People Counting and Flow Estimation with Video Analysis and OpenCV Database

Final Report for UROP 2100

NI Xiaoyuan

Abstract:

The project is to conquer the people counting and flow estimation problem with Video Analysis technique. The primary aim for this project is to detect the people flow fluctuation on the staircases beside the CYT building in HKUST and to further analyze the effect of spurring people to walk the staircases after the painting decoration. Considering the nature of heterogeneity of pedestrian figures and relatively fixed trajectories that may exist in the video, the project applies a threshold to detect the number of pedestrians passing through the staircases, either walking upwards and downwards. The algorithm basically utilizes the object tracking that is already trained by OpenCV and set a threshold to determine the direction of people heading to. This project may be further optimized in a way to fine-tune the algorithm and to better solve the problem like the fixed angel of the video camera, the cases of figure overlapping as well as the people flow pattern difference between weekdays and weekends and vocations.

Background:

According to different requirement and function, different technique would be used for people counting. Considering the threshold for counting, some focus more on the region-based, using the Region of Interest (ROI) [1]: such as in the open shopping region as the clothes shop with multiple entrances in the department store, people focus more on the customer behavior. Other focus more on the flow counting with relatively fixed flow direction (normally 2-way roads), utilizing the Line of Interest (LOI), that line threshold might be applied [2][3].

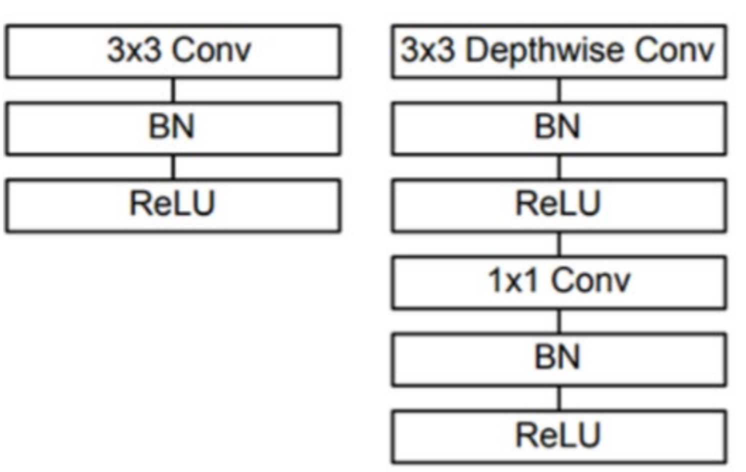
Also, different technique will be used. Most popular one is using the Regression Model [4] that captures the features of the pedestrian figures and further used for people counting. This technique is normally able to subtract the background to improve the computation efficiency and using the regression model to detect the flow size and conquer the figure overlapping problem. Owing to its perfect performance of flow computation in the scenario of mass pedestrian, it was first prompt for the underground people flow computation [5] and can also be applied on various crowded scenario such as the public roads or department stores.

There are also methods that utilize clustering model to realize the people counting task. The model normally suits for the case that simultaneous objects and trajectories moving in the videos. The model [6] clusters the multiple trajectories and thus able to predict the moving directions of pedestrians and count the people entering or exiting certain area. This clustering model are more applicable to Region of Interest scenario and suits more complicated cases that pedestrians are crossing over in multiple directions.

Implement Platform:

The project is implement based on Python3 and CV2 for computer vision package, considering the CV2 package might be easier to implement and work for the purpose. Later, if more flexibility is demanded, other models such as YOLO3 may be considered to improve the changeability of the codes.

The project also uses MobileNets for deep learning the figure of typical pedestrian figures, since it is a more efficient neural network and thus provides with higher speed for video analyze and results in less delay.



Graph 1: Layer Structure of MobileNets

MobileNets is further combined with Single Shot Detectors (MobileNets_SSD) and the deep neural network (DNN) provided in OpenCV in order to finish the task of object detection.

Object Tracking Algorithm:

The project conquers the object tracking problem with 5 steps:

