Indoor Crowd Density Estimation for Efficient Monitoring System

Phong Nguyen and Takayuki Akiyama

Center of Technology Innovation
Hitachi, Ltd. Research & Development Group, Tokyo, Japan

phong.nguyen.kj@hitachi.com; takayuki.akiyama.hv@hitachi.com

Abstract. We propose a system to estimate crowd density in an indoor environment with an efficient number of sensors. There is no conventional system capable of identifying where to concentrate sensors for measurement. Our system uses pre-collected data to analyze crowd density correlation coefficients and identify representative areas for placing sensors. In addition, our system uses multiple linear regression models to estimate crowd density in other areas where no sensor placed, using the data in representative areas and total number of people in a venue. We tested our system in Tokyo International Forum, a 5000 m2 exhibition center. Our system identified 3 representative areas for placing sensors among a total of 56 areas in the exhibition hall. The average estimation error rates for the areas without sensors were 1.46 to 1.49 people per minute, which are sufficiently accurate for visualizing people density on a heat map. Our system thus can operate with efficient operating and maintenance cost..

Keywords: Crowd Density Estimation, Correlation, Grouping, Linear Regression.

1 Introduction

The estimation of crowd density is widely used in various applications such as safety monitoring, traffic control, and smart guiding to ensure a pleasant experience for passengers. In open space indoor environments like airports, shopping malls, or train stations, a monitoring system, using heat maps for visualizing the crowd density, can help the authorities to keep track of the flow of passengers and identify congested areas in order reorganize the distribution of crowds. In order to calculate the crowd density at a specific area, most current solutions are using measuring sensors such as security cameras (for visual analysis), ultra-sound sensors, and laser sensors (for positioning analysis). However, each sensor can only measure a certain limited area, so many sensors are required for full coverage of an open space venue. Moreover, each sensor is associated with a recursive operating and maintenance cost. To reduce these costs, a system to estimate crowd density of multiple areas in the whole indoor environment with an efficient number of sensors is desired. Existing work estimates crowd density in an area using features of that area such as census data to estimate population density of a city.

But these data are not suitable to estimate a dynamic area in an indoor environment due to the difference in estimation resolution and the explanatory data are static. Other work estimates crowd density of an area by using the past data of the same area. Nevertheless, there is no work proposing a way to identify suitable areas for placing sensors and use the crowd density data from the measured areas to estimate crowd density in other areas.

Therefore, we propose a system capable of identifying where to place sensors in an indoor environment and use the data from the sensors to estimate the crowd density in the whole venue. Our system exploits supervised learning technique to reduce number of sensors. We have a data collection step to capture the crowd density in different areas in different time, and use those correlations in a grouping technique to identify representative areas for concentrating the placement of sensors. We then deploy linear regression models constructed from the collected data to estimate crowd density in other areas and create a heat map of the indoor environment.

To test our system, an experiment in Tokyo International Forum, a 5000 m2 exhibition center, was performed. We used 56 laser sensors to measure crowd density in the whole venue for a day, then reduced the number of sensors to 5 for another day to estimate the crowd density of the whole venue. Each sensor could measure the crowd density in an area of 10 m × 10 m. The average estimation error rate was 1.46 to 1.49 people per minute. Such results are sufficient to visualize the crowd density in a heat map accurately. Our result has shown that our system can work well in indoor environments where crowd density in different areas are highly correlated, e.g. exhibition center, museum, aquarium, etc.

This paper is a summarized version of another more detailed paper, which can be accessed at: http://bit.ly/ipin2019-nguyenp [13]

2 Proposed system

In this section, we describe in detail how our system works. We use a supervised learning method, which includes a step for data collection and a step for building models and deploying them. In data collection step, we use crowd density measuring sensors such as security cameras and laser sensors for a full coverage monitoring in a targeted venue for a day. We use the data in the collection step to calculate correlations of crowd density among all areas in the venue. Based on the correlation value, we group areas and shortlist representative areas for keeping the measuring sensors. We then construct linear regression models for other areas by least square method, using the data in the representative areas and total number of people in the venue as the explanatory variables. In the deployment step, we use the data collected from representative areas to estimate the crowd density of other areas. We have a hypothesis that the linear regression models built by the data collected previously are still accurate for a certain time. In our experiment section, we will explain our validation on our models by using them to estimate the crowd density of the whole venue based on data from the previous day.

The details in each step of our system are explained in these following steps:

Step 1. Data collection and correlation coefficients calculation.

Step 2. Areas grouping and representative areas identification.

Step 3. Estimation with linear regression models.

2.1 Data collection and correlation coefficients calculation

In order to select representative areas for placing sensors, we need to collect data of crowd density in the whole area for a certain time.

