Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University | UNIVERSITY of PENNSYLVANIA

#65 MONITORING AND PREDICTING PEDESTRIAN BEHAVIOR AT TRAFFIC INTERSECTIONS

Final Research Report

Luis E. Navarro-Serment, Ph.D.

The Robotics Institute

Carnegie Mellon University

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

#65 Monitoring and Predicting Pedestrian Behavior at Traffic Intersections

Problem

Most traffic intersections 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. An example 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.

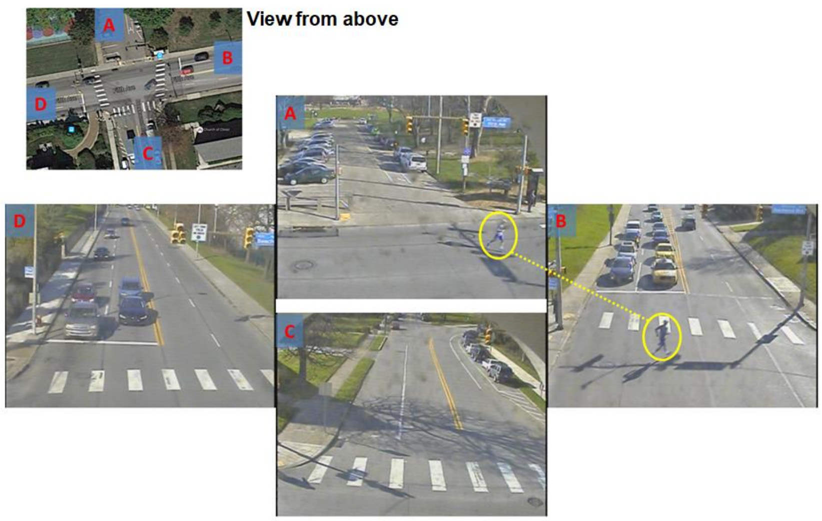

Fig. 1: multiple cameras provide coverage for one intersection. This example corresponds to the intersection of Fifth Ave. and Beechwood Blvd., Pittsburgh, PA 15206.

However, simply detecting a pedestrian is not enough to fully describe the motion of that person within the context of the intersection: it is also necessary to determine *where* that person is located. Furthermore, a person crossing the roads could be seen by several cameras, as in the example shown in Fig. 1. Therefore, detections from different cameras must be associated correctly so that each pedestrian is recognized as unique for the duration of their transit near or through the intersection. Unfortunately, the estimation of depth using a monocular camera is ambiguous. An accurate estimation requires of a second camera (a stereo pair), but this solution is typically avoided to keep costs down. Consequently, other techniques must be used to estimate the distance of objects to the camera.

The correct placement of a person within the intersection requires both the *intrinsic* and *extrinsic* calibration of each camera, so that geometric locations on the surfaces where pedestrians move can be transformed to pixels in each camera's image and vice versa. Traditional approaches to camera calibration work well for the intrinsic part, and can be conducted before the cameras are mounted. However, the extrinsic calibration part requires elements that are not well suited for this application (e.g. collection of multiple images showing a calibration target at different locations), and are time consuming. Furthermore, the calibration task can be affected by traffic interference, and conversely it can affect the flow of traffic itself; the use of common calibration targets is impractical for this application (a very large target is needed given the distance between the camera and the movers); and the surveying equipment—such as total stations and scanners—is costly. Finally, calibration is typically carried out by specialized personnel.

This study intends to reduce the time and amount of effort needed to calibrate and set up a video camera at any location. In this work we focused on developing an approach to calibrate traffic cameras that simplifies the extrinsic calibration work on-site using a low-cost custom-made laser scanner, and that requires minimal personnel training.

Our approach

To reduce the time and amount of effort required to set up a video camera at any location, we developed a methodology based on the use of a low-cost laser scanner. The idea is to create a 3D model of the scene faced by the camera using this scanner. This produces an accurate geometric representation of the environment, which in turn is used to establish correspondences between pixels in the image and locations on the scene. Additionally, this representation is used to model the ground surface; this information is used to determine the location of a person seen from the camera by imposing a simple constraint, i.e. pedestrians walk on this surface, so their feet touch the ground. By knowing in which direction from the camera pedestrians touch the ground, we can determine their location using the model.

Low-cost scanner

We designed and constructed a low-cost scanner. Our design uses a Hokuyo UTM 30LX line scanner, which is attached to a precision servo that manipulates the scanner in a nodding pattern. In this manner, the scanner collects range measurements at different angles; these measurements are merged in to a single point cloud, as shown in Fig. 2. The scanner has a 30 m range, and collects measurements with an angular resolution of 0.25°. This scanner has a price tag of $5,000, and is the costliest part of the assembly, which can be built for $6,500. This is comparatively inexpensive compared with commonly used scanners, whose price ranges from $75,000 to $120,000. The unit was built using COTS components, only requiring a structural support frame which was 3D printed in-house. Additionally, we wrote code to calibrate this scanner, and made it publicly available through Github[1].

