network to predict UAV controls from raw image data for the task of autonomous UAV racing in a photo-realistic simulation. Training is done through imitation learning with data augmentation to allow for the correction of navigation mistakes. Extensive experiments demonstrate that our trained network (when sufficient data augmentation is used) outperforms state-of-the-art methods and flies more consistently than many human pilots. Additionally, we show that our optimized network architecture can run in real-time on embedded hardware, allowing for efficient onboard processing critical for real-world deployment.

2.2 Onboard Hyperspectral Image Compression Using Compressed Sensing and Deep Learning [2]

This paper [2] proposes a real-time onboard compression scheme for hyperspectral datacube which consists of a very low complexity encoder and a deep learning based parallel decoder architecture for fast decompression. The encoder creates a set of coded snapshots from a given datacube using a measurement code matrix.

The decoder decompresses the coded snapshots by using a sparse recovery algorithm. The authors solve this sparse recovery problem using a deep neural network for fast reconstruction. We present experimental results which demonstrate that our technique performs very well in terms of quality of reconstruction and in terms of computational requirements compared to other transform based techniques with some tradeoff in PSNR. The proposed technique also enables faster inference in compressed domain, suitable for on-board requirements.

2.3 SafeUAV: Learning to Estimate Depth and Safe Landing Areas for UAVs from Synthetic Data [5]

The emergence of relatively low cost UAVs has prompted a global concern about the safe operation of such devices. Since most of them can 'autonomously' fly by means of GPS way-points, the lack of a higher logic for emergency scenarios leads to an abundance of incidents involving property or personal injury. In order to tackle this problem, this paper [5] proposed a small, embeddable ConvNet for both depth and safe landing area estimation. Furthermore, since labeled training data in the 3D aerial field is scarce and ground images are unsuitable, the authors captured a novel synthetic aerial 3D dataset obtained from 3D reconstructions. They used the synthetic data to learn to estimate depth from in-flight images and segmented them into 'safe-landing' and 'obstacle' regions. Experiments demonstrated compelling results in practice on both synthetic data and real RGB drone footage.

2.4 Aerial GANeration: Towards Realistic Data Augmentation Using Conditional GANs [6]

Environmental perception for autonomous aerial vehicles is a rising field. Recent years have shown a strong increase of performance in terms of accuracy and efficiency with the aid of convolutional neural networks. Thus, the community has established data sets for benchmarking several kinds of algorithms. However, public data is rare for multi-sensor approaches or either not large enough to train very accurate algorithms. For this reason, this paper [6] proposed a method to generate multi-sensor data sets using realistic data augmentation based on conditional generative adversarial networks (cGAN). cGANs have shown impressive results for image to image translation. The authors used this principle for sensor simulation. Hence, there is no need for expensive and complex 3D engines. The method encodes ground truth data, e.g. semantics or object boxes that could be drawn randomly, in the conditional image to generate realistic consistent sensor data. Their method is proven for aerial object detection and semantic segmentation on visual data, such as 3D Lidar reconstruction using the ISPRS and DOTA data set. The authors demonstrate qualitative accuracy improvements for state-of-the-art object detection (YOLO) using this augmentation technique.

2.5 Metrics for Real-Time Mono-VSLAM Evaluation Including IMU Induced Drift with Application to UAV Flight [3]

Vision based algorithms became popular for state estimation and subsequent (local) control of mobile robots. Currently a large variety of such algorithms exists and their performance is often characterized through their drift relative to the total trajectory traveled. However, this metric has relatively low relevance for local vehicle control/stabilization. In this paper [3], the authors proposed a set of metrics which allows to evaluate a vision based algorithm with respect to its usability for state estimation and subsequent (local) control of highly dynamic autonomous mobile platforms such as multirotor UAVs. As such platforms usually make use of inertial measurements to mitigate the relatively low update rate of the visual algorithm, they particularly focused on a new metric taking the expected IMU-induced drift between visual readings into consideration based on the probabilistic properties of the sensor. The authors demonstrated this set of metrics by comparing ORB-SLAM, LSD-SLAM and DSO on different datasets.

2.6 ShuffleDet: Real-Time Vehicle Detection Network in On-board Embedded UAV Imagery [1]

On-board real-time vehicle detection is of great significance for UAVs and other embedded mobile platforms. In this paper [1] the authors present a computationally inexpensive detection network for vehicle detection in UAV imagery which we call ShuffleDet. In order to enhance the speed-wise performance, we construct our method primarily using channel shuffling and grouped convolutions. We apply inception modules and deformable modules to consider the size and geometric shape of the vehicles. ShuffleDet is evaluated on CARPK and PUCPR+ datasets and compared against the state-of-the-art real-time object detection networks. ShuffleDet achieves 3.8 GFLOPs while it provides competitive performance on test sets of both datasets. We show that our algorithm achieves real-time performance by running at the speed of 14 frames per second on NVIDIA Jetson TX2 showing high potential for this method for real-time processing in UAVs.

