Scene Understanding in UAVs via Semantic Ontology

Naqqash Dilshad, JaeSeung Song*
Department of Computer and Information Security
Sejong University
Seoul, South Korea
dilshadnaqqash@sju.ac.kr, jssong@sejong.ac.kr

***Abstract*—Recently object tracking and detection backed by Unmanned Aerial Vehicles (UAV) have gained a lot of interest specifically in research fields like security surveillance, traffic monitoring, and search and rescue operations in a natural disaster situation. The key challenges in visual tracking with consumer UAVs appear in the form of continuous change in the appearance of the object of interest, target representation, target object detection, and localization in real-time. Consistent object detection mainly depends on various factors that include image noise, obstructions in the line of sight, variance in lightning conditions, posture changes, and hazy or blurriness in the image that might affect object labeling. To counter this problem, this article proposes an approach in which tracked objects in a scene is entirely described by adding contextual information, i.e., natural features, locality of the scene, and general points of interest. Each situation is described semantically by ontological statements that define the environment.**

***Index Terms*—UAV (Unmanned Aerial Vehicle), Scene Understanding, Object Tracking, Semantic Ontology.**

I. Introduction

RECENT applications of UAV such as aerial video surveillance is developing into a crucial safety-critical platform, like traffic congestion, fire detection, search, and rescue operations, among others [1]–[3], The main reason for consumer UAV popularity is due to its cheap deploying cost in comparison with other air to ground surveillance systems for the purpose of emergency alert generator in crucial situations. Though UAVs must assure a quick response, specifically when there is a critical situation of life and death in an accident, at the same time, UAVs must also possibly fly in extreme weather circumstances, which can be extremely hazardous for a manned aircraft. The major problem with UAVs is that getting a high and abstract level description of the scenes appearing in the video feed by employing only object detection, identification along with other techniques. Empowering a UAV to obtain a complete picture of the scenario while flying and then identifying critical situations by looking deeply and classifying a scene is indeed a helpful capability and could be used in many critical applications.

To mitigate this problem, a model that can perform tracking to handle complex tasks such as object recognition and object detection [4]–[6] is needed. Various research studies aim at easing frequent challenges related to UAV video trackings such as camera resolution, camera trembling, illumination change, and appearance change of the background. Although to attain optimum path tracking [7] and lessen automobile routing problems [8]. Although many tracking algorithms can deal with obstruction, separate objects, shadows, and reflection, object tracking suffers in object labeling [9]. Moreover, camera movements add further difficulties to the object tracking system. Most of the techniques depicted in related work on object tracking are limited by a fixed camera, and, while on the other hand on a moving camera, the traditional background subtraction algorithms are not applicable. For this purpose, generally, methods focus on a particular object class detection task, such as pedestrians or crowds, vehicles, or any other related main applications in this field congregate on one

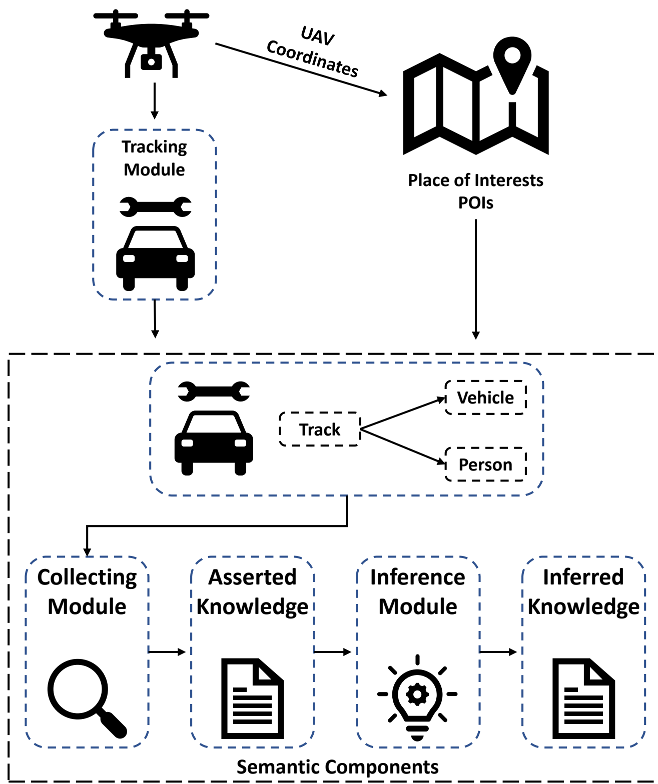

Fig. 1. Logical overview of the framework

particular type of scenario.

