Development of Correction Models for the Accurate Estimation of Pedestrian Volumes Obtained from Infrared Sensors

Kaan Ozbay, *Rutgers, The State University of New Jersey*
Hong Yang (Corresponding Author), *Rutgers, The State University of New Jersey*
Bekir Bartin, *Rutgers, The State University of New Jersey*

Recently, there has been an increasing interest in the development of pedestrian oriented planning models to ensure the sustainability of our cities as well as our transportation system. However, unlike vehicle data, pedestrian data is not always readily available through infrastructure sensors that are deployed throughout the highway network. Transportation agencies are looking for ways of collecting continuous and reliable pedestrian data that can be used to develop pedestrian oriented models. However, manual pedestrian data collection can be quite expensive and given the meager resources of many agencies this might pose a major challenge.

Low-cost methods for collecting bicycle and pedestrian trafficdata will thus be a critical component in developing and calibrating models that include non-motorized travel *(1)*. Emerging sensor technologies provide new opportunities for gathering such datafor project-level use, for instance, pedestrian or bicycle data for a planning or improvement project of a trail, a street, or an intersection. However, the accuracy of the sensor data still needs careful consideration. This paper focuses on the development of a statistically robust procedure to correct the raw sensor counts obtained from an infrared sensor. Field data collection efforts as well as a brief discussion of the collected data are presented next. Then, statistical models based on the relationship between the arrival patterns of pedestrians and total flows are presented. Then, results of tests to show the validity of the developed models are presented. The paper is concluded by discussing our major findings and their implications on long-term pedestrian data collection.

Data Collection using an Infrared Sensor

Instead of using manual data collection methods, automaticcounting technologies have been increasingly employed forlong-term non-motorized travel data collection. Many pedestrian counting technologies such as infrared, microwave, radar, computer vision, and myriad of other technologies, are already commercially available. Among these technologies that have been already deployed, infrared is one of the most widely used technologies. For instance, the city government of Cheyenne, Wyoming, installed an infrared laser counter to record a path counts to justify the usage of the greenway system in 1990s *(2)*. In 2002, the Ohio Licking County Area Transportation Study began installing passive infrared counters along a shared-use path system to provide data for a comprehensive bicycle and pedestrian plan *(3)*. An active infrared counter was also placed above the Norwottuck trail in Amherst, Massachusetts to measure pedestrians and bicycles use in 2001*(4)*. However, lessons learned from these applications show that none of these counters performed perfectly. In fact, many commercially available infrared sensors have well-known error issues *(5, 6)*.

If the raw data obtained from these sensors are directly used for pedestrian travel analysis, for instance, identifying the peak-hour of pedestrian activity, these errors can be an important source of concern. Data collected on busy facilities such as urban sidewalks and intersections should be particularly used with caution. This is due to the fact that most of these sensors will systematically undercount pedestrians that are walking side by side as shown in Figure 1. It has been determined that the missing counts could be more than 20 percent of the actual volume when the sensor was deployed on some busy facilities *(7)*.

(a) Group 1: single arrival (b) Group 2: two simultaneous arrivals (c) Group 3: three simultaneous arrivals

Figure 1. Typical pedestrian arrival patterns

Estimation of Pedestrian Volumes from Sensor Data

On typical pedestrian facilities people randomly arrive in different group sizes. A recent field study conducted by Ozbay et al. *(7)* found that there is a high correlation between the pedestrian volume and the number of people arriving in groups. Careful investigation of the field data showed that pedestrian arrivals can be mainly represented in terms of groups of one, two, and three as shown in Figure 1. It is thus hypothesized that this empirical relationship between arrival patterns and the sensor output can be used to establish a relationship between sensor counts and the actual volumes. Two models specified as a function of flows are proposed to estimate actual volumes:

$$Group2 = \beta_{20} + \beta_{21} \times SensorCounts \quad (1)$$

$$Group3 = \beta_{30} + \beta_{31} \times SensorCounts \quad (2)$$

$$RealCounts \ = SensorCounts + \frac{1}{2} \times Group2 + \frac{2}{3} \times Group3 \quad (3)$$

