

A new approach for pedestrian density estimation using moving sensors and computer vision

Eric K. Tokuda¹, Yitzchak Lockerman³, Gabriel B. A. Ferreira¹,
Ethan Sorrelgreen², David Boyle², Roberto M. Cesar-Jr.¹, and
Claudio T. Silva³

¹University of São Paulo, Brazil

²Carmera Inc., USA

³New York University, USA

July 6, 2020

Abstract

An understanding of person dynamics is indispensable for numerous urban applications including the design of transportation networks and planing for business development. Pedestrian counting often requires utilizing manual or technical means to count individuals in each location of interest. However, such methods do not scale to the size of a city and a new approach to fill this gap is here proposed. In this project, we used a large dense dataset of images of New York City along with computer vision techniques to construct a spatio-temporal map of relative person density. Due to the limitations of state of the art computer vision methods, such automatic detection of person is inherently subject to errors. We model these errors as a probabilistic process, for which we provide theoretical analysis and thorough numerical simulations. We demonstrate that, within our assumptions, our methodology can supply a reasonable estimate of person densities and provide theoretical bounds for the resulting error.

1 Introduction

Pedestrians represent an integral and pervasive aspect of the urban environment. Real estate, consumer patterns, public safety, and other aspects of city life are deeply intertwined with the variations of distribution of people across a city. However, current methods for estimating the distribution of people within a city tend to be expensive and mostly produce a sparse sampling of a few locations.

In this paper we propose a novel method to estimate the pedestrians distribution using street-level pictures of the city. We determine bounds for the

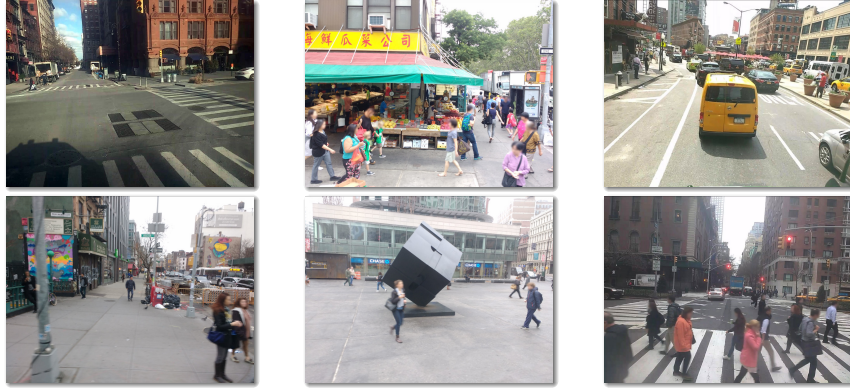


Figure 1: Sample of the images provided by Carmera. More than 10 million images have been used for the people density estimation evaluated in this work.

estimate and apply the methodology using a large and rich collection of street-level pictures from New York city to obtain a dense pedestrian map of the region.

We make use of state-of-art computer vision methods to identify and count pedestrians and we take into account the errors inherent to the detection process, we model it as a *probabilistic detection* and provide a closed form and bounds for the asymptotic error of the sampling process. We compare these formulas to numerical simulations of the sensing process. Our results suggests that computer vision produces usable data, despite the inherent noise. We utilized over 40 million street-level images provided by Carmera ¹. These images were obtained via their partnerships with high coverage fleets operating daily on city streets that traveled through the region of Manhattan Island in New York City over the course of a year. A deep-learning based algorithm for pedestrian detection was utilized in a case study to map the density of pedestrians in Manhattan.

The contributions of this paper can be summarized as

1. A new method for the analysis of the distribution of people using state of the art, but imperfect, computer vision algorithms.
2. A closed form function and bounds for the asymptotic error of the resulting person densities.
3. The results of simulations validating the sampling process and the derived asymptotic error.
4. A case study demonstrating the resulting densities for a collection of images from the City of New York.

¹<http://www.carmera.com>

1.1 Related Work

There are many ongoing efforts on the use of urban data to achieve citizen-centered improvements [69]. Governments and organizations in urban environments collect a vast amount of data daily [61] encompassing a large assortment of information including mobility, crime and pollution. The collection and use of this information has been attracting attention from the academics, governments and corporations [62]. Arietta et al. [3] explores the correlation of visual appearance of pictures and the attributes of the region it pertains. They collected images from Google Inc. [19] and also indicators from multiple regions and trained a model [7] to predict the indicator based on images. The city attributes include violent crime rates, theft rates, housing prices, population density and trees presence. Results show that the visual data can be efficiently used to predict the region attributes. Additionally, the regressor trained in one region showed reasonable results when tested in a different city. In a different problem, Zemene et al. [68] proposes a technique of determining the geo-location of the scene depicted in an image. Local and global features are extracted and used in a posterior features matching step. They report high accuracy and speed performances compared to existing approaches.

A citywide count of people has numerous applications for urban planners including the design of public transport network and of public spaces [66]. One approach is to have people scattered around the city manually counting the people nearby. However, this approach is laborious because it requires dedicated people to perform the measurements. Another possibility explored in Reades et al. [49] is to use cellphone use data to perform the person count. One clear limitation of this approach is that these data are not public and their coverage are restricted to the places where the carrier signal is present. Additionally, it is hard to know whether the cellphone signal is from someone in a building or in a car.

Alternatively, we can consider the visual task of finding people in city images. A remarkable work in this task consists in using the histogram of oriented gradients as the features vector and a support vector machines for the classification task [13]. In the context of deep neural networks [30, 56], the work of Ren et al. [50] introduced an approach that tries to solve this task by using a unified network that performs region proposal and classification. In this way, the method accepts annotations of multiple sized objects during the training step and during the testing stage, it performs classification of those objects in images of arbitrary sizes. In Dai et al. [12] the authors follow the two-stage region proposal and classification framework of Ren et al. [50] and proposes the Region-based Fully Convolutional Networks (R-FCN) which incorporate the idea of position-sensitive score maps to reduce the computational burden by sharing the per-RoI computation. Such speed alterations allow the incorporation of classification backbones such as Ren et al. [21]. Alternatively to the traditional person detection, head detection [17, 52] approaches have also been explored for efficient person count estimates. Given that information from the body is disregarded, however, it is natural to expect an inferior accuracy in

complex environments.

There are several city images repositories that contemplate people, some of them obtained using static cameras [63, 43] and others obtained using dynamic ones [18, 11, 38]. Such configuration of sensors arrangement have long been studied in the *sensor network* field [1, 2, 45] and an important aspect of these networks is whether the sensors are static or mobile. In Wang et al. [65] the authors explore the setting of a network composed of both static sensors and of mobile sensors. The holes in the coverage of the static sensors network are identified and the mobile sensors are used to cover the holes. A common problem in sensor networks is the *k-coverage problem* defined in Huang et al. [23], that aims to find the optimal setting of sensors such that any region is covered at least by k sensors. In Yang et al [67] the authors perform the task of counting people based on images obtained through a wireless network of static sensors.

Apart from controllable mobile sensors network, many works explore data collected from collaborative uncontrolled sensors [4] such as from vehicles GPS [54, 26], mobile phones sensors [53, 31, 48] and even from on-body sensors [10].

The work of Li et al. [35] considers the problem of using GPS data from a network of uncontrolled sensors to reconstruct the traffic in a city. They do that in two steps: initial traffic reconstruction and dynamic data completion. Such approach allowed the authors to get a complete traffic map and a 2D visualization of the traffic.

