ATC pedestrain tracking

Yan Lu, Mengzhuo Lu, Dingyu Yao

Electrical engineering

Columbia University

yl3406@columbia.edu, ml3806@columbia.edu, dy2307@columbia.edu

Abstract—**In this project, we analyzed the 3-D range sensor dataset of Asian Pacific Trade Center, a shopping center in Osaka, Japan. By processing the dataset, we analyzed the flow rate in the shopping center using Pig, and did data visualization to compare the flow rate under different circumstances. We also did clustering by Spark and data visualization for each property in the 3-D range sensor dataset, and dug the underlying relations between them as well as got many interesting founds. We also trained a Mahout classifier to distinguish if a pedestrian walk in a group or individually. Based on these results and explored information, we found some invisible fact concealed in boring pedestrian flow data and come up with useful suggestions and strategies.**

Keywords- Big Data, Pedestrian Tracking, Mall, Sale, Flyer

I. INTRODUCTION

Asian Pacific Trade Center (ATC), located in Osaka Japan, is the largest international mall complex in Kansai. In 2013, a Japanese research team set up a tracking environment in the shopping center. The system consists of multiple 3D range sensors, covering an area of about 900 m2. The top view of the shopping mall is illustrated below.

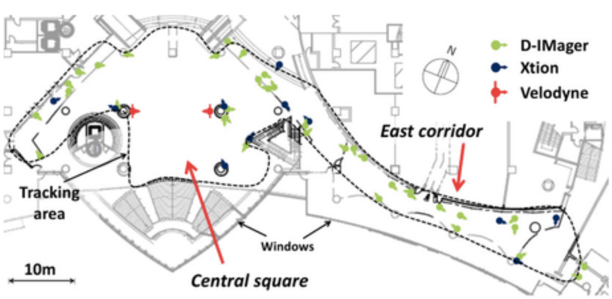


Figure 1 Top view of the shopping mall

The data recorded by these 3D range sensors was made public by the team for research purposes. Our project is based on these datasets, and intends to distinguish target client and come up with store deploy strategy for ATC shopping mall and propose strategies for robot flyer senders and security. We analyzed popular time and area in the shopping mall; clustered pedestrians based on their coordinates distribution, and classified them based on their group behavior.

II. RELATED WORKS

In 2012, a Japanese research team set up 3-D range sensor in ATC shopping mall, they recorded the data collected from 9:40-20:20 throughout the year, and collected in total 92 days data [1]. This dataset has been explored in many ways. Zanlungo, Brscic and Kanda researched in characteristics in group pedestrians, from which we can apply our result on classifying pedestrians into group shoppers and individual shoppers [2]. This study found group shoppers and individual shoppers different in terms of their facing angle and angle of motion, which we can use and applied to classify shoppers.

Kanda and Mascot proposed an idea of using robots for flyer sending in the shopping mall, which our project can help [3]. Kidokoro and Shiomi researched in people's behavior when encounter a robot in a mall area, which is done by simulation [4]. A similar topic was introduced by Hagita, who talked about the potential influence of using robots to do public tasks in a city environment [5]. In our project we will propose suggestions and strategies to make the robot flyer sending process more effective.

III. SYSTEM OVERVIEW

The datasets we used in this study are "ATC pedestrian tracking dataset" and "Pedestrian tracking with group annotations" which were obtained by JST/CREST in japan [1]. The project primary purposed in enabling mobile social robots to work in public spaces.

The dataset was collected between October 24, 2012 and November 29, 2013, Wednesday and Sunday, 9:40-20:20. The dataset consists of 92 days in total. The data of each day is provided as CSV files, which each row in a CSV file corresponds to a single tracked person at a single instant, and it contains the following fields:

time [ms] (unixtime + milliseconds/1000), person id, position x [mm], position y [mm], position z (height) [mm], velocity [mm/s], angle of motion [deg], facing angle [deg]

```
1366502566.671,9022500,-12731,2036,1442.000,1136.300,2.781,2.739
1366502566.671,9022402,-15561,2159,1467.000,1348.700,2.833,2.777
1366502566.671,9022400,-10500,3106,1576.000,1177.000,3.080,3.092
1366502566.671,9022305,13567,-8081,1544.000,1987.900,2.382,2.407
1366502566.671,9022303,-9395,2032,1500.000,1250.700,-2.958,-3.017
1366502566.671,9022301,-8253,6490,1378.000,954.000,2.630,2.661
1366502566.671,9022300,-5569,2883,1155.000,1445.500,2.884,2.866
```