The whole venue is sliced and divided into a grid, where each cell is monitored by a crowd density measuring sensor. The size of each cell in the grid is decided by the measuring capability of a sensor. For example, a security camera can monitor the area with the size of 5 m × 5 m, or a laser sensor can capture the location of people in a 10 m × 10 m area, therefore, the size of a cell can be decided based on type of sensors and their specification. Each cell monitored by a sensor is referred to by the grid reference. On a two-dimensional map, a cell is identified by its row and column number. In this paper, we number the rows from top to bottom, and the columns from left to right as illustrated in Fig. 1.

During the data collection step, we collect data of crowd density measured by each sensor in each cell continuously. For each time interval, we count how many people in average inside a cell. Therefore, a data vector is used to store the average number of people in a cell:

$$D_{(i,j)}[\, d_{t1}, d_{t2}, d_{t3}, \dots , d_{tn}]$$

where:

- $t1, t2, \dots, tn$ is the sampling time;
- $D_{(i,j)}$ is a vector of crowd density in cell at row i and column j.
- n is the total number of data points collected;
- d_t is the number of people inside cell (i,j) at time t.

After the data is collected for a certain time, we calculate correlation coefficients of crowd density of every pair of cells. The crowd density correlation coefficient between a pair of cells is calculated according to Pearson's correlation [10] formula:

$$r_{(i,j)(k,e)} = \frac{n(\sum D_{(i,j)} D_{(k,e)}) - (\sum D_{(i,j)})(\sum D_{(k,e)})}{\sqrt{[n\sum D_{(i,j)}{}^2 - (\sum D_{(i,j)})^2][n\sum D_{(k,e)}{}^2 - (\sum D_{(k,e)})^2]}} \tag{1}$$

where:

- n is the total number of data point collected in a cell.
- $D_{(i,j)}$ is a vector of crowd density in cell at row i and column j.
- $D_{(k,e)}$ is a vector of crowd density in cell at row e and column k.
- $r_{(i,j)(k,e)}$ is the correlation coefficient between $D_{(i,j)}$ and $D_{(k,e)}$

The correlation coefficients measure the linear relationship between two cells, giving a value between -1 and +1 inclusive, where 1 is a perfect positive linear relationship, 0 is no relationship, and -1 is a perfect negative linear relationship. The correlation coefficients between a pair of cells has commutative property (e.g. correlation coefficient of D(1,2) and D(3,4) is the same with correlation coefficient of D(3,4) and D(1,2)), so we can remove the redundant calculations when we calculate the correlation coeffcients of every pair of cells. The results of calculating correlation coefficients can be stored in the upper half of a correlation matrix or correlation coefficient table.

2.2 Areas grouping and representative areas identification

We want to identify the strength in relationship among pairs of cells, therefore we convert all the correlation coefficients into their absolute value. The highly correlated pairs have the absolute correlation coefficients close to +1. By using a correlation coefficient threshold γ from 0.5 to 1, we can filter out those pairs whose relationships are weak. For example, if we choose the threshold γ to be 0.7, all pairs with absolute value of correlation coefficients lower than 0.7 will be removed from the table. The correlation coefficient table after filtering should only have the pairs with strong linear relationship. We call the table after filtering "highly correlated areas table".

In this table, a cell is linked with another cell with a high correlation coefficients. By iterating through all the links, we put linked cells into the same group. Two groups of cells will be distinguished if there is absolutely no link between any pair of cells belonging to each group. After the iteration, m groups can be identified, with $0 \leq m \leq N/2$, where N is the total number of cells of the grid. Number of groups m depends on the value of threshold γ, as γ is higher (closer to 1), m is smaller because more links are filtered out the highly correlated areas table.

In each group, a representative cell with the highest sum of the absolute value of correlation coefficients in highly correlated areas table is selected. The representative cells are the most suitable areas for concentrating the placement of sensors, and use the data collected in those cells to estimate other cells of the grid.

After grouping cells in the highly correlated areas table, there might be some other cells not belonging to any group because they were filtered out. Such cells are put into their respective groups by comparing their correlation coefficients with the representative cells. And cells which does not have any crowd density (e.g. due to blockage, security, etc), will have the default value of 0 crowd density.

The pseudocode of this step is provided in Fig. 1.