There are many ways to model the movement of mobile nodes in a sensor network, the so-called *mobility models* [8]. A simple one is the *random walk mobility model* [14] where at each instant in time each particles gets a direction and a speed to move. In the *random waypoint mobility model* [25], in turn, particles are given destinies and speeds. They travel toward their goal and once they get the destination a new goal and speed are given. The *Gauss-Markov mobility model* [36] attempts to eliminate abrupt stops and sharp turns present in the random waypoint mobility model. It is done by computing the current position based on the previous position, speed and direction.

Simulation of wireless sensor networks has long been studied [64, 34, 41] because it allows a complete analysis of system architectures by providing a controlled environment for the system [58]. The real-life systems non-determinism is simulated by the use of pseudo random number generators [28]. Among the large number of pseudo random number generators [46], a popular algorithm is the Mersenne Twister [39] due to its efficiency and robustness.

2 Data and proposed method

2.1 Data

The images utilized in this work come from a fleet of camera equipped cars (similar to Lee et al. [33]) traveling through Manhattan. Carmera provides a temporally and spatially dense collection of pictures. The volume of the sample utilized is at least an order of magnitude higher compared to the most popular

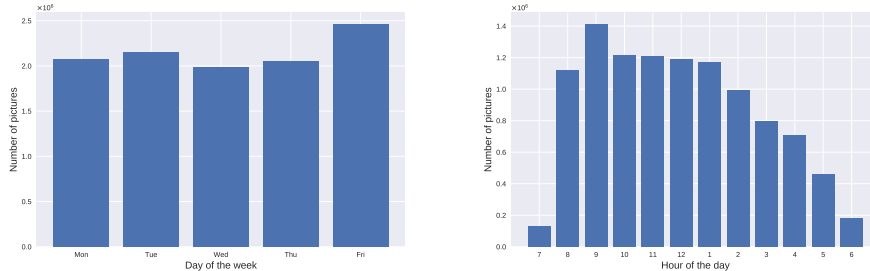


Figure 2: Temporal distribution of the collection of images by day of the week (left) and by hour of the day (right).

images datasets (see Table 1). It also differs from urban image databases [19, 63, 43, 18, 11, 38] by providing dense temporal coverage in addition to dense spatial coverage. The orientation of the cameras varies and the nature of the images are similar to street level collections provided by many mapping services. However, the images are not stitched into a 360 degree panorama. Every image is accompanied by metadata including the acquisition time, location, and camera orientation. The images are captured as the vehicle travels, with no control of the content, the illumination, the weather, the traffic conditions, or vehicular speed. The typical image depicts a urban scenario as a background and the city dynamics including person, vehicles and bicycles such as in Figure 1.

Table 1: Comparison of the numbers of images utilized to other popular datasets.

Datasets	Number of images	Resolution
CIFAR 10 [29]	$6.0 \cdot 10^3$	low
Pascal VOC 2007 [15]	$9.9 \cdot 10^3$	middle
Cityscapes [11]	$2.0 \cdot 10^4$	very high
MS COCO [37]	$3.3 \cdot 10^5$	high
ILSVRC 2017 [51]	$1.2 \cdot 10^6$	high
Our sample	$1.0 \cdot 10^7$	high

All images included in the sample have a resolution of 1280×960 . We used a sample of images captured from March 2016 to February 2017 containing 10,708,953 images. This sample presents a dense spatial sampling of the whole region of Manhattan (see Figure 3) over a year and irregular spatio-temporal sampling on a daily basis (see Figure 2) All resulting heatmaps are weighted sampling according to this distribution. Despite the large scale of the collection utilized in this work, it represents a small portion of the full dataset obtained by the company.

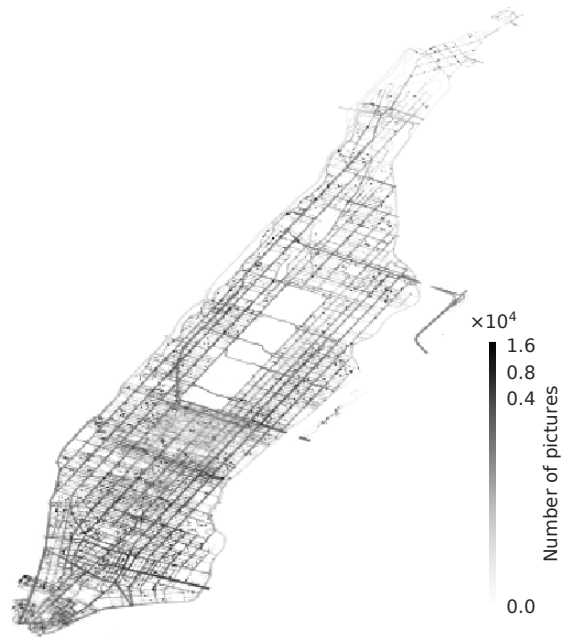


Figure 3: Geographical distribution of the number of pictures utilized in the region of Manhattan. The highest concentration of images are in the bottom and middle regions of map.



Figure 4: Hypothetical person detection. The person on the center was not identified by the detector and is a *false-negative*. The rightmost detection is a *false-positive*. Underlying image provided by Carmera.

2.2 Person detection

We adopt the same metric as Everingham et al. [15] when comparing the detected objects in an image to the ground-truth. A detected object is considered to correspond to a particular ground truth object if there is a minimum ratio of 50% between the overlap of the detected bounding boxes $B_{detected}$ ground-truth bounding boxes B_{gtruth} , and the union of the two areas (see Equation 1).

$$\frac{|B_{detected} \cap B_{gtruth}|}{|B_{detected} \cup B_{gtruth}|} \geq 0.5 \quad (1)$$

The recognition of distant objects in an image is difficult for humans and is even more difficult for computers. We assume that, on average, the size of a person within an image is an indicator of the distance that person to the sensor and try improve accuracy by considering a minimal size of the people detected. Thus, bounding boxes smaller than a parameter threshold are ignored. Another difficulty regards the differentiation among real people, cyclists and disabled people. The tested methods detect people even in the case they are sitting or riding.

We utilized the deep-learning based method R-FCN [12] over the entire dataset to obtain a database with the number of person detected in each image. This database is then aggregated in space and time to create a visualization of the person counts by finding the average number of person per image in each region.

2.3 Sensing model

As current person detection algorithms are far from perfect, it is natural to wonder about the accuracy of any person count resulting from their use. In this

section we provide a theoretical analysis of the effect of algorithmic errors on the final count.

In our model, we assume that the world is modeled by a number of small regions, or buckets, each of which we intend to measure a density. Sensors and people move around a world in some random fashion. At regular intervals, each sensor takes an independent measurement of the nearby person count and updates the recorded density at its current location, x . More formally, each time a sensor takes a sample, it obtains a measurement represented by the random variable $N(x)$. While we do not specify the distribution of $N(x)$, we assume that the expected value follows the formula

$$E[N_i(x)] = pn_i(x) + \lambda \quad (2)$$

Here $n_i(x)$ is the actual number of people in the location and time being sensed. p is a number giving the success rate of the vision algorithm and λ indicating its false positive rate.

