UAVM '23: 2023 Workshop on UAVs in Multimedia: Capturing the World from a New Perspective

Zhedong Zheng
National University of Singapore
Singapore
zdzheng@nus.edu.sg

Yujiao Shi
Australian National University
Australia
yujiao.shi@anu.edu.au

Tingyu Wang
Hangzhou Dianzi University
China
wongtyu@hdu.edu.cn

Jun Liu
Singapore University of Technology and Design
Singapore
jun_liu@sutd.edu.sg

Jianwu Fang
Changán University
China
fangjianwu@chd.edu.cn

Yunchao Wei
Beijing Jiaotong University
China
yunchao.wei@bjtu.edu.cn

Tat-seng Chua
National University of Singapore
Singapore
dcscts@nus.edu.sg

ABSTRACT

Unmanned Aerial Vehicles (UAVs), also known as drones, have become increasingly popular in recent years due to their ability to capture high-quality multimedia data from the sky. With the rise of multimedia applications, such as aerial photography, cinematography, and mapping, UAVs have emerged as a powerful tool for gathering rich and diverse multimedia content. This workshop aims to bring together researchers, practitioners, and enthusiasts interested in UAV multimedia to explore the latest advancements, challenges, and opportunities in this exciting field. The workshop covers various topics related to UAV multimedia, including aerial image and video processing, machine learning for UAV data analysis, UAV swarm technology, and UAV-based multimedia applications. In the context of the ACM Multimedia conference, this workshop is highly relevant as multimedia data from UAVs is becoming an increasingly important source of content for many multimedia applications. The workshop provides a platform for researchers to share their work and discuss potential collaborations, as well as an opportunity for practitioners to learn about the latest developments in UAV multimedia technology. Overall, this workshop provides a unique opportunity to explore the exciting and rapidly evolving field of UAV multimedia and its potential impact on the wider multimedia community.

CCS CONCEPTS

• **Computing methodologies → Scene understanding**; **Visual content-based indexing and retrieval**; **Vision for robotics**.

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10.
https://doi.org/10.1145/3581783.3610937

KEYWORDS

UVA Multimedia Understanding, Drone-based Video Analysis

ACM Reference Format:
Zhedong Zheng, Yujiao Shi, Tingyu Wang, Jun Liu, Jianwu Fang, Yunchao Wei, and Tat-seng Chua. 2023. UAVM '23: 2023 Workshop on UAVs in Multimedia: Capturing the World from a New Perspective. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3581783.3610937

1 BACKGROUND AND MOTIVATION

Unmanned Aerial Vehicles (UAVs), also known as drones, have become increasingly popular in recent years due to their ability to capture high-quality multimedia data from the sky. This has opened up a wide range of possibilities for applications such as aerial photography [26], cinematography [1, 11], mapping [8, 21, 28], agriculture [5, 12], geo-localization [9, 14–18] and delivery [2, 19]. UAVs have emerged as a powerful tool for gathering rich and diverse multimedia content, providing a unique vantage point with less occlusions [10, 24, 25] and the ability to capture data from previously inaccessible or hard-to-reach locations (see Figure 1). The use of UAVs in multimedia applications has become even more significant in recent years with the emergence of new technologies such as machine learning, computer vision, and big data analytics [20, 22]. These technologies have the potential to revolutionize the way UAVs are used to capture and analyze multimedia content, opening up new possibilities for applications such as automated image and video analysis [6, 7, 10, 29], real-time tracking [23, 25], predictive modeling [4, 13], and citywide simulation [27]. The ACM Multimedia conference has been at the forefront of multimedia research for over 31 years, providing a forum for researchers and practitioners to exchange ideas, explore the latest advancements, and discuss the challenges facing the field. In recent years, the conference has also started to cover the topic of UAV multimedia, recognizing the growing importance of this area of research [3, 24, 25, 28]. The

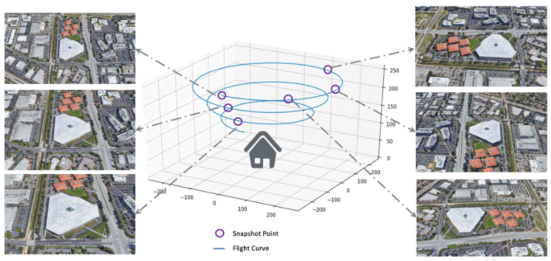

Figure 1: Different from conventional devices, UAV is a controllable aerial information capture platform, and multi-view information helps to establish a robust target model.

Figure 2: Distractor samples in University160k, with diverse building types.

use of UAVs to capture multimedia data is becoming increasingly common, and the data captured by UAVs is becoming an important source of content for many multimedia applications. Therefore, we think it is good timing to hold a workshop for people with different backgrounds to communicate in person.

