

Analyzing the usefulness of public web camera video sequences for calibrating and validating pedestrian dynamics models

Dariusz Pałka^{1[0000–0003–1420–981X]}, Robert Lubaś^{11[0000–0002–7261–5042]}, and
Jarosław Was^{1[0000–0003–2964–745X]}

AGH University of Science and Technology,
Faculty of Electrical Engineering, Automatics, IT and Biomedical Engineering
Mickiewicza 30, 30-059 Krakow, Poland
`jaroslaw.was@agh.edu.pl`

Abstract. Calibrating and validating pedestrian dynamics models is usually conducted using data obtained in experiments with groups of people. An interesting alternative to using data obtained in such a way is using data from public web cameras. The article presents a case study of using public web cameras in the analysis of pedestrian dynamics and social behavior. We have applied YOLOv3 object detector trained on the COCO dataset in order to identify objects in the video frame and to determine their position.

Keywords: Validating pedestrian dynamics models · YOLO · Public web cameras.

1 Introduction

In the process of creating pedestrian dynamics models (or their adaptation to new situations), the determination of parameters for these models and their validation are very important elements. For this to be possible, it is necessary to perform measurements in real pedestrian behavior situations [6]. Typically, in order to obtain the results of such measurements, researchers plan and conduct experiments with groups of pedestrians. An alternative method, presented in this paper, is to use data from publicly available web cameras (webcams).

This paper presents examples of extracting characteristics for a pedestrian dynamics model obtained thanks to a system we implemented that processes data from publicly available web cameras.

In times of the pandemic, methods of assessing social distance based on camera images have gained a lot of importance. In most cases when the dynamics of pedestrians is monitored using the image from cameras, we do not have additional tools (e.g. sensors indicating additional positioning) that would enable a more precise assessment of the configuration of people within the range of a camera. One can point out different approaches to crowd analysis using video cameras. Tran et al. [12] proposed a graph-based framework of grouped pedestrians based on Social Distances Model (SDM) [14] - a Cellular Automata (CA)

based model of crowd based on proxemics rules, where social distances around a pedestrian are applied. CA based methods of crowd dynamics can use data-driven paradigm and adaptable lattice [1]. One can also point out continuous Social Force [4] base crowd dynamics simulations [13]. Another trend is based on using bio inspired methods in crowd analysis [2], while the task of crowd counting and density maps estimating from videos [10] is often carried out using convolutional neural networks [5], [9].

2 Application of web cameras

Advantages of using publicly available web cameras in comparison with conducting pedestrian experiments include: no costs needed for organizing the experiment and using the equipment, access to video data located in different locations, which gives the possibility to acquire data e.g. from different cultures and in different social situations (e.g. pandemic condition).

This approach also has significant drawbacks in the area of both planning and conducting experiments. The most important of these stem from the inability to stage a specific situation. This results in, among other things, an inability to measure parameters for situations that rarely occur in reality (such as evacuation associated with the appearance of a threat) and an inability to repeat measurements for a given situation. In addition, it is not possible to set the position and area of registration of the camera, and the measurement is limited to video data (it is not possible to obtain data from other sensors and to use the sensor fusion technique).

2.1 Object identification

YOLOv3 [7,8] object detector trained on the *COCO* dataset was used to identify objects in the video frame and determine their position. An example of the results of recognition of different types of objects by YOLO for a webcam showing a view of Grodzka Street in Krakow are presented in Fig. 1. As can be seen, most pedestrians were correctly recognized even when they were far from the camera.

2.2 Mapping from 2D camera space to 3D world space

Some characteristics related to pedestrian dynamics can be determined without knowing metric relations in 3D world space, directly from the 2D image from the camera; they include, for example:

- a number of pedestrians in the space observed by the camera
- space utilization (determines how often a location is occupied by a pedestrian)
- the location of Points of Interest (POI)



Fig. 1. The upper picture shows the view from the webcam on Grodzka Street in Krakow. The bottom picture shows the same image with the objects recognized by YOLO, where: p - person (orange frame), b - bicycle (green frame), d - dog (yellow frame), h - handbag (magenta frame), u - umbrella (brown frame), c - car (brown frame).