Figure 2 Dataset of Single Pedestrian

For group data set:

PEDESTRIAN_ID GROUP_SIZE PARTNER_ID_1 ... (list of ids of all other pedestrians in group) NUMBER_OF_INTERACTING_PARTNERS INTERACTION_PARTNER_ID_1 ... (list of all socially interacting partners)

```
10170700 2 10170400 0
10170400 2 10170700 0
10244601 3 10251300 10251101 2 10251300 10251101
10251300 3 10244601 10251101 2 10244601 10251101
10251101 3 10244601 10251300 2 10244601 10251300
10245500 2 10245600 0
10245600 2 10245500 0
10260200 3 10255300 10260900 2 10255300 10260900
10255300 3 10260200 10260900 2 10260200 10260900
10260900 3 10260200 10255300 2 10260200 10255300
```

Figure 3 Dataset of Group Pedestrians

Note, in the two datasets above, the pedestrian ID is the same. Which means if an ID is shown in group dataset, the pedestrian with same ID in single pedestrian dataset should also in the dataset.

Because of the time limitation, we considered 42 days (21 Wednesdays and 21 Sundays) with over 123.57GB of data, and we think this should be enough for the analysis.

IV. ALGORITHM

We applied some advanced tools and algorithms we learnt from lecture including using Pig in Hadoop to filter, sort and count data; using Spark to do k-mean clustering and using Mahout to classify data and make predictions. We also applied some other tools to virtualize and process data including Python, Matlab, NetworkX and iNetSoft.

Our analysis can be separated into 3 parts. The first part is flow rate analysis, in which we used Pig to filter and process raw data. For instance, since the original dataset sort rows with time, we used Pig to resort them by person's ids and grouped data about a same person. After analysis, we used iNetSoft to plot our results.

The second part is distribution analysis and visualization. We firstly used python to pre-processed the dataset and divide the monitoring area into grids. We also calculated many arguments, such as people intensity using python. We then used MATLAB to plot in total 42 days of data focusing on different characteristics. Pig performed filtering of the data. We used Spark to do clustering work and clustered the data based on different categories (for example x, y coordinates and velocity).

The third part involves group behavior analysis and visualization. We firstly used python to pre-process the dataset, selecting those pedestrians appeared in the group dataset area. Then appended another column to the end of each row indicating whether they are involved in a group (this information comes from group dataset). We then used Mahout logistic regression algorithm to train and test models, using datasets from different dates. Then we used NetworkX to plot the group data, and used Fruchterman-Reingold force-directed algorithm to layout the plot.

V. EXPERIMENT RESULTS

i. Pedestrians Group Behaviors Analysis and Classification

1. Background

ATC shopping mall has thousands of visitors each day, and they all have different behaviors. We want to distinguish those behaviors, so we can classify different shoppers, and use the data for various applications.

As we all know, a major behavior difference among shoppers are whether they shop single or with group. Those two kinds of shoppers tend to buy different categories and different amounts of products, and their tolerance of interruptions are also different (for example, interruption can involve being asked to fill in a questionnaire). Therefore, classify shoppers by whether they are in-group or not will be beneficial.

From researches we learned that single shoppers and group shoppers differ in terms of walking angle and looking (facing) angle [6]. These information also appears in our datasets collected by 3D range sensors in ATC shopping mall. Figure 4 shows us the relations between facing and walking angles of shoppers in a group of two.

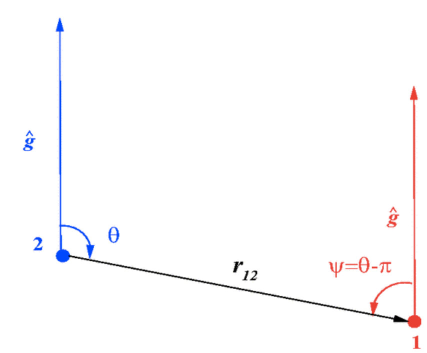
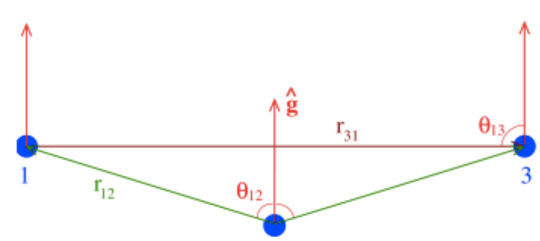


Figure 4 Angles for a two pedestrians [6]

From the figure we can see that in a two pedestrians group, the two shoppers walk in the same direction while keeping each



other in his/her eyesight reachable areas. Same rules are followed by three pedestrians' systems as well, which is illustrated in figure 5.