'SensorCounts' is the sensor output for each time interval. It is assumed that the each sensor count represents a single pedestrian. 'Group2' and 'Group3' represent the estimated number of groups of two and three pedestrians simultaneously arriving within the same time interval. 'RealCounts' is the prediction of actual pedestrian volume. β_{20}, β_{21}, β_{30}, and β_{31} are parameters to be estimated. If two people arrive as a group, there will be one missing counts. Similarly, there will be two missing counts for a group of 3 pedestrians arriving simultaneously. So "1/2" and "2/3" in equation (3) are correction factors for the missing counts of group2 and group3.

Test for Model Validation

EcoCounter *(8)*, an infrared pedestrian sensor was employed. It is expected that the actual pedestrian volumes could be accurately estimated using the sensor output in conjunction with the developed models. To test the performance of estimated models, field data were collected at two high volume trails at Rutgers Busch campus, New Jersey. Detailed information about the field data collection effort is shown in Table 1. Video recordings of the pedestrians were used to determine the true pedestrian volumes as a benchmark by which the sensor data and the predicted pedestrian volumes could be compared.

Table 1. Field data collection

Dataset	Location	Date	Time	Total Flow	Hourly Average Flow
1	Site 1	12-Mar	10:30am-10:30pm	3686	307
2	Site 1	13-Mar	10:30am-10:30pm	2148	179
3	Site 1	10-Apr	10:30am-10:30pm	3103	259
4	Site 1	13-Apr	10:30am-10:30pm	3995	333
5	Site 1	14-Apr	10:30am-10:30pm	3781	315
6	Site 1	15-Apr	10:30am-08:30pm	3299	330
7	Site 2	19-Oct	09:00am-11:00pm	8570	659

Data were aggregated into15-minute and 1-hour time intervals. The datasets 1, 2, 4, 5, and 6 collected at site 1 were used to estimate the parameters of equations (1) and (2). Dataset 3 and dataset 7 were used to test the performance of the estimation method. If the estimated modelsperform relatively well, the predicted results should outperform the raw sensor outputs. All datasets include ten hours or more of continuously collected counts at each site. These datasets are significantly larger than the datasets used in similar previous studies where no more than four hours of data were used to validate the performance of automatic pedestrian sensors *(5, 6)*. However, these sample sizes are still relatively small for conducting realistic planning studies where long-term data collection is required to be able to capture temporal changes in pedestrian travel. To remedy this issue of limited data, a bootstrapping procedure that can expand the available datasets was proposed by Ozbay et al. *(7)*. Using this procedure, two models were estimated for two different time periods namely, 15 and 60 minute time intervals. For 15-minute intervals, estimated models are as follows:

$$\widehat{Group}2 = 0.106 + 0.371 \times SensorCounts \quad (4)$$

$$\widehat{Group}3 = -0.187 + 0.097 \times SensorCounts \quad (5)$$

For 1-hour intervals, estimated models are as follows:

$$\widehat{Group}2 = 1.944 + 0.365 \times SensorCounts \quad (6)$$

$$\widehat{Group}3 = 0.897 + 0.091 \times SensorCounts \quad (7)$$

Where$'SensorCounts'$ is the raw sensor output, and $\widehat{Group}2$ and $\widehat{Group}3$ are estimated results.

Using the above equations (3) through (7), the actual volume could be predicted.

Results and Discussions

Comparisons between the cumulative sensor counts and the observed cumulative volumes shown in Figure 2 illustrate that the infrared sensor clearly undercounts. By the end of the test periods, the sensor outputs are 20.5 percent and 22.1 percent less than the ground truth data collected on April 10 and October 19, respectively. In contrast, estimated models applied to raw sensor outputs significantly reduce the gap between the actual and corrected volumes. Estimated volumes are 0.7 percent and 2.7 percent less than the overall actual volumes on April 10 and October 19, respectively.