2.3 Estimation with linear regression models

The representative cells are where the crowd density measuring sensors are placed in order to collect data for explanatory variables in our regression models. Another explanatory variable in the linear regression model is the total number of people in the venue. To calculate the total number of people, measuring sensors are set up at entrances

```
Input: cell1, cell2 are 2 vectors storing the name of cells under tuple type (i,j) in highly correlated ar-
eas table, where cell1[i] and cell2[i] have correlation coefficients higher than γ.
Initialize count = 0, group = dictionary((i,j),0) with (i,j) is a tuple in cell1 and cell2, groupName = 1
  for i = 1..len(cell1) do:
    if group[cell1[i]] == 0 and group[cell2[i]] ==0 then:
        group[cell1[i]] = groupName;
        group[cell2[i]] = groupName;
          groupName += 1;
  elif group[cell1[i]] != 0 and group[cell2[i]] ==0 then:
        group[cell2[i]] = group[cell1[i]] ;
  elif group[cell1[i]] == 0 and group[cell2[i]] !=0 then:
        group[cell1[i]] = group[cell2[i]] ;
  elif group[cell1[i]] != 0 and group[cell2[i]] !=0 then:
      if group[cell1[i]] == group[cell2[i]] then: next;
      elif group[cell1[i]] > group[cell2[i]] then:
       group[ which group.value()==group[cell2[i]] ]=group[cell1[i]];
          next;
      elif group[cell1[i]] < group[cell2[i]] then:
       group[which group.value()==group[cell1[i]]]=group[cell2[i]];
         next;
```

Fig. 1. Pseudocode for putting cells into different group

and exits of the venue in order to count how many people enter and how many people leave the venue.

In the other words, our linear regression models use total number of people in the venue and the crowd density in a representative cell to estimate crowd density in other cells of the grid. The expression of the linear regression model is explained as follows:

$$D_{(i,j)}(t) = \alpha + \beta_1 D_{(c,f)}(t) + \beta_2 P(t) \tag{2}$$

where:

- $D_{(i,j)}$ is a vector of crowd density in cell at row i and column j, which is in the same group with representative cell at row c and column f.
- $D_{(c,f)}$ is a vector of crowd density in representative cell at row c and column f.
- t is the time step measured.
- $P(t)$ is the total number of people in the venue in time t.
- β_1 and β_2 are the coefficients of a regression model for each estimated cell.
- α is a constant default value of crowd density in a regression model for each esimated cell.

Each cell has its own model, using crowd density data in the representative cells and total number of people in the venue as the input. A heat map can be visualized after crowd densities in all cells are estimated to help the authorities identify congested areas and monitor the distribution of crowd in an indoor environment.

3 Experimental Setup

To validate our method, we set up an experiment to collect crowd density data in an exhibition center. We chose Hitachi Innovation Forum (HIF) event, a large exhibition show, as the experimental field for evaluating our proposed method. The event was held in Tokyo International Forum [11]. Table I summarizes the detailed specifications of the experiment.

According to size of the field, we divided it into a grid with 72 cells (6 rows x 12 columns), with each cell size is 10 m × 10m. However, we only placed laser sensors in 56 cells because 16 cells were impassable to visitors. Each laser sensor can capture the number of visitors and exhibitors within the size of a cell. We counted the average number of people in each cell in every minute.

We collected data of the exhibition event for 2 days as we used the first day for identifying the representative cells and building the linear regression models, and the second day for testing and validating our proposed method.

Table 1. Experimental settings

VENUE	TOKYO INTERNATIONAL FORUM
DATE	30^{th} – 31^{st} October 2014
TIME	9:00am – 6.30pm
POSITIONING TECHNOLOGY	Laser Sensors
SENSOR MAKER	SICK [12]
NUMBER OF SENSORS	56
VENUE SIZE	5000 m^2

4 Results and Evaluations

We calculated the correlation matrix among all the cells in the first day of the exhibition, 30th October 2014. The correlation matrix included the correlation coefficients of all the pairs of cells.

We use threshold $\gamma = 0.7$ to filter out cell pairs which have absolute value of correlation coefficients smaller than 0.7. Only 16 pairs of cells are identified in the highly correlated cells table.

After that, we put highly correlated cells into groups so that in one group all the cells were linked with each other by their correlation coefficients. According to the data in the first day (30th October 2014), we could identify three groups: group 1 included

(2,3), (2,4), (3,4) and (4,6); group 2 included (2,6), (3,5), and (3,6); group 3 included (2,8), (2,9), (3,7), (3,8), (3,9) and (4,8). In each group, the cells which have highest sum of absolute value of correlation coefficients are (3,4), (3,6), and (4,8). We chose those 3 cells as the representative cells for keeping the sensors, and removed other sensors from all other cells. The remaining cells were put into the three groups by comparing the correlation coefficients with the representative cells.

Based on the data collected on 30th October 2014, we constructed a linear regression model for each cell based on formula (2), excluding the representative cells. We evaluated our estimation by 10-fold-cross-validation (10-fold CV) technique on 30th October and used the models training on whole data on 30th October to estimate the crowd density on 31st October with only using data from representative grids. The average error rates are 1.49 people per minute on 30th October and 1.48 people per minute with new data on 31st October. Fig. 2 shows the example of the estimation results for (3,7) in each day.