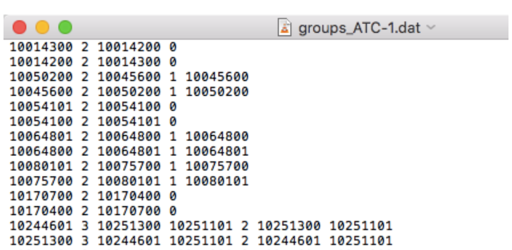
Based on these information, Aanlungo, etc. [2] concludes that that 2 pedestrians, identified as i and j, are socially interacting while walking towards a common goal, given by a vector $\bar{g}$. If we write their relative positions in polar coordinates (r, θ), we can write their relative positions as $r_{ij} = r_i - r_j$. Then angle between this vector and vector $\bar{g}$ can therefore be written as:

$$\psi = \begin{cases} \theta - \pi \ \ if\ \theta > 0, \\ \theta + \pi\ if\ \theta \leq 0, \end{cases} \quad (1)$$

As a group, pedestrians will want their common goal in their vision field, while they also expect their partners in their vision field [7]. Moreover, they would also want themselves being in their partner's vision field [8]. Single shoppers, however, don't have such constrains in their walking and facing angles. Using these information, we can classify pedestrian's behavior using their walking and facing angle data provided by 3D sensors in ATC shopping mall.

2. Analysis of group and person datasets

Since we have found the underline pattern and rules of the relations between pedestrians' behavior and their waking and facing angle data, we would like to analysis the datasets and give a classifier for future usage. Researchers from the 3D range sensor team manually denoted group information of pedestrians in part of the shopping center (approximately in the west most one third of the corridor, close to the central hall) to part of datasets, and this part of data is also open for download [9] [2] . This information corresponds with the person's dataset, as introduced before, which means the ids in the group dataset mean the same person in the person's dataset of that specific date. The example of the group dataset was shown in figure 6.


groups_ATC-1.dat

```
10014300 2 10014200 0
10014200 2 10014300 0
10050200 2 10045600 1 10045600
10045600 2 10050200 1 10050200
10054101 2 10054100 0
10054100 2 10054101 0
10064801 2 10064800 1 10064800
10064800 2 10064801 1 10064801
10080101 2 10075700 1 10075700
10075700 2 10080101 1 10080101
10170700 2 10170400 0
10170400 2 10170700 0
10244601 3 10251300 10251101 2 10251300 10251101
10251300 3 10244601 10251101 2 10244601 10251101
10251101 3 10244601 10251300 2 10244601 10251300
```

Figure 6 Group dataset example

As shown in the figure, the group dataset goes by following rules: the first column lists the id of the currently interested person, and the second column shows has/her group size. The flowing columns will be the ids of other people in this group, until another column of a single integer number appears. The rest information in this row is not important for our purpose. They denote who in the group are the interested person are current interacting with.

Note people's ids in this dataset correspond to the person's dataset. Therefore, we used Python to combine the two

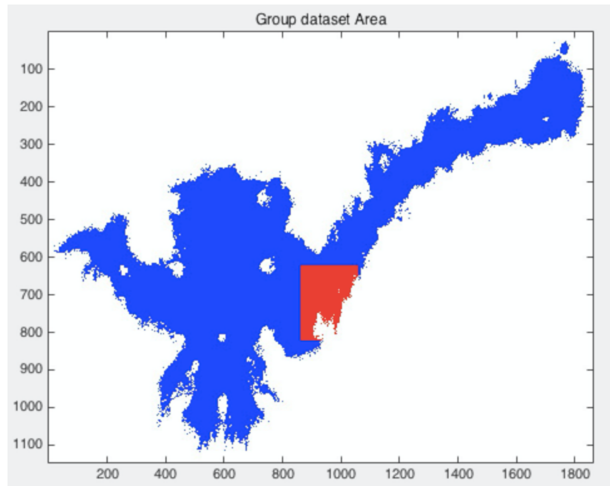


Figure 7 Inspecting area for group data

datasets together, by appending another column to the end of each row of person's dataset, denoting whether this person involves in a group or not. This extra column will be useful in Mahout classification, which will be introduced later.

