Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University | UNIVERSITY of PENNSYLVANIA

#19 MONITORING AND PREDICTING PEDESTRIAN BEHAVIOR USING TRAFFIC CAMERAS

Final Research Report

Luis E. Navarro-Serment, Ph.D.

The Robotics Institute

Carnegie Mellon University

November 25, 2018.

Disclaimer

#019: Monitoring and Predicting Pedestrian Behavior using Traffic Cameras

Problem

This project addresses the need for timely and accurate information about pedestrian traffic in urban areas. This is particularly important at locations where it is not uncommon to find more pedestrians than vehicles during certain times of the day. Through the work conducted in this project, we move closer to deploying systems for detecting, tracking, and forecasting human behavior in dynamic environments. This information can be used by other systems as part of an infrastructure-based framework to effectively protect pedestrians: the more vulnerable traffic participants.

Most intersections lack awareness of pedestrian traffic: their perception infrastructure—when available—is usually limited to the detection of vehicles at very specific places. In this study we consider the use of video cameras to monitor pedestrian traffic in settings where a static camera that has an unobstructed view of the road is used to detect and track pedestrians. An example set up is shown in Fig. 1, where four monocular cameras mounted on the traffic light structures monitor a busy intersection. Typically, a single camera cannot cover the entire area, so multiple cameras are used at each intersection; this increases costs, and consequently it is desirable to use as few cameras as possible.

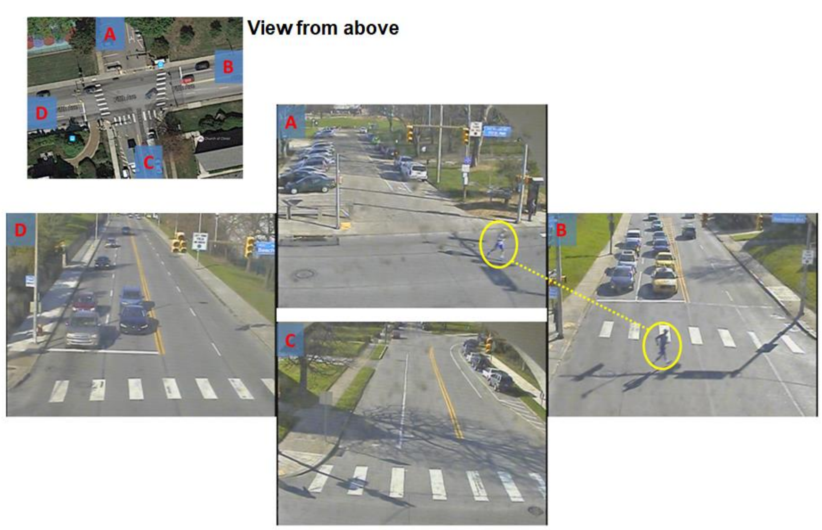

Fig. 1: multiple cameras provide coverage for one intersection.

Intersection monitoring starts by detecting pedestrians from video image frames. However, the detection alone is not enough to fully understand the motion of people within the context of the intersection: it is also necessary to determine *where* people are located, and also to *predict* where they are going.

In previous research efforts we have addressed the detection and placement of pedestrians present at the intersection. In this effort we focused on the prediction aspects of the monitoring process. In particular, we focused on creating interaction models that take into account influences among other agents in the scene (i.e. other pedestrians), rather than independent models for each single agent as we had done before in preliminary developments. Our previous work focused on producing predictions for each mover independently of other pedestrians, which is adequate for sparsely populated environments. However, in more crowded scenarios, it is necessary to take into consideration the interactions with and among other movers. This is a very challenging problem, which in its most general form constitutes an unsolved research problem. Therefore, in this effort we focused on extending our prediction framework to accommodate these multi-agent interactions. Finally, predictions must be generated in real time to be useful. Previous implementations required lengths of time that were unsuitable for this requirement. Therefore, the last objective of this work was producing a software implementation capable of generating predictions in real time.

To accomplish our goals within the project's constraints, we leveraged different elements developed in some of our other projects. We applied methods to extract priors from external sources of information, which provide a richer initial context for predictions at a particular site, e.g. potential destinations for pedestrians, like crosswalks and bus stops. Additionally, we adapted methodologies for intrinsic and extrinsic camera calibration for use at the site of each traffic camera. Finally, we leveraged our recent work on modeling interactions between pedestrians to extend our prediction framework, originally capable of producing predictions for each mover independently of other pedestrians, to generate multi-agent predictions using simple heuristics.