The result of this process is the density of people at each location, $\psi(x)$.

$$\psi(x) = \frac{1}{k} \sum_i N_i(x) \quad (3)$$

For comparison, the ground truth density $\phi(x)$, defined respectively by (where k is the number of steps and samples),

$$\phi(x) = \frac{1}{k} \sum_i n_i(x) \quad (4)$$

We show in Appendix B, Equation 15 that the expected value of $\psi(x)$ is

$$E[\psi(x)] = p\phi(x) + \lambda \quad (5)$$

In other words, $\psi(x)$ is a biased estimator of $\phi(x)$. Unless our sensing algorithm precisely follows Equation 5, we are unable to transform this biased estimator into an unbiased one. Furthermore, even in the ideal case, p and λ may not be known. Instead, we directly utilize $\psi(x)$ and attempt to find a relative histogram. That is, we expect to get a number proportional to the density of the number of people at a location and not the actual density. As such, for any constant a , our density is equivalent to one scaled to $\psi'(x) = a\psi(x)$. Treating the distribution as a vector, we measure the direction but not the magnitude. In the terminology of group theory, our measurement suggests a density within the equivalent class:

$$[\psi] = \{a \in \mathbb{R}_+ \mid a\psi\} \quad (6)$$

To validate our measurement we need a metric that indicates how well the equivalent class compares to the ground truth distribution $\phi(x)$. To do that, we compare the ground truth to the unique closest element within the equivalent class. As a vector projection, this minimum element is (see Appendix A for a proof):

$$\psi' = \begin{cases} \psi \frac{\langle \psi, \phi \rangle}{|\psi|^2} & |\psi| \neq 0 \\ 0 & |\psi| = 0 \end{cases} \quad (7)$$

which we can then compare using the usual euclidean metric $|\psi' - \phi|$. However, this metric depends on the number of locations in the map, as well as the number of people. As such, we normalize the metric to between 0 and 1, to obtain a final metric:

$$\frac{|\psi' - \phi|}{|\psi'| + |\phi|} \quad (8)$$

In Appendix B, we find a bound additionally parameterized by $h > 0$ (the average density of people) and $c \geq 1$ (a function of the distribution of ϕ). Both of these parameters depend on the resolution of the heat map in addition to person distribution. In many cases c can not be determined, as such we can use the inequality in Equation 25 of Appendix B to conclude that over long periods of time:

$$\lim_{k \rightarrow \infty} \frac{|\psi' - \phi|}{|\psi'| + |\phi|} \leq \frac{\sqrt{c^2 - 1}\lambda}{2c^2hp} \leq \frac{1}{4} \frac{\lambda}{hp} \quad (9)$$

It is important to note that h represents the ground truth density of people, in the same units of ϕ . If only the sampled average density, $\hat{h}$, is know, the unbiased estimator of h , $\frac{\hat{h} - \lambda}{p}$ can be used. This leads to the bounds

$$\frac{1}{4} \frac{\lambda}{hp} \approx \frac{1}{4} \frac{\lambda}{\hat{h} - \lambda} \quad (10)$$

This final formula is only dependent on the false positive rate of the sensing algorithm and the average density of sensed objects measured by process, making it suitable for practical sensing applications. We wish to emphasize that this inequality is true whenever Equation 5 holds regardless of the underling probability distribution. This function is only useful when $\lambda \leq \bar{h}$ and in that domain, it is a monotonically increasing function of λ .

2.4 Simulation

The real-life acquisition process lacks some of the simplifications we used in our model. For example, samples taken in spatial and temporal proximity are correlated. To examine the performance of the sensing systems in the face of these non-ideal circumstances, we created a *discrete event simulation* [32] to compare sensed distributions to a known ground truths.

As illustrated in Figure 5, we simulated a number of mobile sensors that detect nearby particles. Sensors simulate the person detection event and their movement simulate the movement of the camera-equipped fleets responsible for the acquisition of data. Each sensor has a circular coverage of radius r . Collision among particles and sensors are ignored for simplicity. Sensors and particles

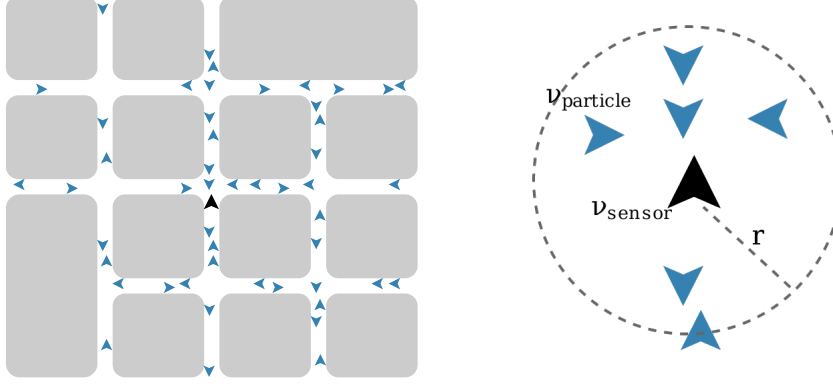


Figure 5: Left: An illustration of our simulation containing a sensor (center) moving through an environment with numerous person. Right: Each sensor moves with uniform speed ν_s and is able to sense people within a radius of r . Each sensing operation has a probability p of correctly detecting each person and, on expectation, finds λ false positives. People move with uniform speed ν_p .

move with uniform speeds ν_{sensor} and ν_{particle} respectively. The simulation world is mapped as a graph, as in Tian et al. [57]. Each node in the graph is a traversable point by both sensors and particles and edges represent a path between the end nodes.

We assume that, in each time step, sensor has an independent chance, p , of detecting each of the $n(x)$ person within range along with an independent chance per location to obtain a false positive. These assumptions lead to $N(x)$ being sampled from the sum of a binomial distribution with mean p and a Poisson process with a given expected number λ . A calculation of the expected value indicates that Equation 2 is satisfied and within the proposed bounds.

Algorithm 1: Mobility model of sensors and particles of the simulation.

```

Initialization(map, currentposition, destination);
while indefinitely do
    if destination =  $\emptyset$  then
        destination  $\leftarrow$  random(map)
        path  $\leftarrow$  A*(currentposition, destination, map)
    else
        currentposition  $\leftarrow$  pop(path);
    end
end

```

The system state can be described by various state variables: sensors and

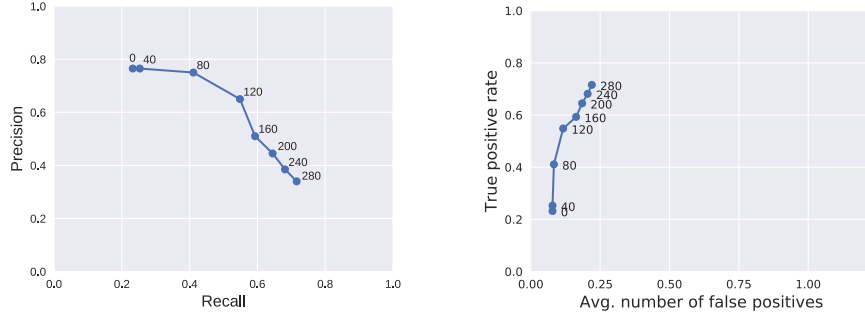


Figure 6: Evaluation of R-FCN [12] for different ground-truth height thresholds. The utilized model has a Resnet-101 backbone [21] trained on the Pascal VOC 2007 dataset [15].

particles positions, sensors and particles waypoints, real density of particles and sensed density of particles. Sensors and particles move with a variation of the random waypoint model [25], differing to it by the fact that sensors and particles are not allowed to change speeds; they have fixed speed given by the system parameters ν_{sensor} and $\nu_{particle}$. When a new destination is randomly picked, the trajectory on the map graph is computed using the A* algorithm [20] and the points of the trajectory are pushed to a heap (please refer to Algorithm 1).