As mentioned in the recent studies and observations, our main idea is to enhance the object tracking task in visual data by augmenting the tracking information with contextual information, i.e., complementary data related to the surrounding objects as well as the objects present in the background of the scene, just to develop a complete scene data to lessen tracking problems. Data from tracking techniques and supplementary information in the background are accumulated and coded into ontological reports: the part of the semantic web technologies in the development of tracked objects and their interactions with other spatial objects in the surrounding is critical for object classification and labeling, specifically when camera motion is involved.

II. Related Work

Although high-level abstract knowledge-based scene understanding from a mobile visual sensor perspective is an interesting topic for enhancing the assessment of procedures and assisting video-tracking events such as moving object detection and tracking [4], [5], [10], there are a few related works in literature focusing on this problem. This hinges on the point that the mobile camera will always cause a shortage of important reference points in the video scene which influences both tracking activities and object detection in a bad way, and high-level scene presentation. Subsequently, the majority of the investigations concentrate only on low-level data retrieved from video.

The proposed logical framework presented in Fig. 1 shows an abstract level workflow of the framework, along with all the modules and their main interactions. The key essence of the framework is signified by the semantic modules that, in the figure, are enclosed in a dark dotted square. The main input is a video filmed by an airborne UAV equipped with a visual sensor (top left of Fig. 1). The filmed video feed is taken as an input by the tracking module that separates and extracts the possible tracks of the moving objects in the video. Furthermore, the second input of the framework is environmental data (top right in Fig. 1) it is comprised of certain places labeled as POIs, that are permanent Geo-Referred coordinate points or regions recovered with any reputable map service, lying in the area where the UAV is airborne. The video sequence and the object trajectories, altogether with the POIs retrieved from the map service are forwarded to the semantic mapping module.

This module in succession translates tracking and contextual data in semantic statements according to Track POI ontology, an Ad-hoc constructed ontology to prototype the observed scenarios. The semantic module tracks semantic mapper that maps the moving object in the video scene and frame data along with their minute details about their identity, real shape, pace, and location. The POI semantic mapper aims at defining assertions on the POIs data retrieved with Google Maps query. The collected information on tracks, POIs, and their interrelations is forwarded to a collecting module, which gathers and chooses only the important assertions from the semantic mappers to improve the model and enhance the perspectives in the developing scene. Ultimately, the gathered information is forwarded to the inference module, which removes the irrelevant assertions on the trail, POIs, and relations, to develop a thorough knowledge base, for an in-depth high-level scene understanding.

III. Conclusion

This article presents a novel approach utilizing semantic reasoning and ontology for scene understanding in UAVs. The places and objects of interest in a given scenario are properly described by adding contextual data like features about time, nature, and locality of the scene as well as general points of interest. Each scene is explained by the assessment of ontological reasoning and statements that describes the scene.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00959, Fast Intelligence Analysis HW/SW Engine Exploiting IoT Platform for Boosting On-device AI in 5G Environment). Corresponding author: Prof. Song JaeSeung.

References

[1] R. Shakeri, M. A. Al-Garadi, A. Badawy, A. Mohamed, T. Khattab, A. K. Al-Ali, K. A. Harras, and M. Guizani, "Design challenges of multi-uav systems in cyber-physical applications: A comprehensive survey and future directions," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3340–3385, 2019.

[2] N. Dilshad, J. Hwang, J. Song, and N. Sung, "Applications and challenges in video surveillance via drone: A brief survey," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 728–732.

[3] A. Sikdar, A. Sahu, D. Sen, R. Mahajan, and A. S. Chowdhury, "Unconstrained vision guided uav based safe helicopter landing," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 8328–8335.

[4] H. Kim and J. Ben-Othman, "A collision-free surveillance system using smart uavs in multi domain iot," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2587–2590, 2018.

[5] N. Dilshad, J. Hwang, J. Song, and N. Sung, "A brief survey of uav applications in video surveillance," , pp. 1172–1173, 2020.

[6] Y. Bazi and F. Melgani, "Convolutional svm networks for object detection in uav imagery," *Ieee transactions on geoscience and remote sensing*, vol. 56, no. 6, pp. 3107–3118, 2018.

[7] R. Cui, C. Yang, Y. Li, and S. Sharma, "Adaptive neural network control of auvs with control input nonlinearities using reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 6, pp. 1019–1029, 2017.

[8] K. Dorling, J. Heinrichs, G. G. Messier, and S. Magierowski, "Vehicle routing problems for drone delivery," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 1, pp. 70–85, 2016.

[9] S. Kwak, M. Cho, I. Laptev, J. Ponce, and C. Schmid, "Unsupervised object discovery and tracking in video collections," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3173–3181.

[10] N. H. Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based internet of things services: Comprehensive survey and future perspectives," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 899–922, 2016.