Also, since the group dataset only involved groups in part of the area in the shopping mall (which is actually area of $x \in [-1,9999]$ and $y \in [-1,9999]$), as illustrated in Figure 7. We filtered the person's dataset with Python to make sure it only contains people in this area, which ensures the accuracy of the classification. As shown by the figure below.

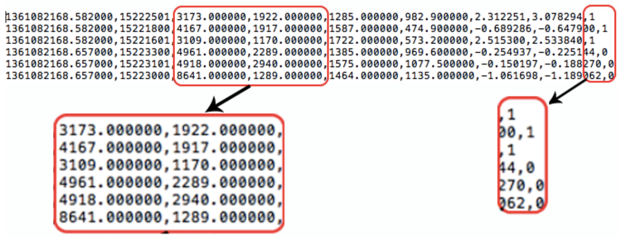


Figure 8 Python pre-processed data, with location filtered and group indicator added

We firstly analyzed the pre-processed datasets using NetworkX, a big-data visualization tool. We plotted individual and group datasets on random-selected Wednesdays and Sundays. The plot was layout using Fruchterman-Reingold force-directed algorithm. So larger groups will be in the middle, while pairs and unconnected shoppers are on the peripheral. The details can be found in the figures below.

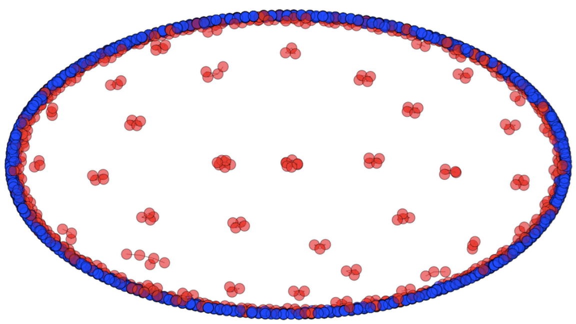


Figure 9 NetworkX visualization of Wednesday data

In the figures, red nodes represent those people who involved in groups, while blue nodes corresponds to people whom walk on their own. After analysis we found the proportion of group shoppers are around 30%.

It can be found that among all of the groups, most of the groups have 2 or 3 people. Those nodes on the peripheral are pairs (2-people groups) and individual pedestrians; we can refer to this local figure below for details:

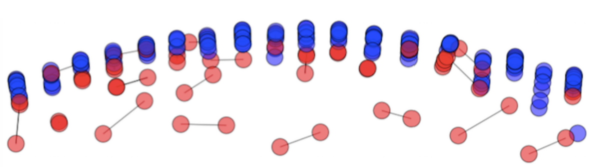


Figure 10 Part of Figure 5

Same attempt was made on Sunday datasets as well. From the figure below we can find there are significantly more groups than Wednesday.

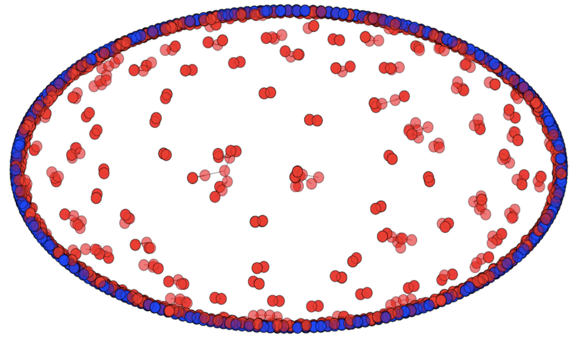


Figure 11 NetworkX visualization of Sunday data

Afte r analyzing, we found the proportion of group pedestrians to individual pedestrians on Sunday is around 52%, which is much larger than Wednesday. Also there tends to be more large-groups, but most of the groups are still off two to three people.

3. Classification by Mahout

To develop a classifier to distinguish groups, we used Mahout to do classification. We firstly used the Python pre-processed data for training. W used in total 970MB data for training, and tested in total 683MB data. We set the column ‘group’ as the classification target, and tried various predictors, including facing angle, waking angle, velocity, x, y coordinates and differnet combinations of them, and finally came up with a best performing classifier.

