UAV-based Person Re-Identification and Dynamic Image Routing using Wireless Mesh Networking

Neelabhro Roy and Sauranil Debarshi

February 15, 2020

UAV-based Person Re-Identification and Dynamic Image Routing using Wireless Mesh Networking

Neelabhro Roy
Department of ECE
IIIT Delhi
New Delhi, India
neelabhro16171@iiitd.ac.in

Sauranil Debarshi
Department of Aerospace Engineering
IISc Bangalore
Bangalore, India
sauranild@iisc.ac.in

***Abstract*—Person Re-Identification (PRID) has been one of the most challenging tasks in intelligent video surveillance. Most existing PRID based surveillance methods rely on a single camera mounted on emplacements. This approach assumes that any scrutinized person appears again in the field of view of that camera, which is unreliable in real-world settings. In this paper, we propose a system wherein we explore on-board re-identification of persons using cameras mounted on Unmanned Aerial Vehicles (UAVs). The captured images are dynamically shared between multiple UAVs by virtue of wireless mesh networking, also making use of the Robot Operating System (ROS) for on-ground control of the UAVs from a control station, facilitating the exchange of images between them. We demonstrate our approach on an institutional dataset we created, and our experimental results show that the system could be indispensable in aiding airborne surveillance operations.**

***Index Terms*—Person Re-Identification, Cross View Image Retrieval, Unmanned Aerial Vehicles, Mobile Ad Hoc Network, Wireless Mesh Network**

I. Introduction

Cross-View image retrieval has been one of the most important tasks of computer vision, wherein, given an instance as the probe, the objective is to find the most similar instances from a large set of gallery.

Person re-identification [1] is one of the most prominent problems of cross-view image retrieval, and is the task of identifying the same person across camera networks with non-overlapping views. It has become one of the most fundamental subjects in the area of intelligent video surveillance. Applications include security monitoring, pedestrian searching [2], cross-camera tracking [3] etc. For example, in a cross-camera tracking scenario , when a person of interest disappears from one camera view, we have to identify him/her from another view. This matching task is performed with the help of person re-identification. The captured pedestrian images are usually of low quality and resolution. As a result, the Re-id task has to rely upon the appearance information of pedestrians. A person's appearance can vary over different camera angles, owing to large variations in images caused by occlusions, viewpoint, illumination and poses. All of these taken into account together, make person re-identification a challenging task.

With the rapid development of Unmanned Aerial Vehicles (UAVs) in recent years, UAV-based surveillance has become very attractive. The authors in [4] discuss a multi-UAV based aerial surveillance system based on complex urban environments. In [5], the authors present the first mobile re-identification dataset, built using a drone, thereby opening avenues to explore airborne PRID. Arne and Tobias [6] present an approach that uses color and texture image features in aerial video data to re-identify a person of interest. A distributed, communication aided surveillance system using autonomous swarm drones that patrol and intercept detected targets has been presented in [7].

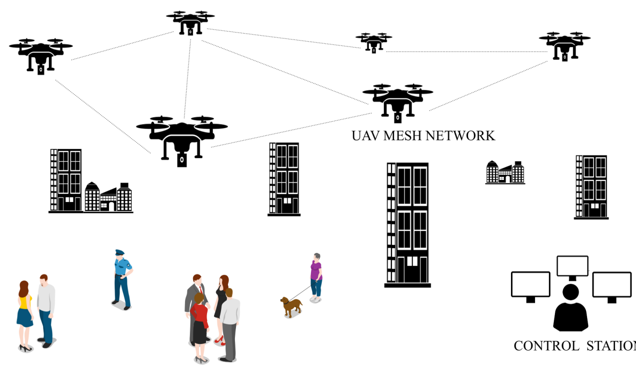

Fig. 1. Wireless mesh network consisting of UAVs and a base station.

In this paper, we present a multi-UAV PRID system. Our prototype system ensures that our implementations are fast and efficient, thereby allowing us to effectively perform person re-identification in a resource-constrained environment. The UAVs create a wireless mesh network to communicate and are controlled by the open-source Robot Operating System (ROS) [8] from a base station. The network, being decentralized, reduces complexity and dynamically routes information from one UAV to another.

