Demo: Tagging IoT Data in a Drone View

Lan-Da Van, Chun-Hao Chang, Kit-Lun Tong, Kun-Ru Wu, Ling-Yan Zhang*, Yu-Chee Tseng
Department of Computer Science, National Chiao Tung University
Hsinchu, Taiwan
lyzhang@nctu.edu.tw

ABSTRACT
Both cameras and IoT devices have their particular capabilities in tracking moving objects. Their correlations are, however, unclear. In this work, we consider using a drone to track ground objects. We demonstrate how to retrieve IoT data from devices, which are attached on human objects, and correctly tag them on the human objects captured by a drone view. This is the first work correlating IoT data and computer vision from a drone camera. Potential applications of this work include aerial surveillance, people tracking, and intelligent human-drone interaction.

CCS CONCEPTS
• **Security and privacy → Usability in security and privacy**; • **Computing methodologies → Vision for robotics**; **Object detection**; **Object recognition**; **Object identification**; **Feature selection**; **Activity recognition and understanding**; • **Hardware** → *Sensor applications and deployments*; • **Computer systems organization** → *Neural networks*.

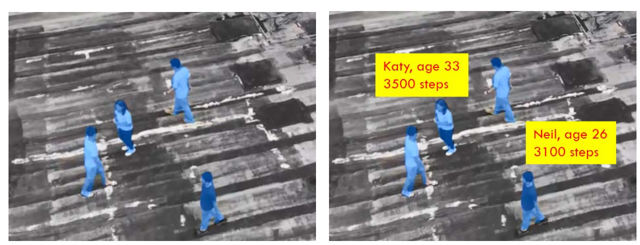
KEYWORDS
drone, IoT, computer vision, people identification, data fusion

ACM Reference Format:
Lan-Da Van, Chun-Hao Chang, Kit-Lun Tong, Kun-Ru Wu, Ling-Yan Zhang*, Yu-Chee Tseng. 2019. Demo: Tagging IoT Data in a Drone View. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19), October 21–25, 2019, Los Cabos, Mexico.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3300061.3343378

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
MobiCom '19, October 21–25, 2019, Los Cabos, Mexico
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6169-9/19/10.
https://doi.org/10.1145/3300061.3343378

1 INTRODUCTION
Drones (known as Unmanned Aerial Vehicles, UAVs) have been applied to a wide range of security and surveillance applications. When conducting video surveillance, one fundamental issue is person identification (PID), where human objects in videos need to be immediately tagged by their IDs and personal profiles. Traditional technologies such as RFID and fingerprint/iris/face recognition have their limitations or require close contact to specific devices. Hence, they are not applicable to drones with changing height and view angle. More challenges of identifying people from a drone view are discussed and presented in [2].

In this work, we present an approach to detecting human objects in the videos taken from a drone and correctly tagging their personal profiles retrieved, through wireless communications, from their wearable devices on those human objects. Fig. 1 shows what our system can achieve in a drone view as compared to existing systems. Through combining IoT and AI, we can display IoT data directly on recognized image objects (Fig. 1(b)).



(a) (b)

Figure 1: (a) State-of-the-art instance segmentation of a drone image; (b) Our conceptual result after tagging IoT data.

Our target application is future aerial surveillance with much richer capability. For example, in Fig. 1, the displayed information can include the past activities of users even before they entered the camera view. To achieve this goal, we propose a data fusion approach that combines inertial sensors with videos. As shown in Fig. 2, our system integrates an aerial camera and users' wearable devices. There is a drone equipped with a camera to continuously record videos. A

