Communications in Transportation Research 1 (2021) 100002

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Editorial

Integrated computer vision algorithms and drone scheduling

1. Introduction

Computer vision algorithms have attained significant accuracy in the past decade, among which arguably the most important one is deep neural networks. Unmanned aerial vehicles, commonly called drones, equipped with cameras, offer a convenient, efficient, and cost-effective way of collecting a large set of images. Combining drones and computer vision algorithms can automate the monitoring and surveying of infrastructure systems, for example, car detection (Maria et al., 2016), pedestrian and bicycle volume data collection (Kim, 2020), and road degradation survey (Leonardi et al., 2018). However, the existing research has been largely driven by two independent streams of expertise: computer vision and drone scheduling. Computer scientists strive to design more accurate computer vision algorithms without much consideration of how the images are collected, whereas operations researchers endeavor to design drone routing algorithms to collect a *given* set of images in the most efficient manner. We suggest that the planning of images to collect (number and locations of images, amongst others) and the design of—more often than not, the choice of—computer vision algorithms should be determined holistically instead of independently. Section 2 presents an example to show the number of images to collect depends on the accuracy of the computer vision algorithms. Section 3 lays out the roadmap for future research direction.

2. Illustrative example

In the example, we consider the construction of a wall of 300,000 (1250*240) black and white mosaics. According to the design, 102,229 mosaics are black and form the word "COMMTR," the acronym of *Communications in Transportation Research*, as shown in Fig. 1(a). However, in practice, construction workers may misplace some mosaics (use a black one when a white one should be used, or vice versa) due to carelessness; for instance, Fig. 1(b) shows a wall with 100 mosaics misplaced. Drones can be used to take photos of different parts of the wall and image classification algorithms (a basic type of computer vision algorithm that predicts the category of an image from a given set of categories) can be used to predict whether an image contains misplaced mosaics, and if yes, corrective actions are taken.

All our experiments are based on simulated data. Each mosaic has 10*10 pixels, and hence the wall has a total of 12,500*2400 pixels. The training data is obtained as follows. (i) 100 walls are generated, each containing 100 randomly misplaced mosaics (i.e., 0.03 % of the mosaics are misplaced). (ii) 100 images of 200*200 pixels are randomly generated from the image of each wall. If an image contains at least one pixel from a misplaced mosaic, the image is labeled "misplaced" (see Fig. 1(c)). Otherwise the image is labeled "conforming" (see Fig. 1(d)). Since most images are conforming but we are more concerned with misplaced images, we randomly delete 85 % of the conforming images to obtain a balanced training dataset. Eventually, we have 1344 conforming images and 1320 misplaced images. We train a deep neural network model using transfer learning from a pre-trained model MobileNetV2 (Sandler et al., 2018). The test set is obtained as follows. (i) 100 walls are generated, each containing 100 randomly misplaced mosaics. (ii) n photos of 200*200 pixels are randomly taken from the image of each wall, $n = 10$, 20, 50, 100, 200, 1000. Suppose that if a photo in the test set is predicted to be in the class "misplaced," it will be manually checked and a misplaced mosaic will be identified if it has at least one pixel in the photo (more than one misplaced mosaic can be identified by checking one photo). The classification accuracy, i.e., the total number of images correctly predicted divided by the total number of images in the test set, is 94 %. The results averaged over the 100 walls[1] in the test set are reported below.

n	Average number of images predicted "misplaced" per wall (these images can be conforming or misplaced)	Average number of images predicted "misplaced" and are actually misplaced	Average number of misplaced mosaics identified in the misplaced images that are correctly predicted
10	1.49	1.15	1.24
20	3.14	2.38	2.54
50	7.65	5.95	6.12
100	15.49	12.16	12.10
200	30.29	23.69	22.31
1000	145.50	115.30	69.50

The first column n is the number of photos taken, which affects the cost of using drones to collect the images and the cost of calling the deep neural network model to classify the images. The first cost is the sum of long-haul cost (travel cost between the drone's depot and the wall) and detour cost approximately proportional to $\sqrt{n}$ (Daganzo, 2005) and the second cost is generally negligible. The second column is proportional to n and reflects the labor cost of manually checking the photos. The third column is the number of images that we are interested in, but it should not be treated as the benefit of the drone-based image classification system, because many images have overlaps. In fact, the last column is the benefit of the system. Note that when n is small, e.g., when, $n \leq 200$[2]

[1] For $n = 1000$, we use 10 walls in the test set for computational convenience.
[2] If no two photos are overlapping, 200 photos will cover 27 % of the area of the wall.

https://doi.org/10.1016/j.commtr.2021.100002
Received 4 July 2021; Received in revised form 23 July 2021; Accepted 23 July 2021
Available online 11 August 2021