This classifier uses two predictors, walking angle and facing angle, both of numeric type. The target variable has 2 categories, indicating if the user involves in a group or not. The model trained is as follow:

$$\begin{aligned} group \sim\ & -12.874 \times Intercept\ Term \\ & + 0.953 \times walking\ angle \\ & + 0.791 \times facing\ angle \end{aligned} \tag{2}$$

This model was trained with following parameters: the size of internal feature vector is 20, the number of passes over input is 100, and the initial learning rate is 50. The average AUC on the test datasets is around 0.7.

Therefore, using this model we can classify pedestrians and know if they are walking in groups or not. This information is useful in many ways, which will be introduced in the following chapter.

3. Applications

(1) Recommendations on ATC shopping center after group data analysis and visualization

Based on the datasets analysis in previous chapters, we can make many recommendations for the restaurants and shops in ATC shopping mall. For instance, since group shopper’s proportion is much larger on weekends then it's on weekdays, the restaurants in ATC shopping center can set more single tables on weekdays and more couple or larger tables on weekends.

Moreover, promotions of lover and family related products (for example flowers, household appliances and children's toys) could also be done more on weekends then weekdays to make more profits.

(2) Applications of group classifier

Individual and group shopper's behavior is apparently different in many ways. And defining groups will be helpful in taking advantage of these differences. For instance, a Japanese research group is trying to use robots to send flyers [10]. Group shoppers tend to be easier to accept flyer's from robots while individual shoppers tends to avoid them [4]. Furthermore, by sending flyers to a shopper group we can make the flyer sending process more effective. Therefore, by automatically knowing which person is in groups, we can make the robot flyers sending more effective.

(2) The classifier can also be used for criminal detection. For example, if policeman was inspecting a group crime case, he can use this classifier to quickly filter out all of the individual pedestrians and it is easier for his to track the walking path of suspects.

ii. Distribution Virtualization

1. Pedestrian distribution analysis

We know from SYSTEM OVERVIEW, that the datasets we used contains x, y coordinates of each pedestrian. Using this information, we did visualization for each day in the dataset and plotted the pedestrian distribution. We found many interesting patterns in these plots.

Firstly, it is obvious that the people intensity on Sundays is much higher than on Wednesdays. As shown in figure below (in which red denotes high intensity and blue means

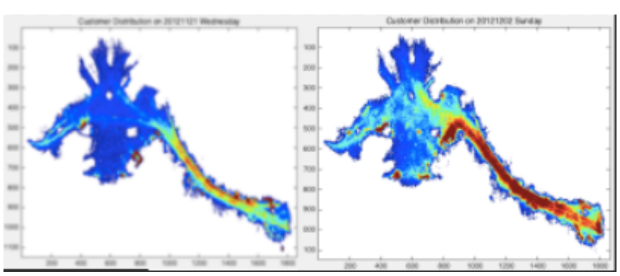


Figure 12 People intensity on Wednesday (left) and Sunday (right)

low intensity).

Moreover, we found that on Wednesdays, more people tend to walk in an invisible path, because there are clearly 3 paths in the figure that can be found with large intensity. This path is marked black in the figure 5. This is probably because most people on weekdays are just walking by, and they tend choose the shortest route (invisible path). Therefore, for robots giving flyers as introduced in last chapter, they can stay on these paths and handing out flyers not directly related to in-mall information (because this people will not be very likely to shopping). However, on Sundays, as shown in the bottom picture of figure 5, we can see that people are more randomly distributed compared to Wednesday, so more people will be actually shopping then just walking by. Then sending flyers about in-mall information (e.g. sale news) will be more effective. And an ideal place for the robots to stay, in this section, is in the black-circled area in the figure below. This area has high people intensity and is not likely to cause congestion

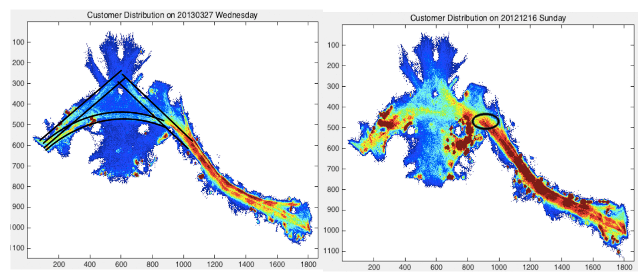


Figure 13 Invisible Path & Sale Area

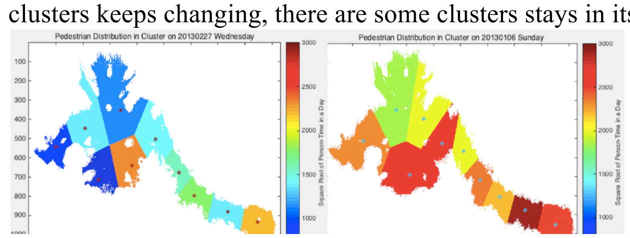
(because there are spaces on the sides).