II. Network Architecture and System Components

A. Network Architecture

Consider a network consisting of N distributed UAVs, where each UAV has one wireless mesh radio, an Odroid-XU4, and a USB camera. These UAVs constitute a mesh network, with the

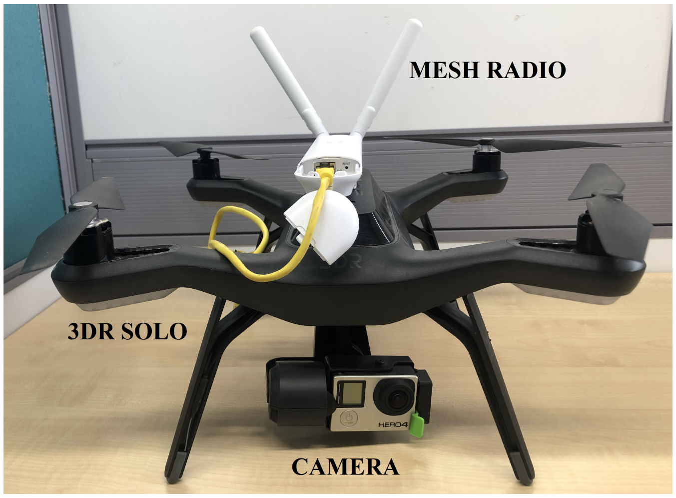

Fig. 2. Quadcopter equipped with a wireless mesh radio and a camera.

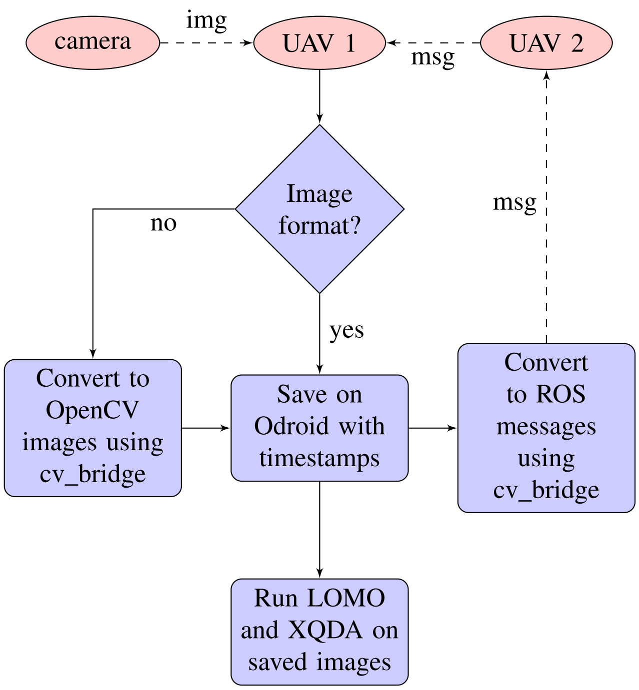

Fig. 3. Sequence of events during image routing from one UAV to another.

base station on the ground acting as the control station. This is illustrated in Fig. 1. The network serves two purposes: First, it functions as a bridge for the base station b_t to monitor and send appropriate commands to the UAVs using the ROS framework. Second, it transmits the camera images from one UAV to another as ROS messages. This ensures synchronization of images between all the UAVs. The UAVs once deployed in the air, can click images in real time, from various positions, thereby incorporating viewpoint variations for subsequent re-identification to take place.

B. System Components

For the purpose of testing the system, we used three 3DR Solo quadcopters, each equipped with a Unifi Mesh UAP-AC-M radio from Ubiquiti Networks. Fig. 2 shows one of the quadcopters. The radios are compact and lightweight, and consist of dual-band omni-directional antennas , making it suitable for our application. All the UAVs have a GoPro Hero 4 camera and an Odroid-XU4 computing device on-board for storing and processing the images.

III. Implementation & Algorithm

A. Image routing

Our algorithms ensure that images are clicked periodically and stored in the on-board Odroid of each UAV with timestamps. Before the transmission process, all images are converted to a ROS compatible format, and then transmitted to the rest of the UAVs as ROS messages. At the receiver end, these messages are converted back to OpenCV based images such that the processing can be done on them. The conversion between messages and images is done by the cv_bridge library of ROS. The complete sequence of events that happen during the routing process is shown in Fig.3. UAV 1 receives data from UAV 2 in the form of ROS messages, which are converted back to images and stored on the Odroid. The images from the on-board camera of UAV 1 are directly saved without any conversion, and then sent to UAV 2.

B. Person Re-identification

For the saved raw images, our feature extraction strategy is based on Local Maximal Occurrence (LOMO) [9]. The LOMO feature analyzes the horizontal occurrence of local features, and maximizes the occurrence to make a stable representation against viewpoint changes. It also handles illumination variations well. For ease of computation, we first apply Principal Component Analysis on the data, to project it onto a 20 dimensional space, from a 26960 dimensional space. We then apply Cross View Semantic Projection Learning (CSPL) [10] on the reduced dimensionality set of features. To devise the most apt algorithm for the dataset, we also apply

Fig. 4. Sample identities from the dataset.