1) For the video figures, use the pre-trained CNN network to detect the objects

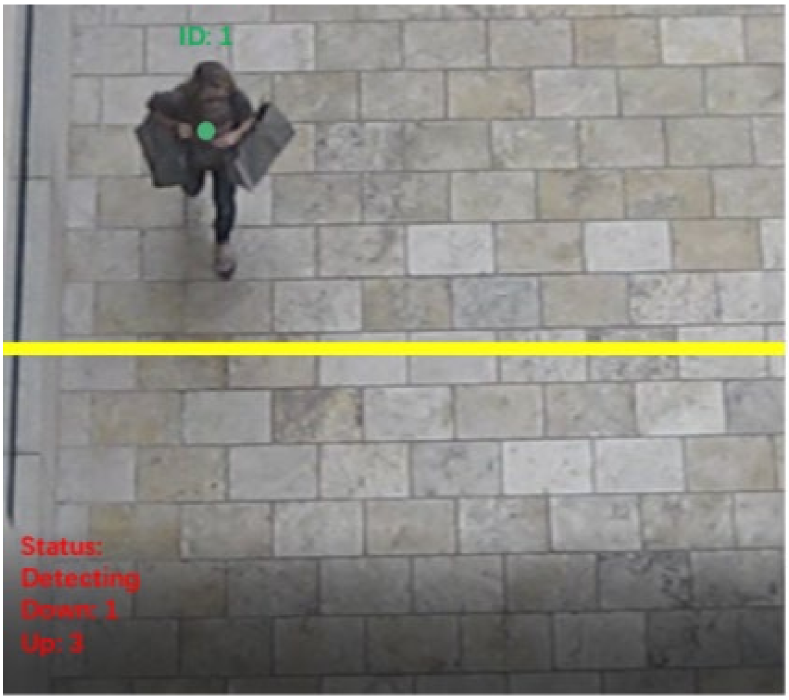
and their bounding boxes. Then, take the central points of each bounding boxes and consider it as the current position of certain object.

2) Follow step 1) in the next frame of the videos. Taking in consider the high cost of CNN computation, normally some frames will be omitted to optimize the analyze. In this case, the author chooses the omission frame to be 40, in other words, the interval the object detection for video frames is set to be 40. For further optimization, the frame interval may be automatically changed to be longer in case to objects exist in the previous frame and during the holiday time and during night when the number of people passing by is more likely to be lower.
3) Compare the positions found in step 1) and 2), regard the point having the smallest Euclidean distance and the distance is less than the largest possible threshold to be the same object.
4) For the unmatched points, if they are detected in the previous frame but don't match the points in the next frame, the objects (people) are regarded to leave the detection area, thus the objects are deleted. In the other way, new object would be added in and object ID would be assigned.
5) For all the objects recorded, the location information (the x, y coordinates) will be updated and prepared for further comparation.

Result Demo:

The algorithm is operated based on CV2 database to detect the pedestrian figure and to count the number of people passing throught the threshold, and tested based on the video from other databases than the real camera video from CYT Building. Though it works perfectly on the cases when few people appear in the scene [graph 1 and graph 2], problems may occure when people are carrying large luggages or 2 people are walking too closed with each other, whith will result in the error [graph 3]. Also, the algorithm doesn’t fit in the case when cars or bicycles are passing through in the video, which is unlikely to happen in the case of staircases.

Graph 1: The successful counting demo



Graph 2: The successful counting demo

Graph 3: Pedestrian with large luggage cannot be detected

Further Work and Application:

The project should be further tested on the real dataset collected from the camera video captured by CCTV camera besides the staircases. Since the angel of the camera is tilted rather than at the roof top, the model in the CV2 should be further trained to get accustomed to the new figures in order to spur better performance. Also, taking the figure overlapping that may happen into consideration, one more camera may have to be install in another angle for further improvement of accuracy. Further researches should also be conducted on how to combine the results of multiple cameras.

Also, this same technique may also serve in various scenario, ranging from the flow detection and monitoring for public infrastructure and management to customer behavior research. For the government, smart people counting system can help analyze the people flow pattern and thus improve the efficiency of public transportation and avoid the serious accident typically occurred in crowded areas. For the personal usage, the retailors may apply this technique to spur their sales and better serve the customer need, the advertisers may better target at potential customer without disturbing other people. Yet, much work must be done, especially to deal with the serious people figure overlapping problem and the classification of figures (pedestrians, cars, etc.).

References:

[1] A. Chan, L. J. Zhang-Sheng, V. Nuno, “Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking,” 2008 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2008.

[2] S. ShiJie, A. Naveed, S. ShengHuan, Z. ChaoYang, L. JianXin and M. Ajmal, “data and method for real-time people counting in cluttered scenes using depth sensors,” [Online]. Available: arxiv.org/pdf/1804.04339.pdf.

[3] Y. Cong, H. Gong, S. C. Zhu, and Y. Tang, “Flow mosaicking: Realtime pedestrian counting without Scene-specific learning,” 2009 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2009, pp. 1093–1100.

[4] L. Del Pizzo, P. Foggia, A. Greco, G. Percannella, and M. Vento, “Counting people by RGB or depth overhead cameras,” *Pattern Recognition Letters*, vol. 81, pp.

41–50, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2016.05.033

[5] N. Paragios and V. Ramesh, “A mrf-based approach for real-time subway monitoring,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 1034–40.

[6] I. S. Topkaya, H. Erdogan, and F. Porikli, “Counting people by clustering person detector outputs,” *2014 IEEE International Conference on Advanced Video and Signal-Based Surveillance*, AVSS 2014, 2014, pp. 313–318.

[7] H. Andrew, Z. Menglong, C. Bo, K. Dmitry, W. Weijun, W. Tobias, A. Marco and A. Hartwig, ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” [Online]. Available: https://arxiv.org/abs/1704.04861