There is also one interesting pattern we found: in the narrow corridor in right-down site of the plot, there tend to be more people on the upper side then down side on Wednesdays, while on Sundays the intensity are basically the same. This is because on the upper side there are mainly shops, but on the down side there are mainly shop goods display. This means shops are more attractive then goods display. Therefore, to balance pedestrians on both sides, we suggest the shopping mall to move some of the shops from upper side to down side, or set some new shops on the down side.

2. Clustering on pedestrian intensity using Spark

To analysis popular areas in the mall, we conducted clustering on each day's dataset using Spark. We use K-mean algorithm to do the clustering, and each clustering went through 10 iterations. We divide all the pedestrians in the mall into 10 clusters and assigned color to them based on the intensity. Some examples are shown in the figure 14. It can be clearly seen that Sunday has larger pedestrian intensity then Wednesday, and most of the clusters lies in east corridor.

There is one interesting thing worth noticing. Although the clusters keeps changing, there are some clusters stays in its



place and continues to have low intensity unless special events happens (details will come up in later chapters). We suggest the shopping mall put effort into these areas, such as try to move some popular restaurants or shops there to balance the pedestrian intensity.

Knowing the locations of cluster centers, we can derive many applications. For instance, security services can stay around each center of clustering, so that it distances to all of people in the cluster will remain controllable. Moreover, this information is also useful for promotions. By putting important promotions (including robot flyer senders) on the cluster center of clusters with high intensity will help improving the promotion effect. Of course, this strategy will base on the principal of not producing congestion.

3. Pedestrians' speeds analysis and clustering

Based on pedestrians' speeds and their x, y coordinates, we plotted pedestrians' speeds comparison between randomly selected Wednesdays and Sundays. As shown in the figure below.

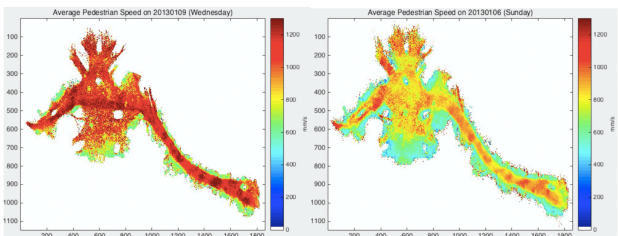
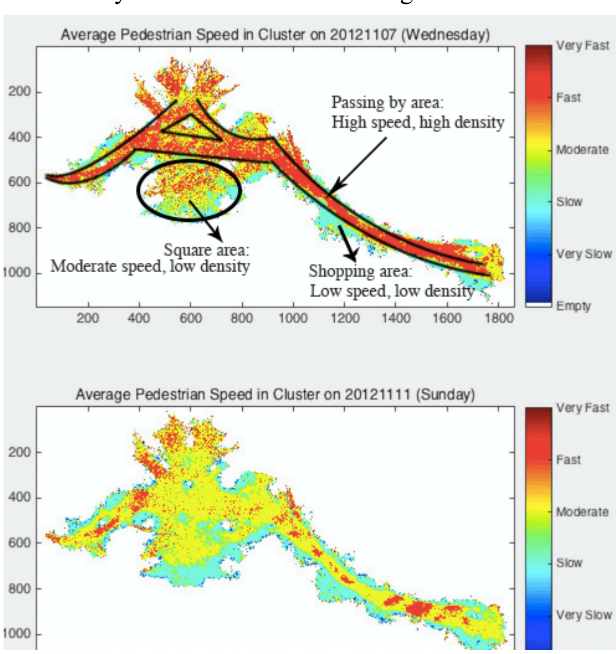


Figure 15 Pedestrians' speeds plotting

However, as we can see from the figure 15, the speed is linearly related to the color, so it is difficult to distinguish patterns in the figures. Therefore, we did clustering on speeds using Spark and plotted the clustering results by people's x, y coordinates. The results are shown in the figure 16 below, where the upper subfigure refers to Wednesday's dataset and lower subfigure refers to Sunday's.