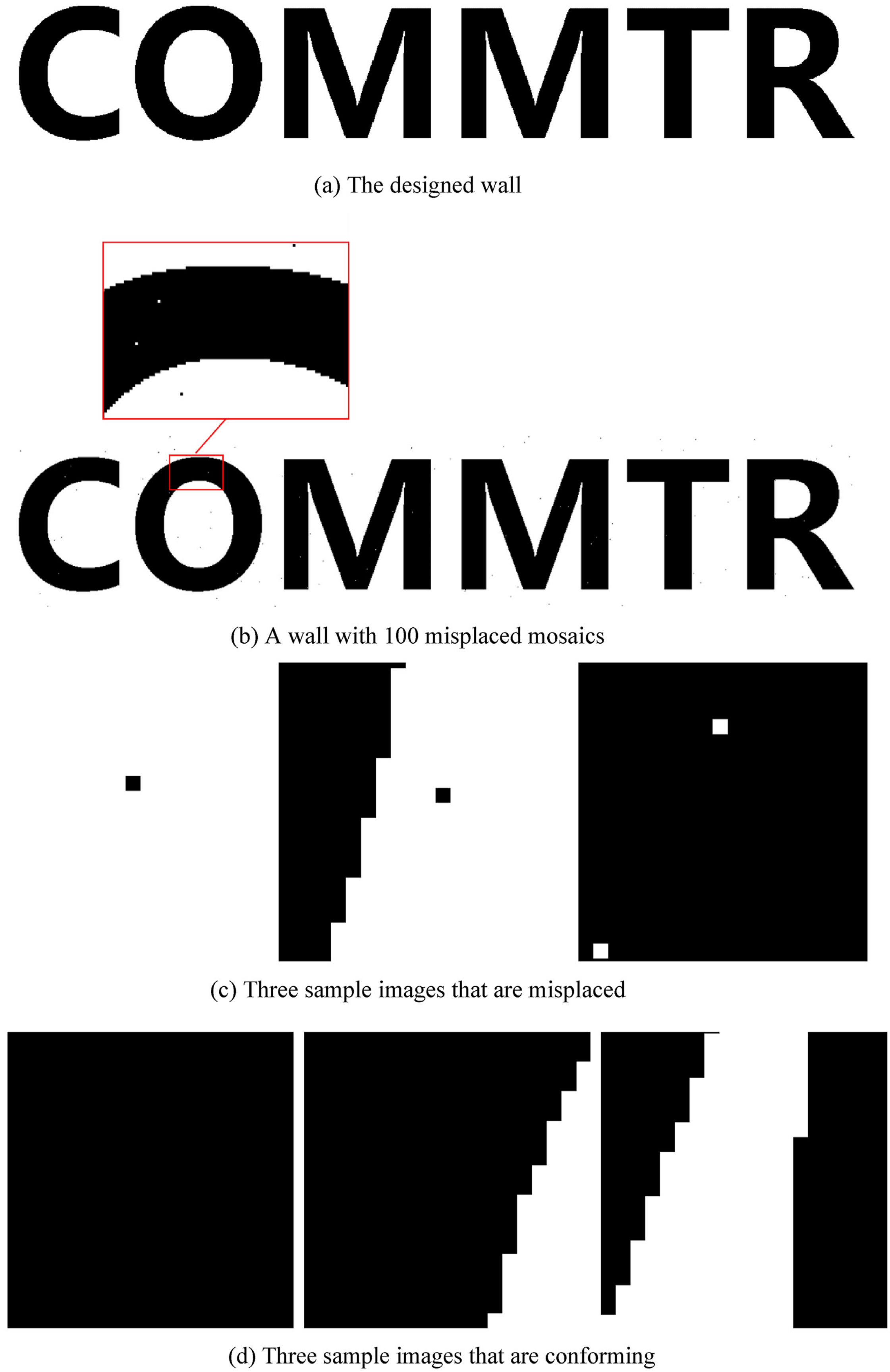

Fig. 1. The designed wall, an actual wall, and photos taken by drone.

the number of misplaced mosaics identified is approximately proportional to n. When n is large, e.g., when $n = 1000$, many images overlap and hence the increase in the number of misplaced mosaics identified slows down with the increase in n. To formalize the above findings, the profit of using the drone-based image classification system is approximately $f(n,\gamma) - (a_0 + a_1\sqrt{n}) - b(\gamma)n$, where γ is the accuracy of the image classification algorithm, $f(n,\gamma)$ is a function reflecting the benefit of identifying misplaced mosaics and is increasing concave in n and increasing in γ, a_0 is the long-haul cost, a_1 is a coefficient related to detour cost, and $b(\gamma)$ is a function reflecting labor cost of manual image checking and is decreasing in.$\gamma\rho b(\gamma) = \gamma\rho + (1-\gamma)(1-\rho)\gamma\rho \ll 0.5$[3] This equation clearly shows that the optimal number of images to collect that maximizes the profit is dependent on the accuracy of the image classification model γ.