As time progresses we obtain a 2D histogram for the sensed density as well as the ground truth density of particles. We are primarily interested in the difference between them, as given by the metric in Equation 8.

3 Experimental results

This section is divided into three parts: person detection and density map generation; calibration of the simulation; and simulation. The full source code is public available ².

3.1 Person detection results

We manually tagged 600 images to use as a ground truth. The R-FCN algorithm [12] was trained on the Pascal VOC 2007 dataset with the 101-layers neural network architecture proposed by He et al. [21]. We evaluated the method with detection scores ranging from 0.0 to 1.0, spaced by 0.1.

None of evaluations achieved recalls exceeding 80% (see Figure 6) and this fact is inherent to the difficulty of object detectors in detecting small objects as discussed in Section 2.1. To mitigate such issue, the detection model we propose assumes a finite radius of coverage (see Figure 5) and thus, we establish a limit on the size of the objects detected in the image. Figure 6 shows the results

²<https://github.com/VIDA-NYU/pedestrian-sensing-model>



Figure 7: A comparison of the ground truth and the measured person count from the 600 tagged test images. While the actual true positive and false positive counts do not match the expected statistics (left), the total measured person count can be close to approximated as linear (right). It should be noted that this is only an approximation as, even taking sampling errors into account, the mean measured count do not fit a linear model. Error bars are the 95% confidence interval of the mean, calculated by assuming the sampling process described in Section 2.3.

of the adopted detector over our sample as we vary the minimum acceptable height. As we can see, the higher the ground-truth height threshold, the higher the precision and specially the recall of the method.

We utilized as our detector the R-FCN method [12] with a residual network of 101 layers architecture [21] trained on Pascal VOC 2007 [15]. The model was trained using a weight decay of 0.0005 and a momentum of 0.9. Assuming a method minimum score of 0.7 and height threshold of 120 pixels, overall 7,474,623 people were detected.

As shown in Figure 7, the total number of people detected is approximately linear satisfying the requirements of the model proposed in Section 2.3, despite stastically significant deviations. These stem from the visions algorithm’s better than expected performance for images without any person and worse than expected performance for images with a single person. While we do not know how these deviations would effect the error bounds given in Equations 9 and 10, we hypothesize that the two deviations would cancel themselves out and bound may still approximately hold with a slightly larger equivalent λ .

A visualization of the density of people in entire Manhattan can be seen

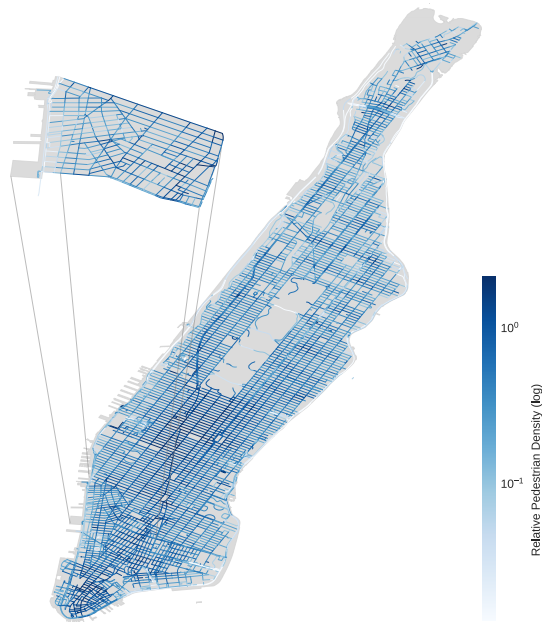


Figure 8: Visualization of the density of people in Manhattan. The scale of colors represent the relative density of people. The main figure depicts the heatmap over the island of Manhattan. The same heatmap is enlarged in the top-left of the figure to show the details of midtown and surrounding areas. *Underlying map data taken from OpenStreetMap [44]. Not drawn to scale.*

in Figure 8. For each analyzed picture, the number of people is determined using the detection algorithm and these counts are aggregated by streets. The topology of the streets were obtained from OpenStreetMaps [44]. The samples were organized in a KD-tree [6] and aggregated by street segments. To account for the imbalance of the number of images per segment, the average of the counts is considered.

Distribution of the number of people, like ours, can be useful for city planing, commercial, and other purposes. Taxis seeking riders, food trucks seeking customers, and businesses seeking storefronts all benefit from large crowds. However, traffic and self-driving cars do not. A knowledge of the distributions of the number of people can allow city planers, civil engineers, and traffic engineers to make better decisions.

Our person map can also show the effect that features of the city have on it's people. In addition to populated neighborhoods, subway stations, and attractions like the Metropolitan Museum of Art are all associated with a spike in the person densities. These spikes might be too localized to be detected using traditional methods. Further studies of vision based person counts may lead to a better understanding of the interplay between a city's environment and it's occupant's walking habits.

Special cases

A visual inspection of the detections showed that the methods are scale-sensitive, having difficulty to detect tiny-scale objects. However, when the objects have representative sizes, even in crowded scenes, the method proved to be robust. Another difficulty of the method regards the differentiation between pedestrians and non-pedestrians, such as cyclists and disabled people (see Figure 9). Such differentiation would require a novel computer vision method that considered the context of the scene. We performed a manual inspection over our annotated dataset and cyclists represented 2.4% of all annotated pedestrians. This low rate supports the argument that such limitation of the method does not invalidate the results statistically.

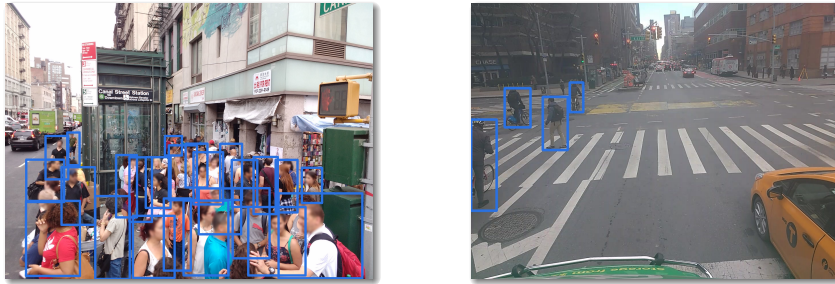


Figure 9: Person detection by the method in special cases. On the left, person detection in a crowded scene. On the right, pedestrians and non-pedestrians (a cyclist) are equally identified by the method.

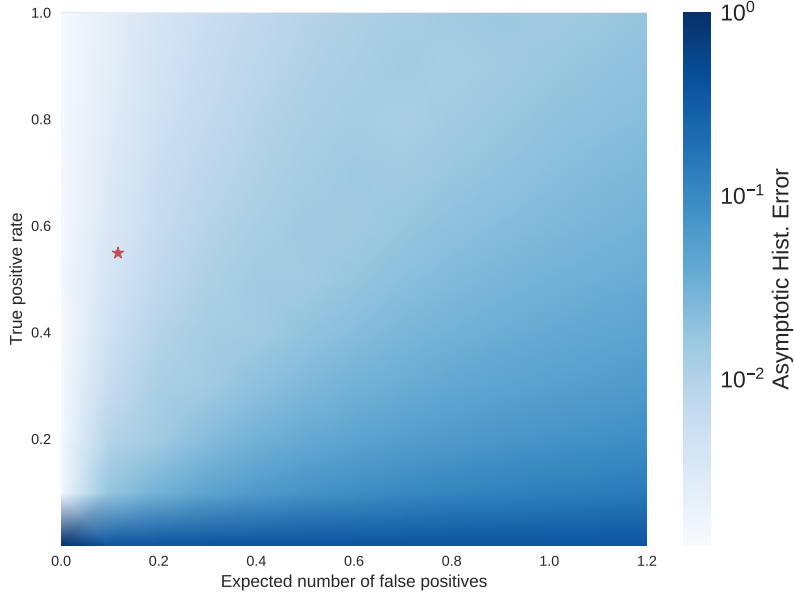


Figure 10: Asymptotic error metric between the sensed and ground truth histograms in our simulation, plotted as a function of the sensors true positive and expected number of false positives. The star in red depicts the results of the vision algorithm utilized with an asymptotic error of 0.062.