Fig. 5. Experimental setting for building dataset.

the Cross-view Quadratic Discriminant Analysis (XQDA) on the extracted set of features. [9]. Most of the other PRID algorithms use PCA, which is unsupervised, and hence, it may not preserve the discriminative ability between classes. Moreover, learning a metric on PCA features where inter classes already overlap, may not help. On the other hand, XQDA performs both dimensionality reduction and learns the metric simultaneously, leading to the discriminative ability in the native features being preserved. The run-time of XQDA ideally suits a resource constrained environment such as ours.

TABLE I
CMC ACCURACY ON OUR DATASET

S. No.	Rank	XQDA Accuracy	CSPL Accuracy
1	1	83%	50%
2	5	98.5%	90%
3	10	100%	95%
4	15	100%	100%
5	20	100%	100%

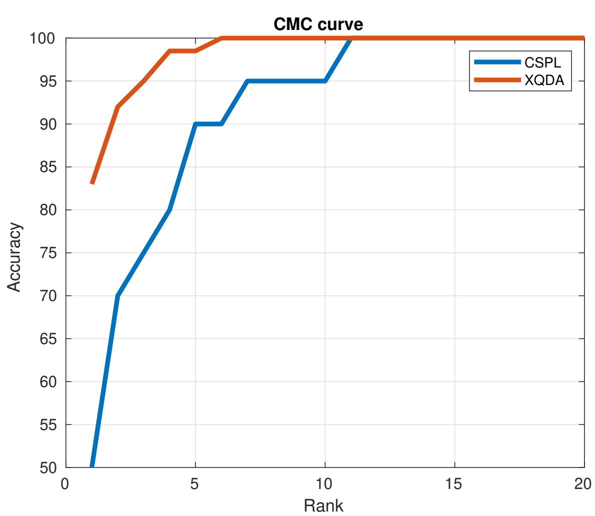

Fig. 6. CMC Curves for XQDA and CSPL.

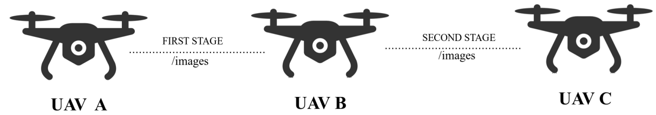

Fig. 7. Illustration of our experimental setup.

IV. EXPERIMENTS AND RESULTS

A. XQDA & CSPL

We built an institutional dataset, to test our prototype system. Fig. 4 shows multiple identities from our dataset, and Fig. 5 shows the experimental setting in which the dataset was captured. We made sure that our dataset covers multiple viewpoint variations. This was built by flying the UAVs at an altitude of 12m, and it covers 40 distinct and discernible identities, which are subsequently divided into probe and gallery. All the images have been normalized to 128x48 pixels for the experiments.

Once the dataset was built, we used LOMO for feature extraction. We use the CMC (Cumulative Match Characteristic) [11] curves for evaluation. We randomly sample half the images for training, and the remaining for testing. This process is repeated for 20 folds, and the resultant average curve is shown in Fig. 6, which shows the results for both CSPL and XQDA. Various accuracies with respect to the ranks for both the algorithms are shown in Table I.

B. Network Throughput

We measured the performance of our multi node network in terms of throughput, when exchange of information from one UAV to another takes place. Three UAVs, each equipped with the aforementioned equipments have been used. Fig. 7 gives an illustration of our setup. All the UAVs are flown equidistant from each other. UAV B acts as the intermediate node between UAV A and UAV C. In the first stage, exchange of information between UAV A and UAV B takes place.

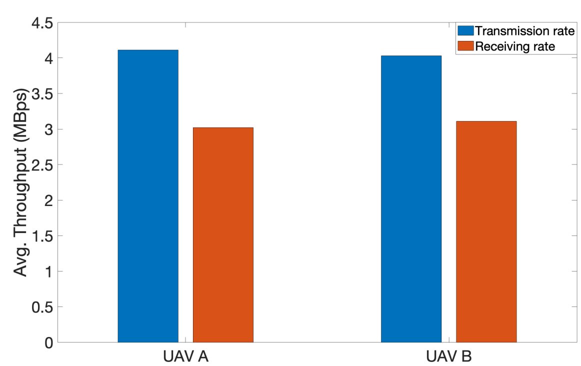

Fig. 8. Average throughput during the first stage.