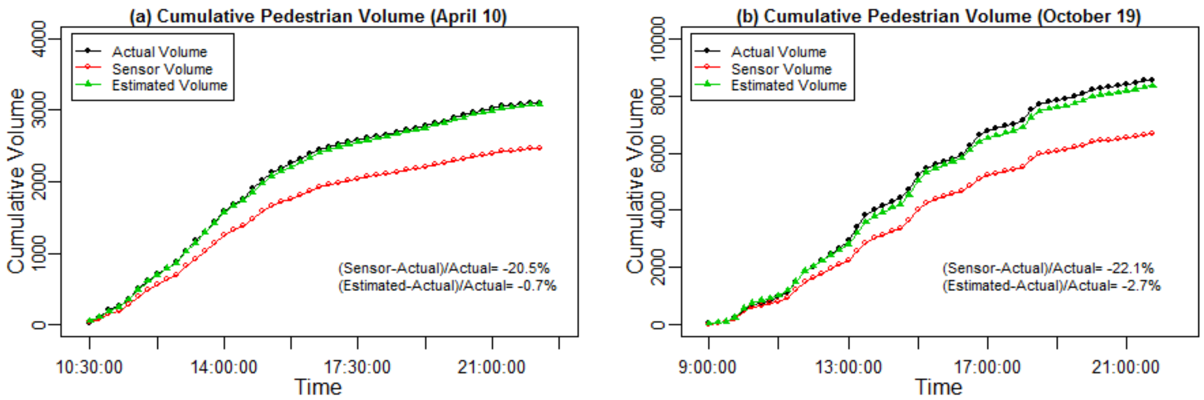

Figure 2. Cumulative pedestrian volume of the field tests

In addition to the overall improvements, estimated counts were investigated in detail. Figure 3 presents the distribution of estimated counts over time. It can be seen from Figure 3 (a), 3 (b), 3 (c), and 3 (d) that the estimated pedestrian volumes over time match the trend and level of the actual volumes significantly better than that of sensor outputs. To statistically confirm this observation, Wilcoxon matched-pairs signed-rank test was conducted to test the difference between the observed volume and sensor outputs, and between observed and estimated volume at a significance level of 0.05. The p-values for the comparisons between observed and sensor counts are all less than 0.05. This result is confirmed for both the 15-minute interval scenarios and the 1-hour interval scenarios using both test datasets. Moreover, test results indicate that the direction of the difference—the original sensor outputs are significantly less than the actual observed volumes. However, the p-values are all more than 0.05 for the comparisons between actual volumes and the estimated which suggest a good match between estimated and observed counts.