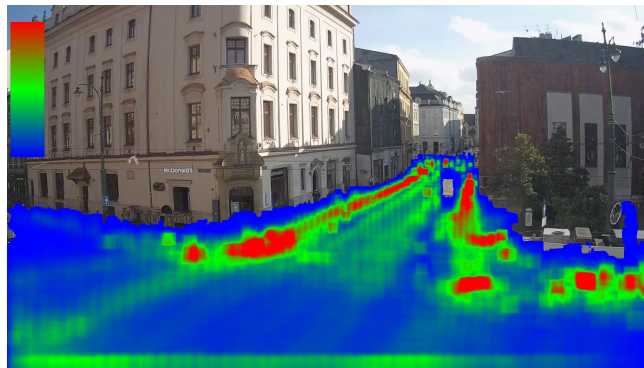


Fig. 2. Grodzka Street relative space utilization. The upper left part of the image: the palette used to indicate the degree of space utilization: red color - the highest space utilization, blue color - the lowest.

Fig. 2 shows an example of calculated space utilization for the visible portion of Grodzka Street. Points of Interest can be determined on the basis of space utilization distribution. They correspond to the areas with the highest values of space utilization (i.e. places where people stay the most often and for the longest time).

There are, however, many characteristics that require the determination of metric spatial relationships, such as:

- density of people in a given area (number of people / area)
- absolute speed of movement of individuals
- distribution of absolute and relative pedestrian speeds
- minimum distances between pedestrians

In order to determine spatial relationships between objects (e.g. people), it is necessary to define the mapping from 2D camera space to 3D world space. The mapping is possible thanks to the camera calibration process, i.e. the projection matrix from 3D points (in world space) to 2D points (in camera space). Camera calibration methods can be classified into two main categories [11]:

- methods based on known calibration objects
- methods that do not depend on prior knowledge of camera scenes, so called ‘camera self-calibration’

In the context of using public webcams to determine parameters of pedestrian dynamics models, a particularly useful method is camera self-calibration based on the video of walking persons [3,11].

In the case of the camera showing the view of the AGH Main Street, a procedure was carried out to determine the mapping from 2D camera space to 3D world space. The procedure was performed based on metric measurements for characteristic elements present in the image, such as street width, the height and distance between the street signs, the size of the pedestrian crossings, the sizes of the benches, etc.

The method for determining minimum distances between pedestrians is shown in Fig. 3 and Fig. 4. Fig. 3 shows two pedestrians passing each other on the sidewalk. The successive pedestrian’s positions are presented with an interval of 4 seconds (every 100 frames; the frame rate for this camera is 25 frames per second).

Fig. 4 shows the moment of the greatest proximity between the pedestrians in Fig. 3. The yellow and blue frames present the area in which the pedestrians are located recognized by YOLO. The red markers are used to indicate the pedestrian center points on the ground plane. The distance in the image between the red markers (pedestrian center points) is 14 pixels, which corresponds to about 150 cm after taking into account the mapping from 3D world space to 2D camera space, while the distance between the pedestrian occupied areas (represented by the yellow and blue frames) in the image is 14 pixels, which corresponds to about 67 cm in 3D space.

Another characteristic which is important in the context of calibration and validation of pedestrian dynamics models is related to the motion paths for



Fig. 3. Two pedestrians passing each other on the sidewalk of the AGH Main Street. The image shows the frames from a video sequence recorded every 4 seconds.



Fig. 4. The closest approximation between two passing pedestrians. The yellow and blue frames were generated by YOLO and show the area in which the pedestrian was identified. The red markers show the calculated center point of the pedestrian on the sidewalk plane.

individual pedestrians. These paths allow us, among other things, to verify if the applied model of interaction between pedestrians (based on e.g. social distances model) is correctly determined. An example of determined motion paths based on images from webcams is shown in Fig. 5.



Fig. 5. Example of the motion paths detected for two pedestrians (highlighted in red and green, respectively). The time between successive positions of the pedestrians presented in the image is 2 seconds.