3.2 Simulation calibration

Simulation parameters have been calibrated according to the computer vision experiments. In Figure 10 we can see in red the expected number of false positives and the true positive rate of the computer vision method utilized. Given the average people density ($\hat{h}$) of 0.587 obtained in the previous section, following Equation 9, an error of 0.062 is obtained.

We compared the number of measured person counts as a function of the average number of ground truth people in each of the 600 manually labeled images to test the linear assumption used in Equation 5. Error bars for the mean were computed using the 5% to 95% values of the median of the appropriate sample process given in section 2.3. We measured the true positive rate (p) to be 0.54 and the average number of false positives (λ) to be of 0.117.

3.3 Simulation results

We evaluated different true positive rates and expected number of false positives of the sensors, p and λ and we used a Mersenne Twister pseudo number generator [39]. The various values for the parameters used in our experiments are listed in Table 2. For the 143 possibilities combination of values, we ran the

Table 2: Parameters of the simulation along with the values we used in the experiments.

Parameter	Symbol	Values
Number of people	N_{people}	50000
Person speed	ν_{person}	1
Sensor speed	ν_{sensor}	3
Number of sensors	$N_{sensors}$	10000
Sensor true positive rate	p	{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}
Sensor exp. number of false positives	λ	{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2}
Sensor range	r	1

simulation for 20,000 time steps 20 independent times.

The code is primarily implemented in Python with performance-sensitive sections implemented in Cython [5]. The average time to run a single experiment of this optimized code is of 11,718 seconds. Running the experiments on a single machine would take roughly one year of computation. By running them in parallel, it took 11 days.

For each experiment we examine the decay of the metric given by Equation 8 as a function of the cumulative number of samples captured by all the sensors. We assume the error continues to decay until it reaches an asymptotic minimum error within the 20,000 simulation time steps. Afterwards, we take the average decay curve of all 20 runs for each settings configuration and take the average of the last 200 values to find the asymptotic value.

We can visualize the results from the simulations in Figure 11 which shows how the variation of true positive rate (p) and number of false positives (λ) affect our histogram error. If we take a profile of say 0.2 of true positive rate ($p = 0.2$) we can see how the errors are greatly affected by the variation of the expected number of false positives, varying from very low to high error values (represented by the variation on the color saturation). We compare these values to our theoretical formulas (see Equation 9), and show they are approximately equal. Finally, We show that they are within the bound given by Equation 9 (red mesh).

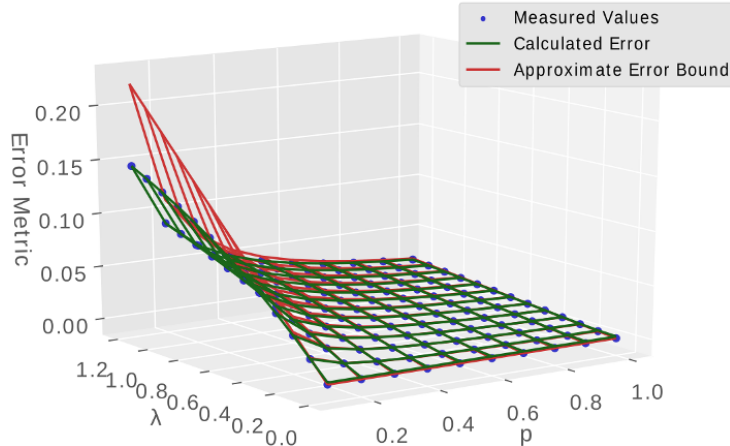


Figure 11: Comparison of the asymptotic histograms error measured from the simulation experiments (blue dots) to the theoretical close form function and the approximate bound (please refer to Equation 9), the red mesh.

4 Conclusion

In this project we applied computer vision techniques over a large and novel set of street-level images to obtain a pedestrian density map of the region of Manhattan. Inspired by the density estimation procedure we proposed a sensing model and established theoretical bounds for the density estimation error. Results give evidence that even when considering faulty sensors such as abstracted by our detection algorithms, reliable density maps can be obtained under our model assumptions.

Besides the results presented, there are other potential future avenues of studies as discussed next. First, we should caution that any application of our methodology should perform statistical tests to ensure that their results are statistically significant. While we set bounds on the asymptotic error after the sampling process converges, we have only provided case studies and heuristics for the time to convergence. It would be interesting to find a formal bound on time to convergence as well as provide guidelines for the appropriate statistical tests to validate the data post collection.

Our experiments could be extended to consider alternative mobility models [8], higher performance computer vision approaches, possibly semi-supervised ones [9, 60] as well as dynamics models including macroscopic ones [22, 24]. We could also use data completion algorithms [16, 35] or semi-supervised approaches [70] to reconstruct a city-wide people map.

The people distribution generated will then be able to be combined with other urban datasets such from Socrata [42], weather, crime rate, vandalism [59], census data, public transportation [55], bicycles and shadows [40]. We additionally aim to explore apparently disparate datasets such as from wind and from garbage collection.

Another future work is incorporation of advances such as from Microsoft Inc. [47] to visualize our images in the context of the city and use this visualization to gain additional insights into other datasets.

Additionally, we hope to use 3D pop-ups and/or photo based rendering to fully enhance the images in the three dimensional environments.

Acknowledgements

We thank Carmera for their collaboration. We also thank Harish Doraiswamy, Fabio Miranda, Alexandru Telea and Helio Pedrini for providing insights, comments, and suggestions that greatly contributed to this work. This work was supported in part by: NSF awards CNS-1229185, CCF-1533564, CNS-1544753, CNS-1730396, CNS-1828576; FAPESP (grants #14/24918-0 and #2015/22308-2); CNPq and CAPES; the Moore-Sloan Data Science Environment at NYU, and C2SMART. C. T. Silva is partially supported by the DARPA D3M program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

References

- [1] Ian F Akyildiz, Tommaso Melodia, and Kaushik R Chowdhury. A survey on wireless multimedia sensor networks. *Computer networks*, 51(4):921–960, 2007.
- [2] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. A survey on sensor networks. *IEEE Communications magazine*, 40(8):102–114, 2002.
- [3] Sean M Arietta, Alexei A Efros, Ravi Ramamoorthi, and Maneesh Agrawala. City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 20(12):2624–2633, 2014.
- [4] Stefano Basagni, Alessio Carosi, and Chiara Petrioli. Controlled vs. uncontrolled mobility in wireless sensor networks: Some performance insights. In *Vehicular Technology Conference, 2007*, pages 269–273, Maryland, USA, 2007. IEEE.