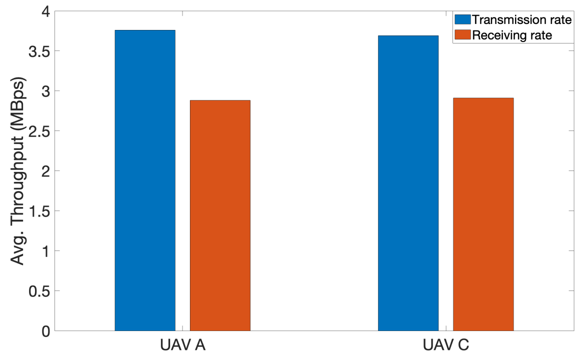

Fig. 9. Average throughput during the second stage.

The second stage comprises of data transfer between UAV A and UAV C, through the intermediate UAV B. The average throughput in each of the stages have been calculated.

C. Throughput in the first stage

Exchange of the respective camera images in the form of ROS messages, takes place in the first stage. The average transmission and reception rates, when a transmission of forty images from UAV A to UAV B, and vice versa, occurs is shown in Fig. 8. Each of them had a transmission and reception rate of approximately 4 MBps and 3 MBps respectively, in the first stage.

D. Throughput in the second stage

We measured the average throughout when the exchange of images from UAV A and UAV C takes place in the second stage. These UAVs are connected to each other through the intermediate UAV B. It has been observed that the average transmission rates reduces marginally. This is because the information shared by UAV A and UAV C has to go through the intermediate stage in between. Fig. 9 shows the transmission rates of UAV A and UAV C during the second stage of transfer. A similar reduction in the reception rates have been observed.

V. Conclusion

We proposed and implemented a multiple UAV based Person Re-Identification system wherein, the UAVs capture images of an identity from multiple angles, taking into account various complexities like angular differences, occlusions, viewpoint and illumination changes etc. The UAVs then communicate with each other by virtue of wireless mesh networking aided by our dynamic image routing algorithms. Our prototype integrates IEEE 802.11 single and multi-hop wireless networks. The ROS middleware provides a robust mechanism to transfer data using WMNs. Moreover, our system can also be used to obtain real-time imagery information using a camera mounted on the UAVs along with the GPS coordinates imprinted on the images to aid person tracking, and transfer those images to other nodes, which are a part of the mesh network, without requiring an active internet connection, and at reasonable transmission rates. We tested our proposed algorithm on an institutional dataset we built, and performed detailed experiments using various PRID algorithms. Our results prove that the usage of XQDA can be of significance, given the robust nature of the scenario in question. From the throughput measurements calculated for single-hop and multi-hop networks, it has been observed that by using aerial nodes like UAVs, the transfer rate of our algorithm significantly improves. This system could be indispensable in real time cross camera tracking.

References

[1] "S. gong, m. cristani, s. yan, c.c. loy (eds.), person re-identication, advances in computer vision and pattern recognition, springer, 2014," https://www.springer.com/in/book/9781447162957.

[2] Chen Change Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1988–1995.

[3] Xiaogang Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 3 – 19, 2013, Extracting Semantics from Multi-Spectrum Video.

[4] E. Semsch, M. Jakob, D. Pavlicek, and M. Pechoucek, "Autonomous uav surveillance in complex urban environments," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Sep. 2009, vol. 2, pp. 82–85.

[5] Ryan Layne, Timothy M. Hospedales, and Shaogang Gong, "Investigating open-world person re-identification using a drone," in *Computer Vision - ECCV 2014 Workshops*, Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, Eds., Cham, 2015, pp. 225–240, Springer International Publishing.

[6] Arne Schumann and Tobias Schuchert, "Person re-identification in UAV videos using relevance feedback," in *Video Surveillance and Transportation Imaging Applications 2015*, Robert P. Loce and Eli Saber, Eds. International Society for Optics and Photonics, 2015, vol. 9407, pp. 305 – 312, SPIE.

[7] M. Gótzy, D. Hetényi, and L. Blázovics, "Aerial surveillance system with cognitive swarm drones," in *2016 Cybernetics Informatics (K I)*, Feb 2016, pp. 1–6.

[8] Morgan Quigley, Ken Conley, Brian P. Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng, "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[9] S. Liao, Y. Hu, , and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 2197–2206.

[10] Ju Dai, Ying Zhang, Huchuan Lu, and Hongyu Wang, "Cross-view semantic projection learning for person re-identification," *Pattern Recogn.*, vol. 75, no. C, pp. 63–76, Mar. 2018.

[11] Doug Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *In IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, Rio de Janeiro*, 2007.