Our approach

During this research we incorporated enhancements to our original design of prediction for multiple pedestrians. These enhancements reduced the prediction computation time; this is relevant to our goal of real time operation.

We have made a number of improvements to the prediction code, in order to streamline the generation of predictions for multiple movers. We have tested the prediction performance—in terms of computation time—of this approach using pedestrian data publicly available from the University of Oxford[1]. The dataset contains video collected from an overhead surveillance camera looking at an urban commercial area. People walk through the camera's field of view, and the location of their faces in the image has been annotated. More importantly, pedestrian tracking data is also provided, which is useful as ground truth to evaluate detection and tracking performance. A sample view is shown in Fig. 2, where each pedestrian detected is assigned a unique track ID number. The figure also shows all the pedestrian interactions analyzed at that time, indicated by the blue lines. The heuristic algorithm is based on principles from proxemics and focuses only on those interactions that are potentially meaningful, e.g. it ignores interactions between pedestrians that are far away from each other and therefore unlikely to affect each other's movement. Similarly, it identifies people moving together as a group based on speed, distance between them, and direction of motion; this reduces the number of computations by generating only one prediction for the group instead of individual predictions for each pedestrian.

The dataset includes various groups of pedestrians at any given time, ranging from only a few sparsely distributed individuals to multiple people walking in close proximity to one another. Since the extrinsic camera calibration is also available, and assuming an average person's face height, we estimate each pedestrian's location on the world frame. These locations then serve as input to the prediction algorithm. Additionally, we created an obstacle map of the scene, which includes walls, a couple of benches, and a trash can (these are visible in Fig. 2). This map is represented by a 19 x 37 grid, 1m cell size. Similarly, using information from external maps we obtained a probability map of potential destinations, which are needed by the prediction algorithm.

[1] http://www.robots.ox.ac.uk/ActiveVision/Research/Projects/2009bbenfold_headpose/project.html#datasets

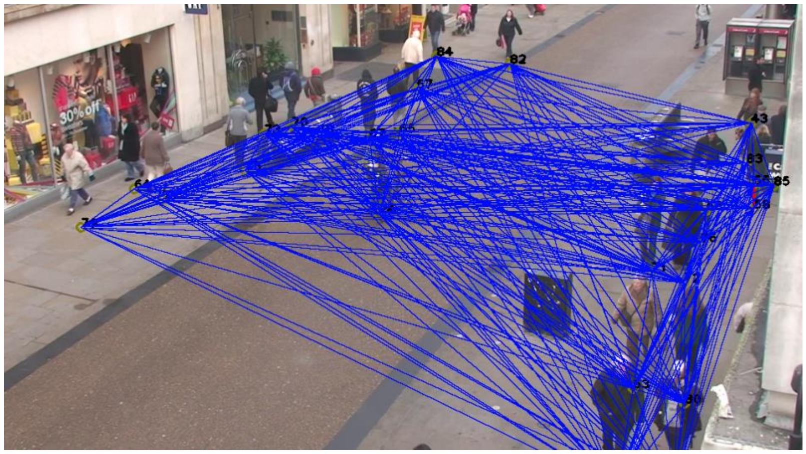

Fig. 2. Analysis of interactions between multiple pedestrians. Only a few interactions will impact the movement of other pedestrians significantly.

As we did in previous studies, we processed 2.3 minutes of video data, and executed the prediction algorithm every 25 frames. For this test, we used a computer with an Intel Core i5-2520M 2.5GHz dual core CPU. For consistency, this computer maintained the same the same Operating System used for our previous characterization. We measured and recorded the prediction computation time for each execution cycle, and compared it against the results obtained with the initial version of our iterative algorithm. The results are shown in Fig. 3. As seen in the figure, the latest version runs considerably faster than the previous two versions. A single prediction, after incorporating the potential conflicts and modifying the obstacle map accordingly, is executed in ~0.011 sec. per target.