- [5] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn, and K. Smith. Cython: The best of both worlds. *Computing in Science Engineering*, 13(2):31–39, 2011.
- [6] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [7] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
- [8] Tracy Camp, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. *Wireless communications and mobile computing*, 2(5):483–502, 2002.
- [9] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [10] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Conference on Human Factors in Computing Systems (CHI)*, pages 1797–1806, Italy, 2008. ACM.
- [11] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, USA, 2016. IEEE.
- [12] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: Object detection via region-based fully convolutional networks. *arXiv preprint arXiv:1605.06409*, 2016.
- [13] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, USA, 2005. IEEE.
- [14] Vanessa Ann Davies et al. Evaluating mobility models within an ad hoc network. Master’s thesis, Citeseer, 2000.
- [15] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.
- [16] Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

- [17] Chenqiang Gao, Pei Li, Yajun Zhang, Jiang Liu, and Lan Wang. People counting based on head detection combining adaboost and cnn in crowded surveillance environment. *Neurocomputing*, 208:108–116, 2016.
- [18] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.
- [19] Google Inc. (<https://maps.google.com>), Last accessed March 2017.
- [20] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.
- [22] Dirk Helbing. A fluid dynamic model for the movement of pedestrians. *arXiv preprint cond-mat/9805213*, 1998.
- [23] Chi-Fu Huang and Yu-Chee Tseng. The coverage problem in a wireless sensor network. *Mobile Networks and Applications*, 10(4):519–528, 2005.
- [24] Tomoharu Iwata, Hitoshi Shimizu, Futoshi Naya, and Naonori Ueda. Estimating people flow from spatiotemporal population data via collective graphical mixture models. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 3(1):2, 2017.
- [25] David B Johnson and David A Maltz. Dynamic source routing in ad hoc wireless networks. *Mobile computing*, 353(1):153–181, 1996.
- [26] Sophia Karagiorgou, Dieter Pfoser, and Dimitrios Skoutas. A layered approach for more robust generation of road network maps from vehicle tracking data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 3(1):3, 2017.
- [27] Leonard Kleinrock. *Queueing systems, volume 2: Computer applications*, volume 66. John Wiley & Sons, USA, 1976.
- [28] Donald Ervin Knuth. *The art of computer programming*, volume 3. Addison-Wesley, USA, 1997.
- [29] Alex Krizhevsky et al. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.
- [30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, Nevada, USA, 2012.

- [31] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *IEEE Communications magazine*, 48(9):140–150, 2010.
- [32] Averill M Law, W David Kelton, and W David Kelton. *Simulation modeling and analysis*, volume 3. McGraw-Hill New York, Arizona, USA, 2007.
- [33] Uichin Lee, Eugenio Magistretti, Mario Gerla, Paolo Bellavista, and Antonio Corradi. Dissemination and harvesting of urban data using vehicular sensing platforms. *IEEE Transactions on Vehicular Technology*, 58(2):882–901, 2009.
- [34] Victor Lesser, Charles L Ortiz Jr, and Milind Tambe. *Distributed sensor networks: A multiagent perspective*, volume 9. Springer Science & Business Media, 2012.
- [35] Weizi Li, David Wolinski, and Ming C Lin. City-scale traffic animation using statistical learning and metamodel-based optimization. *ACM Transactions on Graphics (TOG)*, 36(6):200, 2017.
- [36] Ben Liang and Zygmunt J Haas. Predictive distance-based mobility management for pcs networks. In *INFOCOM’99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1377–1384, New York, USA, 1999. IEEE, IEEE.
- [37] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014*, pages 740–755. Springer, Zurich, Switzerland, 2014.
- [38] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017.
- [39] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.
- [40] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Luc Wilson, Mondrian Hsieh, and Claudio T Silva. Shadow accrual maps: Efficient accumulation of city-scale shadows over time. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [41] Muaz A Niazi and Amir Hussain. A novel agent-based simulation framework for sensing in complex adaptive environments. *IEEE Sensors Journal*, 11(2):404–412, 2011.
- [42] NYC open data. (<https://opendata.cityofnewyork.us/>), Last accessed March 2017.

- [43] Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, JK Aggarwal, Hyungtae Lee, Larry Davis, et al. A large-scale benchmark dataset for event recognition in surveillance video. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3153–3160, Colorado, USA, 2011. IEEE.
- [44] OpenStreetMap. Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>, 2017.
- [45] Mohd Fauzi Othman and Khairunnisa Shazali. Wireless sensor network applications: A study in environment monitoring system. *Procedia Engineering*, 41:1204–1210, 2012.
- [46] Stephen K. Park and Keith W. Miller. Random number generators: good ones are hard to find. *Communications of the ACM*, 31(10):1192–1201, 1988.
- [47] Photosynth. (<https://blogs.msdn.microsoft.com/photosynth/2017/02/06/microsoft-photosynth-has-been-shut-down/>), Last accessed March 2017.
- [48] Rajib Kumar Rana, Chun Tung Chou, Salil S Kanhere, Nirupama Bulusu, and Wen Hu. Ear-phone: an end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 105–116. ACM, 2010.
- [49] Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive computing*, 6(3):30–38, 2007.
- [50] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.
- [51] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [52] Muhammad Saqib, Sultan Daud Khan, Nabin Sharma, and Michael Blumenstein. Person head detection in multiple scales using deep convolutional neural networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018.
- [53] Xiang Sheng, Jian Tang, and Weiyi Zhang. Energy-efficient collaborative sensing with mobile phones. In *INFOCOM, 2012 Proceedings IEEE*, pages 1916–1924, Florida, USA, 2012. IEEE.

- [54] Wenhuan Shi, Shuhan Shen, and Yuncai Liu. Automatic generation of road network map from massive gps, vehicle trajectories. In *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pages 1–6, Missouri, USA, 2009. IEEE.
- [55] Claudio T. Silva, Juliana Freire, Fabio Miranda, Marcos Lage, Harish Doraiswamy, Maryam Hosseini, Eric K. Tokuda, Gabriel Ferreira, and Roberto M. Cesar Jr. Integrated analytics and visualization for multi-modality transportation data. Technical report, C2SMART Connected Cities for Smart Transportation, 2019.
- [56] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, et al. Going deeper with convolutions. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, Massachusetts, USA, 2015.
- [57] Jing Tian, Jorg Hahner, Christian Becker, Illya Stepanov, and Kurt Rothermel. Graph-based mobility model for mobile ad hoc network simulation. In *Simulation Symposium, 2002. Proceedings. 35th Annual*, pages 337–344, California, USA, 2002. IEEE, IEEE.
- [58] Ben L Titzer, Daniel K Lee, and Jens Palsberg. Avroa: Scalable sensor network simulation with precise timing. In *International Symposium on Information Processing in Sensor Networks*, pages 477–482, Tennessee, USA, 2005. IEEE.
- [59] Eric K. Tokuda, Roberto M. Cesar, and Claudio T. Silva. Quantifying the presence of graffiti in urban environments. In *International Conference on Big Data and Smart Computing (BigComp)*, pages 1–4, Japan, 2019. IEEE.
- [60] Eric K. Tokuda, Gabriel B. A. Ferreira, Claudio Silva, and Roberto M. Cesar-Jr. A novel semi-supervised detection approach with weak annotation. In *Image Analysis and Interpretation (SSIAI), 2018 IEEE Southwest Symposium on*, Nevada, USA, 2018. IEEE.
- [61] United States Environment Protection Agency Air Data USEEPAAD. (https://www3.epa.gov/airdata/ad_data_daily.html), Last accessed March 2017.
- [62] Carlos A Vanegas, Daniel G Aliaga, and Bedrich Benes. Automatic extraction of manhattan-world building masses from 3d laser range scans. *Visualization and Computer Graphics, IEEE Transactions on*, 18(10):1627–1637, 2012.
- [63] Roberto Vezzani and Rita Cucchiara. Video surveillance online repository (visor): an integrated framework. *Multimedia Tools and Applications*, 50(2):359–380, 2010.