2 TOPICS AND THEMES

Topics covered in this workshop (but not limited to) is as follows:

- Video-based UAV Navigation
- UAV Swarm Coordination
- UAV-based Object Detection and Tracking
- UAV-based Sensing and Mapping
- UAV-based Delivery and Transportation

3 PAPER SUBMISSION AND REVIEWING

3.1 Challenge Dataset

We also provide a challenging cross-view geo-localization dataset, called University160k, and the workshop audience may consider to participate the competition. The motivation is to simulate the real-world geo-localization scenario that we usually face an extremely large satellite-view pool. In particular, University160k extends the current University-1652 dataset [28] with extra 167,486 satellite-view gallery distractors. We have released University160k on our website, and made a public leader board. These distractor satellite-view images have a size of 1024×1024 and are obtained by cutting orthophoto images of real urban and surrounding areas. The larger image size ensures higher image clarity, while the wider framing range allows the images to contain more diverse scenes, such as buildings, city roads, trees, fields, and more (see Figure 2). In our primary evaluation, the distractor is challenging and make the competitive baseline model, LPN [21], decrease the Recall@1 accuracy from 75.93% to 64.85% and the value of AP from 79.14% to 67.69% in the Drone → Satellite task (Please see Table 1). We hope more audiences can be involved to solve this challenge, and also consider the efficiency problem against a large candidate pool.

Table 1: Here, we show the result on different subsets. The competitive baseline model's performance drops significantly with more distractors added. Therefore, we call for the audience to design a robust algorithm against our challenging large-scale satellite pool to minimize such performance gaps.

Dataset	#Distractor	LPN	
		Drone → Satellite	
		R@1	AP
University-1652 [28]	0	**75.93**	**79.14**
+ Subset1-lian1	18,155	68.09	71.18
+ Subset2-kai1	43,728	71.62	74.42
+ Subset3-lian2	37,522	69.10	72.03
+ Subset4-kai2	6,8081	69.03	71.95
University160k	167,486	64.85 (-11.08)	67.69 (-11.45)

Table 2: Summary of the top-5 valid challenge results.

Rank	Team Name	R@1	R@10
1	Skyy93	95.71	98.91
2	lzf	94.48	98.15
3	lihaoran	92.77	97.03
4	huhuhuhu	90.02	94.83
5	zhangbing	85.76	95.09
-	baseline	64.85	67.89

3.2 Challenge Results

The ACMMM23 Multimedia Drone Satellite Matching Challenge attract 55 registered participants, and 22 teams from various countries provide results in the final phase. The competition server was running on the Codalab [1]. Table 2 summarizes the top-5 valid challenge results that largely outperform the baseline. The solution of top teams has been open-sourced at the leaderboard [2].

4 ORGANIZER INFORMATION

Zhedong Zheng (https://zdzheng.xyz) is a research fellow at School of Computing, National University of Singapore. He received the Ph.D. degree from the University of Technology Sydney, Australia, in 2021 and the B.S. degree from Fudan University, China, in 2016. He received the IEEE Circuits and Systems Society Outstanding Young Author Award of 2021. He has served as the reviewer and program committee (PC) member for multiple conferences and journals, including TPAMI, TMM, IJCV, CVPR, ICCV, ECCV, IJCAI, AAAI and ACM Multimedia, and organized a special session on reliable retrieval at ICME 2022.

Yujiao Shi (https://shiyujiao.github.io/) is a research fellow at the Australian National University where she did her Ph.D. degree. She received the M.S and B.S degree from Nanjing University of Posts and Telecommunications, China, in 2014 and 2017, respectively. Her research interests include multi-modal retrieval, registration and translation, 3D vision, and self-supervised learning. She has published seven first-author papers in aerial image-related tasks in top-tier conferences, including CVPR, NeurIPS, TPAMI *et al.*.

Tingyu Wang (https://scholar.google.com/citations?user=wv3H-F4AAAAJ) is an assistant research professor at the Hangzhou Dianzi University, Hangzhou, China, where he obtained his Ph.D. and M.S. degrees. He received the B.S. degree in automation from Yantai University, Shandong, China, in 2015. His research interests include deep learning, image retrieval and remote sensing.

[1] https://codalab.lisn.upsaclay.fr/competitions/12672
[2] https://github.com/layumi/UAVM2023

Jun Liu (https://istd.sutd.edu.sg/people/faculty/liu-jun) is currently an assistant professor at Singapore University of Technology and Design. He received the PhD degree from Nanyang Technological University, the MSC degree from Fudan University, and the BEng degree from Central South University. His research interests include computer vision and artificial intelligence. He is an Associate Editor of IEEE Transactions on Image Processing and IEEE Transactions on Biometrics, Behavior, and Identity Science, and Area Chair of ICML, NeurIPS, ICLR, and WACV in 2022 and 2023.