According to the histogram, most of the average errors were less than 1 people. In addition, 95% of average errors were less than 2.5 people.

80% of maximum errors were less than 15 people. There is an extreme case where the error has crossed over 80 people.

We also validated our method by using data on 31st October as training data and data on 30th October as test data. The results had average error rates of 1.47 people per minute for 10-fold-cross-validation on 31st October, and 1.46 people per minute for the test data of 30th October. 90% of average errors were also less than 2.5 people. In addition, 75% of maximum errors were below 5 people. These results indicated that our proposed method was reliable for deploying in exhibition center with robust estimation.

5 Conclusion

This paper has presented a system for crowd density estimation using efficient number of sensors in indoor environment. Our system uses a supervised learning approach. We used pre-collected data to analyze correlation coefficients of crowd density among all areas in the venue. Based on the analysis results, we identified representative areas for concentrating the placement of sensors. Using the data collected from representative areas, we could estimate the crowd density of other areas in the whole venue using linear regression models.

To validate our system, we have trailed it in a 5000 m2 exhibition center. After collecting data for a day, we identified 3 areas that have high correlation coefficients with other areas. We used them as representative areas and we placed a sensor for each area. In addition, we placed 2 additional sensors at the entrance and exit to count the total number of people in the venue. Using the data collected from representative areas, we estimated the crowd density of other areas with average error rates ranging from 1.46 people per minute to 1.49 people per minute. These error rates are tolerable for visualizing crowd density of the whole exhibition center on a heat map and useful for monitoring the distribution of crowds in the venue. In other words, our system only used 5 sensors instead of 56 sensors for visualizing crowd density in the exhibition center, therefore theoretically reducing the cost of operating and maintaining the infrastructure.

In the future, we plan to validate our system in various indoor environments such as airports, train stations, shopping malls, etc. The number of identified representative areas may depend on the characteristics of the indoor environment. We also plan to validate our system in a longer period in order to quantify the effect of seasonal drift on our system.

References

[1] McKenna, S., Gong, S., and Raja Y. 1998. Modelling Facial Colour and Identity with Gaussian Mixtures. Pattern Recognition, vol.31, no. 12, pp. 1883-1892.

[2] Zhang D., Ma, J., Chen, Q. and Lionel M. N. 2007. An RF-based System for Tracking Transceiver-free Objects. IEEE PerCom.

[3] Haralick, R. M. Statistical and Structural Approaches to exture. Proceedings of the IEEE, vol. 67, no. 5, pp. 786-804, 1979.

[4] Saleh, S. A. M., Suandi, S. A., and Ibrahim, H., Recent survey on crowd density estimation and counting for visual surveillance. 2015. Engineering Applications of Artificial Intelligence, vol. 41, pp. 103-114. DOI= doi:10.1016/j.engappai.2015.01.007

[5] Yuanm Y., Qiu, C., Xi, W., and Zhao, J. 2011. Crowd Density Estimation Using Wireless Sensor Networks. Seventh International Conference on Mobile Ad-hoc and Sensor Networks.

[6] Liu, X. H., Kyriakidiss, P. C., and Goodchilds, M. F. 2008. Population-density estimation using regression and area-to-point residual kriging. International Journal of Geographical Information Science, Vol. 22, Nos. 4-5, 431-447.

[7] Rdenas, L., Wang, L., Rogerson, P., Wu, C., Feng, T., and Kamphaus, B. D. 2010. Assessing fine-spatial-resolution remote sensing for small-area population estimation. International Journal of Remote Sensing, Vol. 31, No. 21, 5605–5634.

[8] Lu, Z., Im, J., Quackenbush, L., and Halligan, K. 2010. Population estimation based on multisensory data fusion. International Journal of Remote Sensing,31:21, 5587-5604, DOI= 10.1080/01431161.2010.496801

[9] Felkel, R., and Klann, D. 2012. Comprehensive passenger flow management at Frankfurt Airport, Airport Management, Vol. 6, No. 2, 107–124.

[10] Chen, P. Y., and Popovich, P. M. 2002. Correlation. pp. 10-26. Thousand Oaks, CA: SAGE Publications, Inc. DOI= http://dx.doi.org/10.4135/9781412983808

[11] Tokyo International Forum, (online), available from <http://www.t-i-forum.co.jp/en/>, (accessed in June 2016)

[12] SICK, (online), available from http://www.sick.com/, (accessed in June 2016)

[13] Nguyen, P.., Akiyama, T., Indoor Crowd Density Estimation for Efficient Monitoring System, availbale from http://bit.ly/ipin2019-nguyenp (accessed in July 2019)