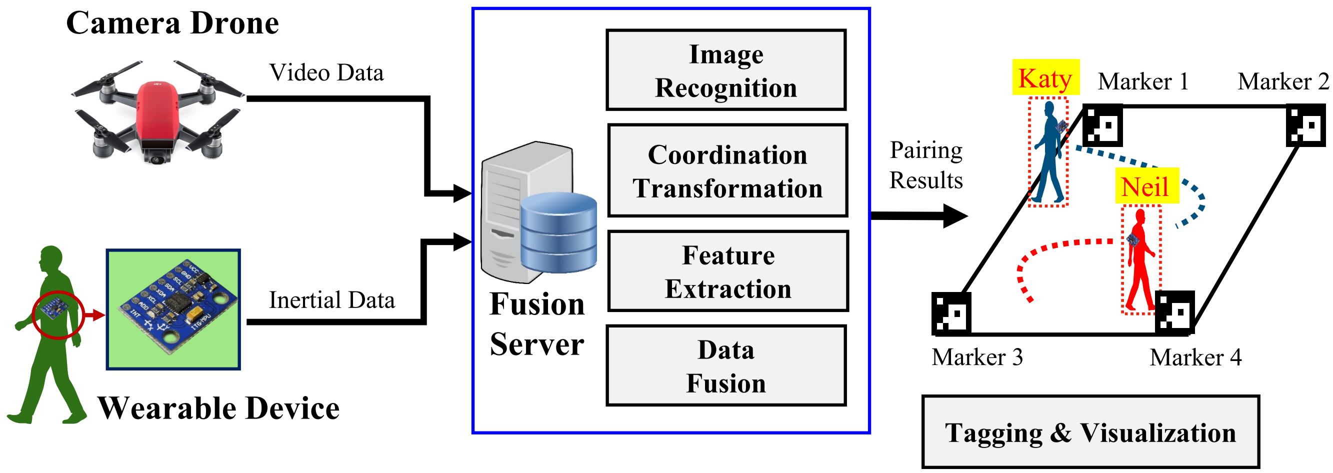


Figure 2: System architecture.

crowd of people are in the drone view and some of them may put on their wearable devices. We assume that each device has a communication interface and some inertial sensors. This allows us to transmit user profiles and motion data. These video data and inertial data are transmitted to a fusion server for correlation analysis.

In order to correlate video data and inertial data, we conduct four major procedures. Firstly, we retrieve human objects by using a deep learning network. Secondly, we use the ArUco algorithm to transform the human objects in a pixel space to a real-world coordinate to handle the change of drone positions. Note that we place four ArUco markers on the ground for the transformation. Thirdly, we extract human motion features from both video data and inertial sensor data. Fourthly, we design a fusion algorithm to measure the similarity of any pair of normalized motion feature series, one from a human object in videos and the other from a wearable device. By quantifying all-pair similarity scores, we are able to couple human objects with their IoT data, achieving our goal of tagging wearable device data on human objects in drone-recorded videos.

2 IMPLEMENTATION AND DEMONSTRATION

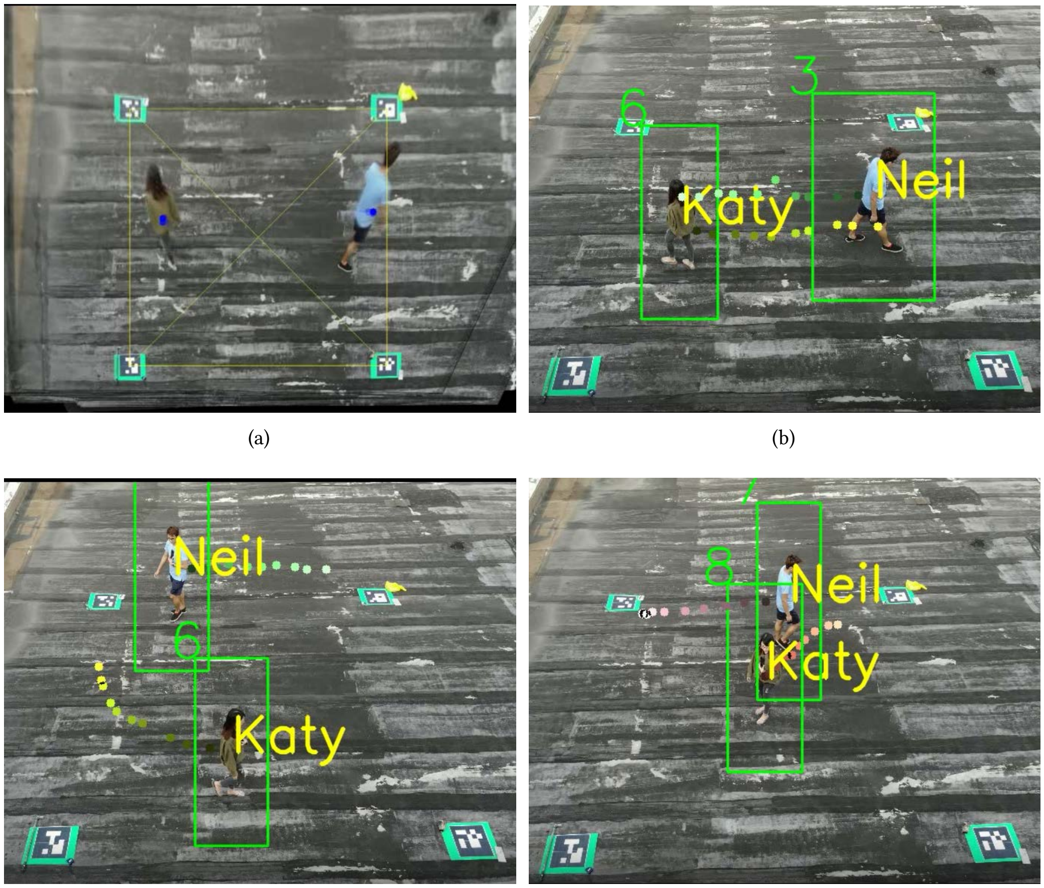
We develop a prototype of our system consisting of the following hardware components.