Jianwu Fang (www.lotvs.net) received the Ph.D. degree in signal and information processing (SIP) from the University of Chinese Academy of Sciences, China, in 2015. He is currently the Director and an Associate Professor with the Laboratory of Traffic Vision Safety (LOTVS) and the Department of Big Data Management and Application, College of Transportation Engineering, Chang'an University, Xi'an, China. He has been the reviewer and program committee (PC) member for top conferences and journals, including CVPR, ICCV, ACM MM, ICRA, AAAI, IJCAI, IROS, TIP, TITS, TCYB, TIV *et al.*. His research interests include computer vision and pattern recognition and their applications intelligent transportation.

Yunchao Wei (https://weiyc.github.io) received the Ph.D. degree from Beijing Jiaotong University in 2016. He is currently an Assistant Professor with Beijing Jiaotong University. Before joining BJTU, he was an Assistant Professor at the University of Technology Sydney. His main research interests include deep learning and its applications in computer vision, e.g., image classification, video/image object detection/segmentation, and learning with imperfect data. He received the Excellent Doctoral Dissertation Award of CIE in 2016, the ARC Discovery Early Career Researcher Award in 2019, and the First Prize in Science and Technology awarded by China Society of Image and Graphics in 2019.

Tat-seng Chua (https://www.chuatatseng.com/) received the Ph.D. degree from the University of Leeds, U.K. He is the KITHCT Chair Professor with the School of Computing, National University of Singapore, where he was the Acting and Founding Dean of the School from 1998 to 2000. His main research interests include multimedia information retrieval and social media analytics. He is the Co-Director of NExT, a joint center between NUS and Tsinghua University.He is the 2015 winner of the prestigious ACM SIGMM Award for Outstanding Technical Contributions to Multimedia Computing, Communications, and Applications. He is the Chair of Steering Committee of the ACM International Conference on Multimedia Retrieval (ICMR) and Multimedia Modeling (MMM) conference series. He is also the General Co-Chair of ACM Multimedia 2005, ACM CIVR (now ACM ICMR) 2005, ACM SIGIR 2008, and ACM Web Science 2015.

ACKNOWLEDGMENTS

We really appreciate Dinnovate Technology (https://www.dinnovate.cn/) providing us the real satellite data that they collected in urban scenes during different weather.

REFERENCES

[1] Yuanjie Dang, Chong Huang, Peng Chen, Ronghua Liang, Xin Yang, and Kwang-Ting Cheng. 2022. Path-Analysis-Based Reinforcement Learning Algorithm for Imitation Filming. *TMM* (2022).

[2] Didula Dissanayaka, Thumeera R Wanasinghe, Oscar De Silva, Awantha Jayasiri, and George KI Mann. 2023. Review of Navigation Methods for UAV-Based Parcel Delivery. *TASE* (2023).

[3] Tzu-Yi Fan, Fangqi Liu, Jia-Wei Fang, Nalini Venkatasubramanian, and Cheng-Hsin Hsu. 2022. Enhancing situational awareness with adaptive firefighting drones: leveraging diverse media types and classifiers. In *ACM Multimedia*.

[4] Jianwu Fang, Lei-Lei Li, Kuan Yang, Zhedong Zheng, Jianru Xue, and Tat-Seng Chua. 2022. Cognitive Accident Prediction in Driving Scenes: A Multimodality Benchmark. *arXiv:2212.09381* (2022).

[5] Payton Goodrich, Omar Betancourt, Ana Claudia Arias, and Tarek Zohdi. 2023. Placement and drone flight path mapping of agricultural soil sensors using machine learning. *Computers and Electronics in Agriculture* (2023).

[6] Pu Jin, Lichao Mou, Gui-Song Xia, and Xiao Xiang Zhu. 2022. Anomaly Detection in Aerial Videos With Transformers. *TGRS* (2022).

[7] Tianjiao Li, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, and Zhiheng Li. 2021. Uav-human: A large benchmark for human behavior understanding with unmanned aerial vehicles. In *CVPR*.

[8] Jinliang Lin, Zhedong Zheng, Zhun Zhong, Zhiming Luo, Shaozi Li, Yi Yang, and Nicu Sebe. 2022. Joint Representation Learning and Keypoint Detection for Cross-view Geo-localization. *TIP* (2022).