Typically, the scanner needs to be placed at multiple locations near the intersection to ensure full area coverage. It takes less than 10 sec. to complete a full cycle; this is repeated at each location. We used publicly available software[2] to align the individual point clouds and merge them into a single model of the intersection.

---

[1] The link is provided in a subsequent section.

[2] We used CloudCompare (https://www.danielgm.net/cc/); other software packages can be used as well.

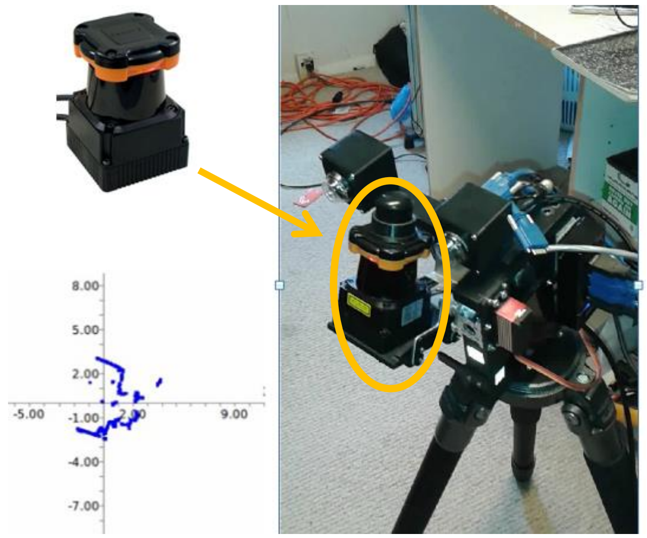

Fig. 2: Low-cost 3D scanner, based on a line scanner (top left) that produces range measurements (bottom left). The scanner is attached to a precision servo motor that rotates the scanner (right), allowing it to collect measurements at different angles that are merged together to create a point cloud.

Calibration of traffic cameras for pedestrian monitoring

Our calibration approach follows a simple modification of the traditional method: we replace the typical calibration target (usually, a high-contrast checkerboard of known size) by the 3D model collected from the intersection. The approach is summarized as follows:

- **Input**: 3D model of the scene in front of the camera
- **Clean up**: remove noise, and all elements of dynamic scene
  - Merge multiple sub-models using alignment algorithms (e.g. ICP, 4PCS).
- Select and locate a set of markers in 3D model's coordinate frame that are seen in the image
- Find matching markers in image coordinates (e.g. pixel coordinates)
- Find calibration parameters using Direct Linear Transformation first, then Levenberg-Marquardt if significant distortion is present
- Verify calibration results by re-projecting 3D points into the image plane
- **Output**: extrinsic calibration matrix, representing a homogeneous transformation $\mathbf{H}_c^w$

In subsequent steps, the ground is segmented and modeled separately using a 3D mesh. This is illustrated in Fig. 4. By detecting the location of people in the images, using the extrinsic calibration is possible to estimate the location of where the person touches the ground and consequently, their location.

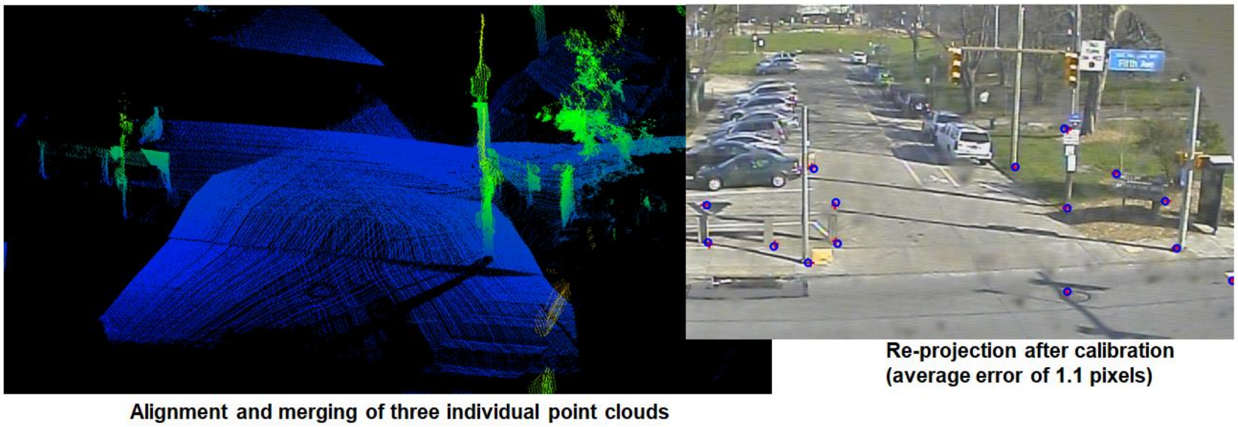

Fig. 3: Left.- a 3D model of the site is used to calibrate the camera that covers it. Right.- the claibration is tested by reprojecting locations in the 3D model (red markers) back into the image plane (blue markers).