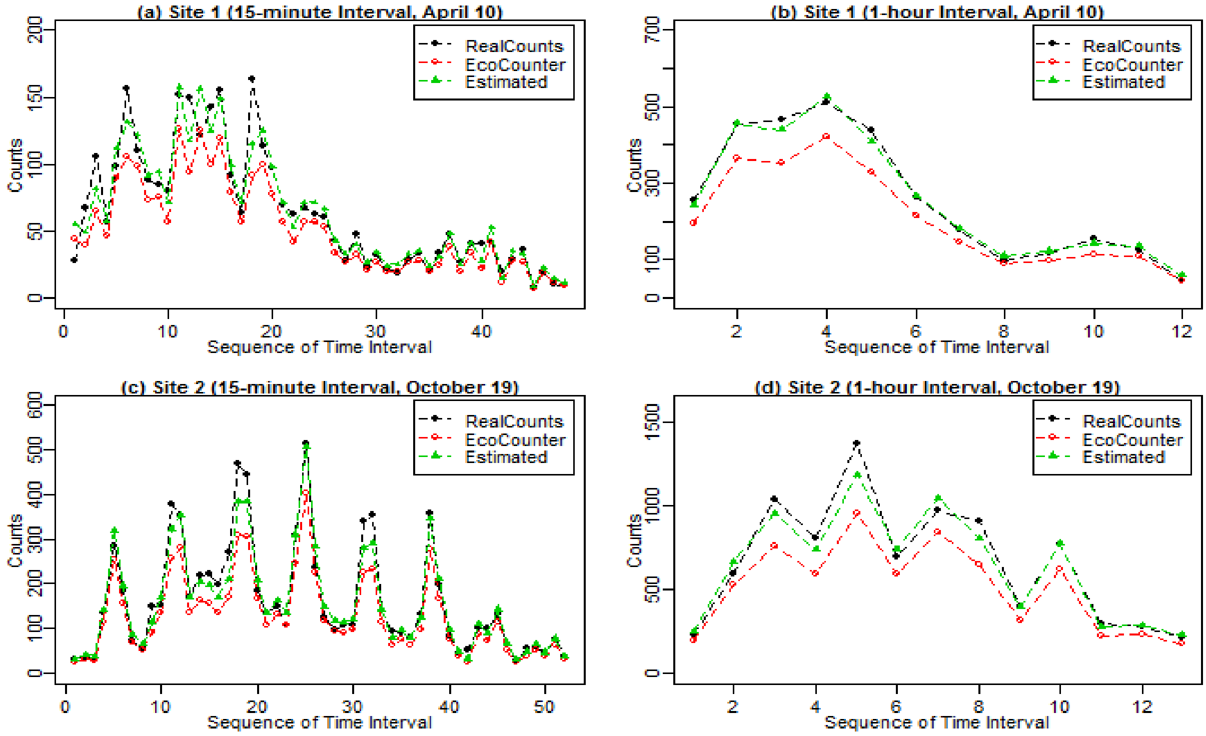

Figure 3. Pedestrian volume over time

CONCLUSIONS

Automatic pedestrian sensors appear to be viable alternative to manual counts for long-term pedestrian data collection. However, in many cases sensor data cannot be directly used due to counting errors caused by the technological limitations of these sensors and arrival patterns of pedestrians. The difference observed between actual volumes and sensor counts will lead to inaccurate understanding of the pedestrian activities if the collected data are used without further processing. To be able to use the sensor data to provide reliable information, correction models based on the relationship between sensor data and arrival patterns of pedestrians are proposed to estimate actual volumes. Test results suggest that the proposed statistical correction approach performed well at two high volume sites. It is expected that the proposed method can be applied to other similar pedestrian facilities given that there is an increasing need to automate the pedestrian data collection process for long-term data collection programs through the use of emerging sensor technologies.

REFERENCES

1. Porter, C., Suhrbier, J., and Schwartz, WL. Forecasting Bicycle and Pedestrian Travel: State of the Practice and Research Needs. In *Transportation ResearchRecord: Journal of the Transportation Research Board, No.1674,* Transportation Research Board of the National Academies,Washington, D.C., pp.10-18, 1999.
2. Schneider, R., Patton, R., Toole, J., and Raborn, C. *Pedestrian and Bicycle Data Collection in United States Communities: Quantifying Use, Surveying Users, and Documenting Facility Extent*. FHWA, U.S. Department of Transportation,, January 2005.
3. Zegeer, C.V., Sandt, L., and Scully, M. *How to Develop a Pedestrian Safety Action Plan*. FHWA Office of Highway Safety, and National Highway Traffic Safety Administration, February 2006.
4. Noyce, D., and Dharmaraju, R. *An Evaluation of Technologies for Automated Detection and Classification of Pedestrians and Bicyclists*. FHWA, and University of Massachusetts Transportation Center, May 2002.
5. Turner, S., Middleton, D., Longmire, R., Brewer, M. and Eurek, R. Testing and Evaluation of Pedestrian Sensors: Texas Transportation Institute, Texas A&M University, September 2007.
6. Greene-Roesel, R., Diogenes, M.C. Ragland, D.R., and Lindau, L.A. Effectiveness of a Commercially Available Automated Pedestrian Counting Device in Urban Environments: Comparison with Manual Counts. Presented at 87thAnnual Meeting of theTransportation Research Board, Washington, D.C., 2008.
7. Ozbay, K., Yang, H., and Bartin, B. Calibration of An Infrared-Based Automatic Counting System for Pedestrian Traffic Flow Data Collection. Accepted for presentation and publication in the proceedings of the89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
8. EcoCounter, People Counter Systems Webpage. http://www.eco-compteur.com/. Accessed on 12/01/2008.