3. Conclusions and future research directions

In summary, the planning of images to collect and the choice of computer vision algorithms should be determined jointly. The illustrative example already demonstrates that the number of images to collect depends on the accuracy of the computer vision algorithm. The distance of the drone to the object whose photo is to be taken, the direction of the drone to the surface of the object, and time of photo taking (which is related to the light on the object surface) all affect the accuracy and the choice of the computer vision algorithm, which then affects the optimal number of images to collect. Even if the photos are taken from a fixed-position camera, the direction and focal distance of the camera may still be adjustable. Integrated planning of images to collect, drone routing, and the choice of computer vision algorithms is a worthwhile future research topic. For instance, to monitor road traffic, the collected road traffic information in the past by drones and other sources (e.g., bus and taxi GPS data) can be used to predict the future traffic, and drones can be routed to collect images on roads whose traffic is highly uncertain or unstable; moreover, the visibility (weather and time of day) and the altitude at which the drone flies affect the number of roads whose traffic can be monitored simultaneously and the resolution of each road, which then affect the performance of the computer vision algorithms. Therefore, the height of the drone should be optimized jointly with routing decisions. Hopefully, more transportation researchers will embark on this fruitful area.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The work described in this paper was substantially supported by the funding for Projects of Strategic Importance of The Hong Kong Polytechnic University (Project Code: 1-ZE2D).

References

Daganzo, C.F., 2005. Logistics Systems Analysis. Springer, Heidelberg.
Kim, D., 2020. Pedestrian and bicycle volume data collection using drone technology. J. Urban Technol. 27 (2), 45–60.
Leonardi, G., Barrile, V., Palamara, R., Suraci, F., Candela, G., 2018. Road degradation survey through images by drone. In: International Symposium on New Metropolitan Perspectives. Springer, Cham, pp. 222–228.
Maria, G., Baccaglini, E., Brevi, D., Gavelli, M., Scopigno, R., 2016. A drone-based image processing system for car detection in a smart transport infrastructure. In: 2016 18th Mediterranean Electrotechnical Conference (MELECON). IEEE, pp. 1–5.
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 4510–4520.

Wen Yi is a Senior Lecturer at Massey University, New Zealand. Her research interests include construction waste transportation, optimization in construction management, digital technology in construction, and prefab product logistics. Her papers are published in *Computer-Aided Civil and Infrastructure Engineering*, *Transportation Research Part A*, *Journal of Transportation Engineering*, and *Advanced Engineering Informatics*.

Shuaian (Hans) Wang is a Professor at The Hong Kong Polytechnic University (PolyU). His research interests include shipping operations management, green shipping, big data in shipping, port planning and operations, urban transport network modeling, and logistics and supply chain management. He dedicates to rethinking and proposing innovative solutions to improve the efficiency of maritime and urban transportation systems, to promote environmental friendly and sustainable practices, and to transform business and engineering education.

Yong (Jimmy) Jin is an Assistant Professor and Assistant Dean in the Faculty of Business, at The Hong Kong Polytechnic University. He received PhD in Business Administration from the University of Florida. His current research interests include interdisciplinary research in information systems, technology management and finance. His work has appeared in leading journals such as *Production and Operations Management*, *Journal of the Association for Information Systems*, *Journal of Management Information Systems*, *Decision Support Systems*, *International Journal of Production Economics*, *Risk*, and others.

Professor Jiannong Cao is currently the Otto Poon Charitable Foundation Professor in Data Science and the Chair Professor of Distributed and Mobile Computing in the Department of Computing at The Hong Kong Polytechnic University (PolyU). He is also the Dean of Graduate School, the director of Research Institute for Artificial Intelligence of Things (RIAIoT) in PolyU, the director of the Internet and Mobile Computing Lab (IMCL). He was the founding director and now the associate director of PolyU's University's Research Facility in Big Data Analytics (UBDA). He served the department head from 2011 to 2017. Prof. Cao is a member of Academia Europaea.

Wen Yi
School of Built Environment, College of Sciences, Massey University, 0632, Auckland, New Zealand
E-mail address: yiwen96@163.com.

Hans Wang*
Faculty of Business, The Hong Kong Polytechnic University, Hong Kong, China

[3] Suppose the actual proportion of images that are misplaced is ρ. Then, $b(\gamma) = \gamma\rho + (1-\gamma)(1-\rho)$, which is decreasing in γ because in reality $\rho \ll 0.5$.

Yong Jin
School of Accounting and Finance, The Hong Kong Polytechnic University, Hong Kong, China
E-mail address: jimmy.jin@polyu.edu.hk.

Jiannong Cao
Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
E-mail address: csjcao@comp.polyu.edu.hk.

* Corresponding author.
E-mail address: wangshuaian@gmail.com (H. Wang).