Findings

Experiments conducted with this approach confirmed its ability to obtain an accurate extrinsic calibration of the site. In the example shown in Fig. 3, the location was modeled by collecting 3D data from 3 different locations; this process was conducted in the evening to minimize interference from traffic and pedestrians, and took less than 5 minutes. The calibration was conducted off line, and achieved an average error of 1.1 pixels which is acceptable for this application.

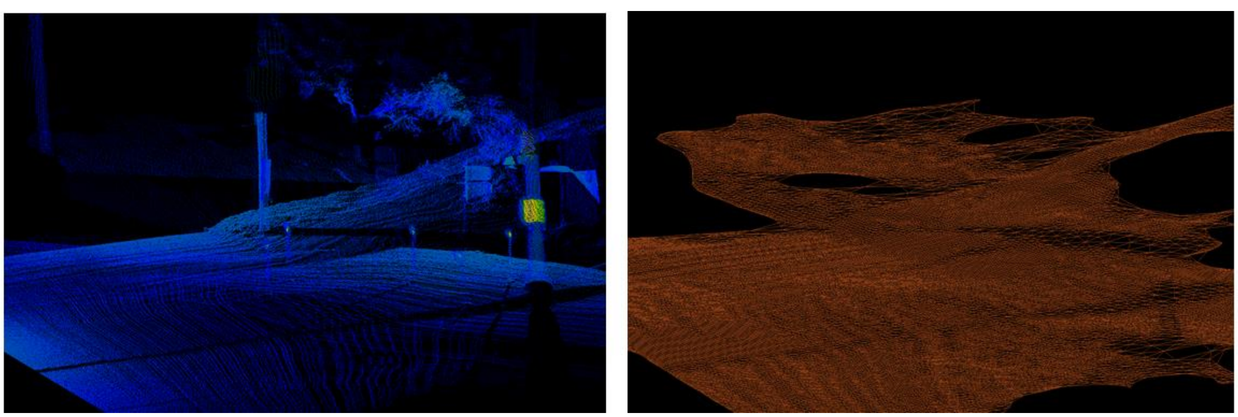

Fig. 4: Left.- a complete 3D model is constructed by merging three separate point clouds. Right.- the ground surface is estimated from the 3D model; this mesh represents the surface where pedestrians move.

Participants

Luis E. Navarro-Serment, PI

Students:

Ana Beisy Cruz
Youngwook Do
Praneet Dutta
Venkatesh Manikavasegam
Meghana Reddy Guduru

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 2016 MASITE/ITSPA Annual Conference, "Calibration of traffic cameras using low-cost 3D scanners", State College, PA, Aug. 29, 2016.

Outcomes

We have developed a video processing pipeline to detect people from images, which is customized for operation with the type of cameras currently used to monitor vehicular traffic.
We have also developed an approach to calibrate traffic cameras on-site, which is inexpensive in terms of time and logistics; does not require expensive instruments or software packages; uses a low-cost custom-made laser scanner; and can be performed by personnel with minimal training.
Finally, building upon the calibration approach, we have developed a methodology to determine the location of a person in an image with respect to the geometry of the traffic intersection, and considering all the cameras covering the intersection.

Other Products associated

To support the calibration approach and the methodology for person location within the intersection, 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. We plan to make our design (i.e. mechanical design and accompanying software) freely available to other researchers in the near future. This will facilitate the adoption of our camera calibration methodology by other agencies.

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. 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. Typically, a single camera cannot cover the entire area, so multiple cameras are used at each intersection. However, simply detecting a pedestrian is not enough to fully describe the motion of that person within the context of the intersection: it is also necessary to determine where that person is located. Furthermore, a person crossing the roads will likely be seen by several cameras. Therefore, detections from different cameras must be associated correctly so that each pedestrian is recognized as unique for the duration of their transit near or through the intersection. The correct placement of a person within the intersection requires both the intrinsic and extrinsic calibration of each camera, so that points on the surfaces where pedestrians move can be transformed to pixels in each camera's image and vice versa. Traditional approaches to camera calibration require elements that are not well suited for this application (e.g. collection of multiple images showing a calibration target at different locations, or use of expensive instruments for dimensional analysis), and are typically carried out by specialized personnel. Our work simplifies the calibration task and therefore makes it easier to set up pedestrian monitoring systems at more locations.

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.