3 Conclusions

As part of the project, we created and tested an application that allows tracking the trajectory of people, determining the distances between people and indicating space utilization and some configuration patterns in the crowd. Thus, we have an important element in the data-driven modeling scheme based on the images from web-cameras. Calibrating and validating pedestrian dynamics models using data from public web cameras has many advantages. These undoubtedly include the high availability of data from different geographical locations (from different cultures) and social situations (e.g. under pandemic conditions). Additionally, thanks to widely available object detectors such as YOLO (even pre-trained on datasets such as COCO), the cost of implementing such a solution is relatively low. However, this approach also has disadvantages with respect to planning and executing experiments involving groups of individuals, the most important of which are related to the inability to plan the observed situations. Therefore, it seems that the best approach for model calibration and validation purposes is to combine these two techniques, which will allow obtaining synergistic effects.

References

1. Bazior, G., Palka, D., J., W.: Using cellular automata to model high density pedestrian dynamics. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **12137 LNCS**, 486 – 498 (2020)
2. Chrysostomou, D., Sirakoulis, G.C., Gasteratos, A.: A bio-inspired multi-camera system for dynamic crowd analysis. *Pattern Recognition Letters* **44**, 141 – 151 (2014). <https://doi.org/https://doi.org/10.1016/j.patrec.2013.11.020>, pattern Recognition and Crowd Analysis
3. Fengjun Lv, Tao Zhao, Nevatia, R.: Camera calibration from video of a walking human. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(9), 1513–1518 (2006). <https://doi.org/10.1109/TPAMI.2006.178>
4. Helbing, D., Molnar, P.: A social force model for pedestrian dynamics. *Phys. Rev. E* **51**, 4284–4286 (1995)
5. Miao, Y., Han, J., Gao, Y., Zhang, B.: St-cnn: Spatial-temporal convolutional neural network for crowd counting in videos. *Pattern Recognition Letters* **125**, 113 – 118 (2019). <https://doi.org/https://doi.org/10.1016/j.patrec.2019.04.012>
6. Porzycki, J., Schmidt-Polończyk, N., Wąs, J.: Pedestrian behavior during evacuation from road tunnel in smoke condition-empirical results. *PLoS One*. **13**(8) (2018), [doi:10.1371/journal.pone.0201732](https://doi.org/10.1371/journal.pone.0201732)
7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection (2016)
8. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. *CoRR abs/1804.02767* (2018), <http://arxiv.org/abs/1804.02767>
9. Sun, Y., Jin, J., Wu, X., Ma, T., Yang, J.: Counting crowds with varying densities via adaptive scenario discovery framework. *Neurocomputing* **397**, 127 – 138 (2020). <https://doi.org/https://doi.org/10.1016/j.neucom.2020.02.045>
10. Tadeusiewicz, R.: Electronic observation and computer monitoring of human behavior in public space. *Napedy i Sterowanie* **12** (2013)
11. Tang, Z., Lin, Y., Lee, K., Hwang, J., Chuang, J., Fang, Z.: Camera self-calibration from tracking of moving persons. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 265–270 (2016). <https://doi.org/10.1109/ICPR.2016.7899644>
12. Tran, K., Gala, A., Kakadiaris, I., Shah, S.: Activity analysis in crowded environments using social cues for group discovery and human interaction modeling. *Pattern Recognition Letters* **44**, 49 – 57 (2014). <https://doi.org/https://doi.org/10.1016/j.patrec.2013.09.015>, <http://www.sciencedirect.com/science/article/pii/S0167865513003516>, pattern Recognition and Crowd Analysis
13. Tytko, K., Mamica, M., Pękala, A., Wąs, J.: Simulating pedestrians’ motion in different scenarios with modified social force model. In: Wyrzykowski, R., Deelman, E., Dongarra, J., Karczewski, K. (eds.) *Parallel Processing and Applied Mathematics*. pp. 467–477. Springer International Publishing, Cham (2020)
14. Wąs, J., Gudowski, B., Matuszyk, P.J.: New cellular automata model of pedestrian representation. In: El Yacoubi, S., Chopard, B., Bandini, S. (eds.) *Cellular Automata*. pp. 724–727. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)