- **Drone.** We use a DJI Spark, which has a camera with a 1/2.3" CMOS sensor of 11.8 megapixel resolution and a viewing angle of 81.9°.
- **Wearable device.** We use Android smartphones (HTC 10), which have built-in accelerometer, gyro, and geomagnetic field sensors, to simulate wearable devices.
- **Fusion server.** We use a server equipped with an Intel Core i7-8700HQ processor, 32GB RAM, and a GeForce GTX1080 Ti GPU.

The following APIs and tools are used for data sensing, people detection, and coordination transformation.

- **Sensor type TYPE_ORIENTATION** supported by Android is used for deriving inclination matrix and rotation matrix from gravity and the geomagnetic field sensors.
- **YOLO (You Only Look Once) v3 and SORT (Simple Online and Realtime Tracking)** are used to detect and track human objects from videos.
- **ArUco markers and transformation matrices** are used to transform the a human object's location from a pixel space to a ground space.

To validate our idea, we conduct a number of experiments. The experiment scene is shown in Fig. 3. We place four ArUco markers on a square ground. There are two users (named *Neil* and *Katy*) and we hang smartphones on their chests. The two users walk in a variety of walking paths (straight line, snake-like, circular, and random). Their walking paths do interleave with each other in many cases, causing occlusion in videos from time to time. However, since our data fusion model also considers long-term inertial data, it is still possible to get correct pairing.



(a) (b) (c) (d)

Figure 3: Experimental scene and results: (a) The drone view after coordination transformation; (b) The pairing result of straight-line walk; (c) The pairing result of circular walk; (d) The pairing result of interleaving walk.

We evaluate our results by *multiple object tracking accuracy* (MOTA):

$$MOTA = \frac{\sum_{\forall t} \text{correct identifications in } t}{\sum_{\forall t} \text{all identifications in } t}. \quad (1)$$

Our experimental results show an average MOTA of 90.06%.

To conclude, this work presents a novel idea for fusing visual data from a drone view and inertial data from wearable devices. Some sample demo videos are available in [4]. The contributions of our work are three-fold: (i) it tackles the problem of changing height and view angle of drones, (ii) it requires no cumbersome process of data labeling, and (iii) it does not rely on finding clear human biological features. These features make it possible to correctly identify people who walk in a crowd, even under the circumstances of severe occlusion or appearance change. The detailed designs of data fusion algorithms are available in [1, 3] (but which are designed for ground cameras).

ACKNOWLEDGMENTS

Y.-C. Tseng's research is co-sponsored by Pervasive Artificial Intelligence Research (PAIR) Labs, 105-2218-E-009-003 (MoST), and "Center for Open Intelligent Connectivity" (MoE).

REFERENCES

[1] Wei-Chun Chang, Cheng-Wei Wu, Richard Tsai, Kate Lin, and Yu-Chee Tseng. 2018. Eye on You: Fusing Gesture Data from Depth Camera and Inertial Sensors for Person Identification. In *IEEE International Conference on Robotics and Automation (ICRA)*. 2021–2026.

[2] Matthew Fysh and Markus Bindemann. 2018. Person Identification from Drones by Humans: Insights from Cognitive Psychology. *Drones* 2, 4 (2018), 32.

[3] Richard Tsai, Ting-Yuan Ke, Kate Lin, and Yu-Chee Tseng. 2019. Enabling Identification Aware Tracking via Fusion of Visual and Inertial Features. In *IEEE International Conference on Robotics and Automation (ICRA)*.

[4] Lan-Da Van, Chun-Hao Chang, Kit-Lun Tong, Kun-Ru Wu, Ling-Yan Zhang, and Yu-Chee Tseng. 2019. Demo videos of paper "Demo: Tagging IoT Data in a Drone View". Retrieved August 2, 2019 from https://youtu.be/X4Hqn7ndQd4