[9] Liu Liu and Hongdong Li. 2019. Lending orientation to neural networks for cross-view geo-localization. In *CVPR*.

[10] Zhihao Liu, Yuanyuan Shang, Timing Li, Guanlin Chen, Yu Wang, Qinghua Hu, and Pengfei Zhu. 2023. Robust Multi-Drone Multi-Target Tracking to Resolve Target Occlusion: A Benchmark. *TMM* (2023).

[11] Ioannis Pitas and Ioannis Mademlis. 2022. Autonomous UAV Cinematography. In *ACM Multimedia*.

[12] Redmond Ramin Shamshiri, Ibrahim A Hameed, Siva K Balasundram, Desa Ahmad, Cornelia Weltzien, and Muhammad Yamin. 2018. Fundamental research on unmanned aerial vehicles to support precision agriculture in oil palm plantations. *Agricultural Robots-Fundamentals and Application* (2018).

[13] Yujiao Shi, Dylan John Campbell, Xin Yu, and Hongdong Li. 2022. Geometry-guided street-view panorama synthesis from satellite imagery. *TPAMI* (2022).

[14] Yujiao Shi and Hongdong Li. 2022. Beyond Cross-view Image Retrieval: Highly Accurate Vehicle Localization Using Satellite Image. In *CVPR*.

[15] Yujiao Shi, Liu Liu, Xin Yu, and Hongdong Li. 2019. Spatial-Aware Feature Aggregation for Image based Cross-View Geo-Localization. In *NeurIPS*.

[16] Yujiao Shi, Xin Yu, Dylan Campbell, and Hongdong Li. 2020. Where am I looking at? joint location and orientation estimation by cross-view matching. In *CVPR*.

[17] Yujiao Shi, Xin Yu, Liu Liu, Dylan Campbell, Piotr Koniusz, and Hongdong Li. 2022. Accurate 3-DoF Camera Geo-Localization via Ground-to-Satellite Image Matching. *TPAMI* (2022).

[18] Yujiao Shi, Xin Yu, Liu Liu, Tong Zhang, and Hongdong Li. 2020. Optimal feature transport for cross-view image geo-localization. In *AAAI*.

[19] Francesco Betti Sorbelli, Federico Corò, Lorenzo Palazzetti, Cristina M Pinotti, and Giulio Rigoni. 2023. How the Wind Can Be Leveraged for Saving Energy in a Truck-Drone Delivery System. *TITS* (2023).

[20] Tingyu Wang, Zhedong Zheng, Yaoqi Sun, Tat-Seng Chua, Yi Yang, and Chenggang Yan. 2022. Multiple-environment Self-adaptive Network for Aerial-view Geo-localization. *arXiv:2204.08381* (2022).

[21] Tingyu Wang, Zhedong Zheng, Chenggang Yan, Jiyong Zhang, Yaoqi Sun, Bolun Zheng, and Yi Yang. 2021. Each part matters: Local patterns facilitate cross-view geo-localization. *TCSVT* (2021).

[22] Tingyu Wang, Zhedong Zheng, Zunjie Zhu, Yuhan Gao, Yi Yang, and Chenggang Yan. 2022. Learning Cross-view Geo-localization Embeddings via Dynamic Weighted Decorrelation Regularization. *arXiv:2211.05296* (2022).

[23] Longyin Wen, Dawei Du, Pengfei Zhu, Qinghua Hu, Qilong Wang, Liefeng Bo, and Siwei Lyu. 2021. Detection, tracking, and counting meets drones in crowds: A benchmark. In *CVPR*.

[24] Cai YuanQiang, Dawei Du, Libo Zhang, Longyin Wen, Weiqiang Wang, Yanjun Wu, and Siwei Lyu. 2020. Guided attention network for object detection and counting on drones. In *ACM Multimedia*.

[25] Haotian Zhang, Gaoang Wang, Zhichao Lei, and Jenq-Neng Hwang. 2019. Eye in the sky: Drone-based object tracking and 3d localization. In *ACM Multimedia*.

[26] LiangLiang Zhao and MinLing Zhu. 2023. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography. *Drones* (2023).

[27] Ou Zheng, Mohamed Abdel-Aty, Lishengsa Yue, Amr Abdelraouf, Zijin Wang, and Nada Mahmoud. 2022. CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins. *arXiv:2208.11036* (2022).

[28] Zhedong Zheng, Yunchao Wei, and Yi Yang. 2020. University-1652: A multi-view multi-source benchmark for drone-based geo-localization. In *ACM Multimedia*.

[29] Zhedong Zheng and Yi Yang. 2019. Unsupervised scene adaptation with memory regularization in vivo. *IJCAI* (2019).