- [64] Meritxell Vinyals, Juan A Rodriguez-Aguilar, and Jesus Cerquides. A survey on sensor networks from a multiagent perspective. *The Computer Journal*, 54(3):455–470, 2011.
- [65] Guiling Wang, Guohong Cao, and Tom LaPorta. A bidding protocol for deploying mobile sensors. In *International Conference on Network Protocols*, pages 315–324, Georgia, USA, 2003. IEEE.
- [66] William H Whyte. *City: Rediscovering the center*. University of Pennsylvania Press, Pennsylvania, USA, 2012.
- [67] Danny B Yang, Leonidas J Guibas, et al. Counting people in crowds with a real-time network of simple image sensors. In *International Conference on Computer Vision Workshops*, page 122, Nice, France, 2003. IEEE.
- [68] Eyasu Zemene, Yonatan Tariku Tesfaye, Haroon Idrees, Andrea Prati, Marcello Pelillo, and Mubarak Shah. Large-scale image geo-localization using dominant sets. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):148–161, 2018.
- [69] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38, 2014.
- [70] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2005.

A Proof of metric formula

Here we derive the formula for the closest point in our class and the ground truth vector. This is equivalent to solving:

$$\min_a |a\psi - \phi|^2 \quad (11)$$

First we expand the distance metric using the euclidean inner-product:

$$|a\psi - \phi|^2 = \langle a\psi - \phi, a\psi - \phi \rangle \quad (12)$$

$$= a^2 \langle \psi, \psi \rangle - 2a \langle \psi, \phi \rangle + \langle \phi, \phi \rangle \quad (13)$$

which is minimized by

$$a = \frac{\langle \psi, \phi \rangle}{\langle \psi, \psi \rangle} = \frac{\langle \psi, \phi \rangle}{|\psi|^2}$$

when $|\psi|^2 \neq 0$. For $|\psi|^2 = 0$, then $\psi = 0$ and

$$\min_a |a\psi - \phi|^2 = |\phi|^2 \quad (14)$$

regardless the value of a .

Substituting this value into a , we obtain our equation above:

$$\psi' = \begin{cases} \psi \frac{\langle \psi, \phi \rangle}{|\psi|^2} & |\phi| \neq 0 \\ 0 & |\phi| = 0 \end{cases}$$

When $\langle \psi, \phi \rangle = 0$ then $\psi' = 0$ and we get the same value.

B Theoretical Asymptotic Error

In this paper we will derive the equation for sensor error that we give in Equation 9. The error bounds only assume that $E[\psi(x)] = p\phi(x) + \lambda$ for some values of p and λ . First, we will show that, for the simulation, the values of p and λ agree with the parameters of the same name.

As noted in Equation 3, the sampling process for the simulation results in the following sampled values for each location:

$$\psi(x) = \frac{1}{k} \sum (T(n(x)) + F)$$

where $T(n(x))$ is a sampled from a binomial distribution with mean $n(x)$ and F is a Poisson process [27] with a mean of λ .

At the same time, the ground truth distribution of people at each location is given by Equation 4

$$\phi(x) = \frac{1}{s} \sum n(x)$$

In this appendix we will make the simplifying assumption that $s = k$

The executed value of the sampled can then be found by (noting that the random variables are all independent)

$$\begin{aligned} E[\psi(x)] &= \frac{1}{s} \sum (E[T(n(x))] + E[F]) \\ &= \frac{1}{s} \sum (pn(x) + \lambda) \\ &= p \frac{1}{s} \sum (n(x)) + \lambda \\ &= p\phi(x) + \lambda \end{aligned} \quad (15)$$

From this point on, all results will only depend on the equations $E[\psi(x)] = p\phi(x) + \lambda$ and not the underlying sampling process.

Let m be the total number of people and r be the sampling location. In a real world scenario, m may not be well defined. As such, we will work in terms of $h = \frac{m}{r}$, the density of people.

We want to calculate $\frac{|\psi' - \phi|}{|\psi' + \phi|}$ and we are going to compute each term separately.

By the law of large numbers, in the limit of $s \rightarrow \infty$, ψ approaches

$$\psi(x) \rightarrow E[\psi(x)] = \phi(x)p + \lambda$$

Using this limit, we can find the asymptotic value of ψ' , as defined by Equation 7, can be found:

$$\begin{aligned} \psi' &= \psi \frac{\langle \psi, \phi \rangle}{|\psi|^2} \\ &\rightarrow (p\phi + \lambda) \frac{\langle p\phi + \lambda, \phi \rangle}{|p\phi + \lambda|^2} \\ &= (p\phi + \lambda) \frac{p|\phi|^2 + \lambda \langle 1, \phi \rangle}{p^2|\phi|^2 + 2p\lambda \langle 1, \phi \rangle + \lambda^2 \langle 1, 1 \rangle} \\ &= (p\phi + \lambda) \frac{p|\phi|^2 + m\lambda}{p^2|\phi|^2 + 2mp\lambda + r\lambda^2} \\ &= \left(\phi + \frac{\lambda}{p} \right) \frac{p^2|\phi|^2 + mp\lambda}{p^2|\phi|^2 + 2mp\lambda + r\lambda^2} \end{aligned} \tag{16}$$

Here, $\langle \cdot, \cdot \rangle$ is the Euclidean inner product and 1 is the vector with all ones. Note that $\langle 1, \phi \rangle = m$ and $\langle 1, 1 \rangle = r$

The magnitude of ψ' can then be found by

$$\begin{aligned} |\psi'| &= \frac{p^2|\phi|^2 + mp\lambda}{p^2|\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{\left\langle \phi, \frac{\lambda}{p} 1 \right\rangle} \\ &= \frac{p^2|\phi|^2 + mp\lambda}{p^2|\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{|\phi|^2 + \frac{r\lambda^2}{p^2} + 2\frac{m\lambda}{p}} \\ &= \frac{p|\phi|^2 + m\lambda}{p^2|\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{p^2|\phi|^2 + r\lambda^2 + 2mp\lambda} \end{aligned} \tag{17}$$

Similarly, the difference between ψ' and the ground truth sampling can be found by

$$\begin{aligned}
\psi' - \phi &= \left(\phi + \frac{\lambda}{p} \right) \frac{p^2 |\phi|^2 + mp\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} - \phi \\
&= \left(\phi + \frac{\lambda}{p} \right) \frac{p^2 |\phi|^2 + mp\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} - \phi \frac{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \\
&= \phi \frac{p^2 |\phi|^2 + mp\lambda - p^2 |\phi|^2 - 2mp\lambda - r\lambda^2}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} + \frac{\lambda}{p} \frac{p^2 |\phi|^2 + mp\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \\
&= -\phi \frac{mp\lambda + r\lambda^2}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} + \frac{\lambda}{p} \frac{p^2 |\phi|^2 + mp\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \\
&= \frac{\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \left(-\phi (mp + r\lambda) + p |\phi|^2 + m\lambda \right)
\end{aligned} \tag{18}$$

The magnitude of which can be found by

$$\begin{aligned}
|\psi' - \phi| &= \frac{\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{\left\langle -\phi (mp + r\lambda), \left(p |\phi|^2 + m\lambda \right) 1 \right\rangle} \\
&= \frac{\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{|\phi|^2 (mp + r\lambda)^2 + \left(p |\phi|^2 + m\lambda \right)^2 r - 2 (mp + r\lambda) \left(p |\phi|^2 + m\lambda \right) m}
\end{aligned} \tag{19}$$