2.7 Joint Exploitation of Features and Optical Flow for Real-Time Moving Object Detection on Drones [4]

Moving object detection is an imperative task in computer vision, where it is primarily used for surveillance applications. With the increasing availability of low-altitude aerial vehicles, new challenges for moving object detection have surfaced, both for academia and industry. In this paper [4], the authors proposed a new approach that can detect moving objects efficiently and handle parallax cases. By introducing sparse ow based parallax handling and downscale processing, they pushed the boundaries of real-time performance with 16 FPS on limited embedded resources (a five-fold improvement over existing baselines),

while managing to perform comparably or even improve the state-of-the-art in two different datasets. They also presented a roadmap for extending our approach to exploit multi-modal data in order to mitigate the need for parameter tuning.

2.8 UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition [8]

Current UAV-recorded datasets were mostly limited to action recognition and object tracking, whereas the gesture signals datasets were mostly recorded in indoor spaces. Currently, there is no outdoor recorded public video dataset for UAV commanding signals. To fill this gap and enable research in wider application areas, this paper [8] presented a UAV gesture signals dataset recorded in an outdoor setting. The authors selected 13 gestures suitable for basic UAV navigation and command from general aircraft handling and helicopter handling signals. They provide 119 high-definition video clips consisting of 37151 frames. All the frames are annotated for the body joints and gesture classes in order to extend the dataset's applicability to a wider research area including gesture recognition, action recognition, human pose recognition and situation awareness.

2.9 ChangeNet: A Deep Learning Architecture for Visual Change Detection [9]

The increasing urban population in cities necessitates the need for the development of smart cities that can offer better services to its citizens. Drone technology plays a crucial role in the smart city environment and is already involved in a number of functions in smart cities such as traffic control and construction monitoring. A major challenge in fast growing cities is the encroachment of public spaces. A robotic solution using visual change detection can be used for such purposes. For the detection of encroachment, a drone can monitor outdoor urban areas over a period of time to infer the visual changes. Visual change detection is a higher level inference task that aims at accurately identifying variations between a reference image (historical) and a new test image depicting the current scenario. In case of images, the challenges are complex considering the variations caused by environmental conditions that are actually unchanged events. Human mind interprets the change by comparing the current status with historical data at intelligence level rather than using only visual information. In this paper [9], the authors presented a deep architecture called ChangeNet for detecting changes between pairs of images and express the same semantically (label the change). A parallel deep convolutional neural network (CNN) architecture for localizing and identifying the changes between image pair has been proposed in this paper. The architecture is evaluated with VL-CMU-CD street view change detection, TSUNAMI and Google Street View (GSV) datasets that resemble drone captured images. The performance of the model for different lighting and seasonal conditions are experimented quantitatively and qualitatively. The result shows that ChangeNet outperforms the state of the art by

achieving 98.3% pixel accuracy, 77.35% object based Intersection over Union (IoU) and 88.9% area under Receiver Operating Characteristics (RoC) curve.

3 Discussion: Trends and Solutions to Common Challenges

Throughout the workshop we identified a number of common concerns for UAV vision applications that multiple authors identified and proposed solutions for. Below we give an overview.

3.1 Potential of Deep Learning for UAV Applications

One main message is that the success of deep learning based techniques also extends towards UAV applications. Almost every author in the workshop made use of deep learning for their specific drone application. For example, Marcu *et al.* [5] proposed a neural network that is trained to detect flat ground surfaces upon which a UAV can land safely. Also, a CNN that is able to detect scene changes from UAV drone images, without being distracted by seasonal effects like snow and fallen leaves was presented by Varghese *et al.* [9]. A remarkable result was shown in the work of Kumar *et al.* [2], where they show that for multispectral data decompression, their proposed deep learning alternative is even substantially faster than the classic mathematical approach.

3.2 Collecting Training Data for UAV Applications

A difficulty many drone vision researcher struggles with is how to gather enough visual training material to train these neural networks with. Indeed, because of the inherent viewpoint freedom a flying drone has, it is very difficult to acquire real UAV image data that has enough variance. Quite a few papers in the Uavision workshop tackled this problem, in very diverse ways.