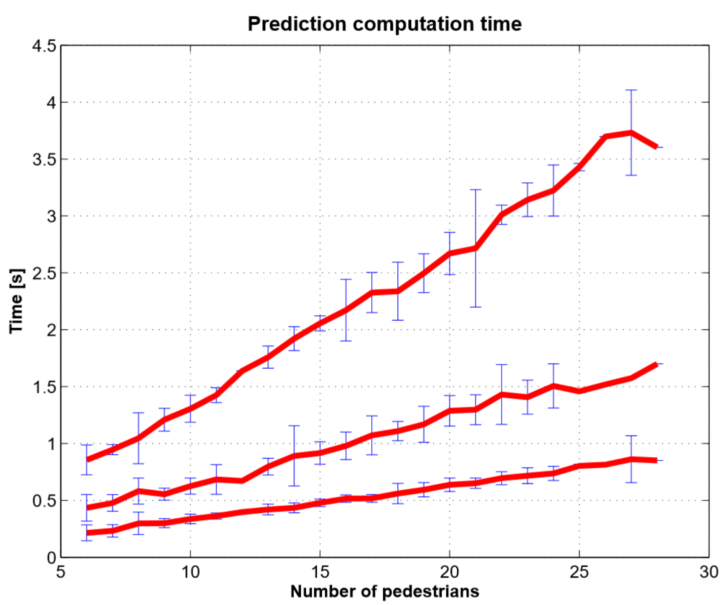

Fig. 3. Performance of the prediction framework. The graph depicts the prediction computation time vs. the number of pedestrians. The top trace illustrates the computation time for the original version; the middle trace represents the first improved version. The bottom trace corresponds to the latest version. The error bars indicate the 2-sigma limits in the times measured for each number of pedestrians.

Findings

The results confirmed the system's ability to generate predictions of pedestrian trajectories. An example is shown in Fig. 4, which depicts predictions for the pedestrian designated with ID=133. The left image shows a map of expectations of future visitation, where the color of each cell indicates the probability that track 133 will step inside that cell as it moves. A red color denotes a high probability; a blue color denotes low probability. This map includes predictions for all cells in the grid; however, this representation can be difficult to interpret for certain applications. The right figure presents a different representation, where specific paths are extracted from the map and ranked according to the probability of being taken. The same coloring scheme is used, so the red path indicates the most likely trajectory.

Fig. 4. Sample predictions. A map of expectation of future visitation (left) and specific paths ranked according to probability (right) are shown. The colors indicate probability, going from red (high) to blue (low).

Participants

Luis E. Navarro-Serment, PI

Students:

Ingrid Navarro-Anaya
Tatsuya Ishihara
Dragan Ahmetovic
Eshed Ohn-Bar
Nick Rhinehart
Arjun Sharma

Deployment Partners/Participating Organizations

Rapid Flow Technologies, LLC
124 South Highland Ave
Suite 206
Pittsburgh PA 15206

Provides: access to video data from traffic intersections, expertise on practical issues.

Other publications, conference papers and presentations

Presentation at the *2017 RCTA World Model Workshop*, "Perception systems for autonomous robots", Philadelphia, PA, May, 2017.

Outcomes

We have developed a video processing pipeline to detect people from images, and to track them over multiple video frames. We have also developed a heuristic-based algorithm to predict their trajectories, taking into account relevant interactions with other movers. Finally, we have implemented the prediction code as a software library, which is capable of operation in real time.

Other Products associated

To support the calibration approach and the methodology for person location within the intersection, in previous effort we designed and constructed a low-cost 3D scanner. This scanner, built around a low-cost 2D laser range finder, allows us to obtain three-dimensional models of traffic intersections quickly and accurately, but at a fraction of the cost of more expensive scanners commercially available. The software used to calibrate the low-cost scanner is also available from:

https://github.com/cmu-navlab/calibrate-scanner

Impact

The approaches developed in this effort move us closer to provide traffic intersections with the ability to monitor pedestrian activity in several ways: detection of humans, tracking their movements, and forecasting their trajectories. Most traffic intersections currently lack awareness of pedestrian traffic: their perception abilities—when available—are usually limited to the detection of vehicles at very specific places. Video cameras can be used to monitor pedestrian traffic in a setting where a static camera that has an unobstructed view of the road is used to detect and track pedestrians. This type of sensor offers a cost –effective alternative over other types of sensors suitable for pedestrian monitoring.

Impact in other disciplines

We anticipate that our research will have an impact on adaptive traffic light control systems, which currently operate entirely based on information pertaining vehicular traffic. Our work will alleviate the need for timely and accurate information about pedestrian traffic. This is particularly important at locations where it is not uncommon to find more pedestrians than vehicles during certain times of the day.