Equations 17 and 19 can be used to find our metric as defined by Equation

8

$$\begin{aligned}
\frac{|\psi' - \phi|}{|\psi'| + |\phi|} &= \frac{\frac{\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{|\phi|^2 (mp + r\lambda)^2 + \left(p |\phi|^2 + m\lambda \right)^2 r - 2 (mp + r\lambda) \left(p |\phi|^2 + m\lambda \right) m}}{\frac{p |\phi|^2 + m\lambda}{p^2 |\phi|^2 + 2mp\lambda + r\lambda^2} \sqrt{p^2 |\phi|^2 + r\lambda^2 + 2mp\lambda} + |\phi|} \\
&= \frac{\lambda \sqrt{|\phi|^2 (mp + r\lambda)^2 + \left(p |\phi|^2 + m\lambda \right)^2 r - 2 (mp + r\lambda) \left(p |\phi|^2 + m\lambda \right) m}}{\left(p |\phi|^2 + m\lambda \right) \sqrt{p^2 |\phi|^2 + r\lambda^2 + 2mp\lambda} + \left(p^2 |\phi|^2 + 2mp\lambda + r\lambda^2 \right) |\phi|}
\end{aligned} \tag{20}$$

However, this equation depends on $|\phi|$, m , and r which would not be known for real applications. To account for these variables, we will introduce a new parameter, c :

$$c = |\phi| \frac{\sqrt{r}}{m} \tag{21}$$

While c may also be unknown, we will be able to take a maximum of the resulting error function to get a bound. Using the bounds of L2-norm in terms of the L1-norm and noting that the L1-norm is equal to m , we obtain the identity

$$1 \leq c \leq \sqrt{r} \quad (22)$$

By substituting, $|\phi| = c \frac{m}{\sqrt{r}}$ and a bit of algebra, we can transform Equation 20 into a final form:

$$\begin{aligned} \frac{|\psi' - \phi|}{|\psi'| + |\phi|} &= \frac{\lambda \sqrt{|\phi|^2 (mp + r\lambda)^2 + \left(p|\phi|^2 + m\lambda\right)^2} r - 2(mp + r\lambda) \left(p|\phi|^2 + m\lambda\right) m}{\left(p|\phi|^2 + m\lambda\right) \sqrt{p^2 |\phi|^2 + r\lambda^2 + 2mp\lambda} + \left(p^2 |\phi|^2 + 2mp\lambda + r\lambda^2\right) |\phi|} \\ &= \frac{\lambda \sqrt{c^2 \frac{m^2}{r} (mp + r\lambda)^2 + \left(pc^2 \frac{m^2}{r} + m\lambda\right)^2} r - 2(mp + r\lambda) \left(pc^2 \frac{m^2}{r} + m\lambda\right) m}{\left(pc^2 \frac{m^2}{r} + m\lambda\right) \sqrt{p^2 c^2 \frac{m^2}{r} + r\lambda^2 + 2mp\lambda} + \left(p^2 c^2 \frac{m^2}{r} + 2mp\lambda + r\lambda^2\right) c \frac{m}{\sqrt{r}}} \\ &= \frac{\frac{1}{r\sqrt{r}} \lambda \sqrt{c^2 \frac{m^2}{r} (mp + r\lambda)^2 + \left(pc^2 \frac{m^2}{r} + m\lambda\right)^2} r - 2(mp + r\lambda) \left(pc^2 \frac{m^2}{r} + m\lambda\right) m}{\frac{1}{r\sqrt{r}} \left(\left(pc^2 \frac{m^2}{r} + m\lambda\right) \sqrt{p^2 c^2 \frac{m^2}{r} + r\lambda^2 + 2mp\lambda} + \left(p^2 c^2 \frac{m^2}{r} + 2mp\lambda + r\lambda^2\right) c \frac{m}{\sqrt{r}} \right)} \\ &= \frac{\lambda \sqrt{c^2 \frac{m^2}{r^2} \left(\frac{m}{r}p + \lambda\right)^2 + \left(pc^2 \frac{m^2}{r^2} + \frac{m}{r}\lambda\right)^2} - 2\left(\frac{m}{r}p + \lambda\right) \left(pc^2 \frac{m^2}{r^2} + \frac{m}{r}\lambda\right) \frac{m}{r}}{\left(pc^2 \frac{m^2}{r^2} + \frac{m}{r}\lambda\right) \sqrt{p^2 c^2 \frac{m^2}{r^2} + \lambda^2 + 2\frac{m}{r}p\lambda} + \left(p^2 c^2 \frac{m^2}{r^2} + 2\frac{m}{r}p\lambda + \lambda^2\right) c \frac{m}{r}} \\ &= \frac{\lambda \sqrt{c^2 h^2 (hp + \lambda)^2 + (pc^2 h^2 + h\lambda)^2} - 2(hp + \lambda) (pc^2 h^2 + h\lambda) h}{(pc^2 h^2 + h\lambda) \sqrt{p^2 c^2 h^2 + \lambda^2 + 2hp\lambda} + (p^2 c^2 h^2 + 2hp\lambda + \lambda^2) ch} \\ &= \frac{\frac{1}{h^2} \lambda \sqrt{c^2 h^2 (hp + \lambda)^2 + (pc^2 h^2 + h\lambda)^2} - 2(hp + \lambda) (pc^2 h^2 + h\lambda) h}{\frac{1}{h^2} \left((pc^2 h^2 + h\lambda) \sqrt{p^2 c^2 h^2 + \lambda^2 + 2hp\lambda} + (p^2 c^2 h^2 + 2hp\lambda + \lambda^2) ch \right)} \\ &= \frac{\lambda}{h} \frac{\sqrt{c^2 \left(p + \frac{\lambda}{h}\right)^2 + \left(pc^2 + \frac{\lambda}{h}\right)^2} - 2\left(p + \frac{\lambda}{h}\right) \left(pc^2 + \frac{\lambda}{h}\right)}{\left(pc^2 + \frac{\lambda}{h}\right) \sqrt{p^2 c^2 + \left(\frac{\lambda}{h}\right)^2} + 2p\frac{\lambda}{h} + \left(p^2 c^2 + 2p\frac{\lambda}{h} + \left(\frac{\lambda}{h}\right)^2\right) c} \quad (23) \end{aligned}$$

Note that this formula is a function of $\frac{\lambda}{h}$, p , and c . By expanding this function as a Taylor series, we find

$$\frac{|\psi' - \phi|}{|\psi'| + |\phi|} = \frac{\sqrt{c^2 - 1}\lambda}{2c^2 hp} - \frac{\sqrt{c^2 - 1}\lambda^2}{2c^4 h^2 p^2} + \dots \quad (24)$$

Noting that $\frac{\sqrt{c^2 - 1}\lambda^2}{2c^4 h^2 p^2}$ is always positive and applying Taylor's theorem, we end up with the inequality

$$\frac{|\psi' - \phi|}{|\psi'| + |\phi|} \leq \frac{\sqrt{c^2 - 1}\lambda}{2c^2 hp} \leq \frac{1}{4} \frac{\lambda}{hp} \quad (25)$$

Where the last step comes from the inequality

$$\frac{\sqrt{c^2 - 1}}{c^2} \leq \frac{1}{2}$$