The straightforward manner is setting up a large data recording campaign with real drones, pilots and actors. This is only feasible for a constrained application because of the manual labour and hence the cost. Perera *et al.* [8] did this and presented on this workshop a newly recorded dataset for gesture recognition from drone images.

However, many authors seek the answer of this in other data sources, which can be used for training a visual drone application. As in other computer vision applications, the use of rendered synthetic data from simulation engines shows potential for UAV too, as demonstrated by Müller *et al.* [7], using Sim4CV to build a virtual environment to train a racing drone.

Another example is the work of Marcu *et al.* [5], in which the authors used 3D Google Earth data as training material for a drone to learn where it is safe to land.

In this workshop, other work from Milz *et al.* [6] showed the potential of cGANs to generate data to train a UAV application, yielding a virtually infinite source of relevant training data.

3.3 Real-Time On-board Processing

The participants of this UAVision workshop all agreed that on-board processing is a must for real-time UAV vision applications. The second speaker [2] stated this very strictly: for hyperspectral video transmission from UAVs, their is simply not enough bandwidth available. On-board compression is hence a necessity. Also, Hardt-Stremayr [3] concluded in his talk about metrics for UAV vision-based SLAM that they need video processing with a frame-rate of at least 10 fps, in order to keep the drift error caused by the IMU low enough.

Many authors showed successful implementations of deep learning based image interpretation algorithms that indeed can run in real-time on embedded hardware. We noticed that the NVIDIA Jetson TX2 platform is a popular choice in this field. For example, Müller [7] estimated that their drone racing model (running at 556 fps on a NVIDIA TitanX), will run at about 50 fps on a Jetson TX2 platform, largely fast enough for real-time processing.

Another example is the presented work of Lezki *et al.* [4], who reached real-time performance with 16 FPS on limited embedded resources (a 5× improvement) for their moving objects detection, by introducing sparse parallax handling and downscaling processing.

Indeed, also Kumar *et al.* [2] demonstrated a speed-up factor of 30× for their hyperspectral decompression algorithm as compared to the baseline, indicating that two-digit speed-up factors can be achieved in many cases.

Last but not least, in their talk on ShuffleDet, Azimi *et al.* [1] pulled out all the stops for developing a ultimately efficient object detector. By exploiting group convolutions, channel shuffling, and depth wise convolutions, they achieved a 14× speed-up as compared to the already very time-optimal YOLO detector.

4 Conclusion

This paper summarized the contributed work which was presented at the UAVision2018 workshop (in conjunction with ECCV2018), and tried to identify common concerns and challenges that were recognized by multiple authors, and their proposed solutions. Three significant trends were discovered. First, the use of deep learning for (embedded) UAV applications seems viable, despite their increased computational complexity. Secondly, the collection of sufficient training data remains difficult, and several authors therefore use synthetically generated images. Finally, although real-time computer vision processing on-board of UAVs on low-power embedded hardware platforms remains challenging, several authors were able to present real-time implementations through extreme software and/or hardware optimizations.

Acknowledgements. This work is supported by the agency Flanders Innovation & Entrepreneurship (VLAIO) and Research Foundation - Flanders (FWO).

References

1. Azimi, S.M.: ShuffleDet: real-time vehicle detection network in on-board embedded. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 88–99. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
2. Kumar, S., Chaudhuri, S., Banerjee, B., Ali, F.: Onboard hyperspectral image compression using compressed sensing and deep learning. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 30–42. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
3. Hardt-Stremayr, A., Schörghuber, M., Weiss, S., Humenberger, M.: Metrics for real-time mono-VSLAM evaluation including IMU induced drift with application to UAV flight. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 73–87. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
4. Lezki, H., et al.: Joint exploitation of features and optical flow for real-time moving object detection on drones. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 100–116. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
5. Marcu, A., Costea, D., Licăreţ, V., Leordeanu, M., Pîrvu, M., Sluşanschi, E.: SafeUAV: learning to estimate depth and safe landing areas for UAVs from synthetic data. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 43–58. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
6. Milz, S., Rüdiger, T., Süss, S.: Aerial GANeration: towards realistic data augmentation using conditional GANs. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 59–72. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
7. Müller, M., Casser, V., Smith, N., Michels, D.L., Ghanem, B.: Teaching UAVs to race: end-to-end regression of agile controls in simulation. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 11–29. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
8. Perera, A.G., Law, Y.M., Chahl, J.: UAV-GESTURE: a dataset for UAV control and gesture recognition. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 117–128. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z
9. Varghese, A., Gubbi, J.: ChangeNet: a deep learning architecture for visual change detection. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11130, pp. 129–145. Springer, Cham (2019). https://doi.org/10.1007/97d8-3-030-11012-3_z