Based on the clustering results, we split the Wednesday shopping mall into 3 areas: passing by area, shopping area and square area.

From the figure we can see that the average speed on Wednesdays is much higher than it on Sundays. Besides, the high-speed area in Wednesdays is larger then that in Sundays. This could because on Wednesdays there are fewer people, so pedestrians can walk faster and have more space to walk. On the other hand, on Sundays the shopping mall is rather crowded and people cannot walk as fast as they intended. And the paths spotted earlier for pedestrians just passing by has higher speed. So trying to block them in these paths is not a wise choice for promotions. If we want to send flyers to people on the path (which is tempting because there are more people there), we should stand on the shopping areas besides the waking path so that we will not slow the pedestrian flow.

4. Pedestrian Walking Angle Analysis

The last two part provide us with distribution information from pure location. In this the next 2 section, we will take velocity, angle of motion and facing angle into consideration and do advanced data mining.

In this part, we filter out people who are walking by their velocity. If the velocity is larger than 0.5m/s, we mark them as walking. And among all walking people, we consider their angle of motion and draw the average angle of motion for people in every grid as shown in figure 17.

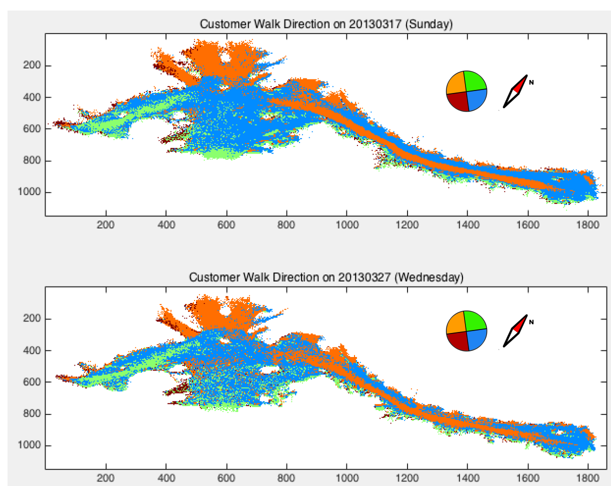


Figure 17 Pedestrian Walk Direction

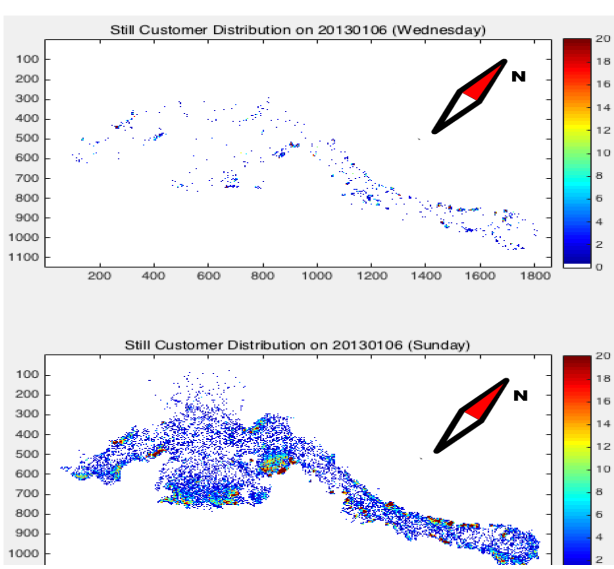
As a result, most people are well behaved, which means they always walk on left in the narrow areas like east corridor. The interesting thing is, on Sunday, although the population of pedestrian increased significantly, people tend to walk in a more consistent direction. This is probably because people are not so busy on Sunday, and they can just follow the crowd in relax. This high consistence of motion of angle in each area can not only reduce the possibility of walking collision (which may cause safety issue), but can help us find the common sense of pedestrian and further improve our sale strategy. Thus, we can make several suggestions from this phenomenon:

1. People may think the traffic of people becomes massed up and may need more security staff, but this is unnecessary. Because on Sunday, people behave more politely, the possibility of walking collision reduces a lot.
2. For the project of robot sending flayer, on weekends, fix the robot near the south wall of east corridor can be an even better strategy comparing with the strategy we came up with in figure 13. On the one hand, from figure 13, there are still some not so hot areas near the south wall of east corridor that we can deploy the robot without influencing to much of pedestrian's walking comfort. On the other hand, the high consistence on weekend means almost all people go on the left in the east corridor, and people walk on the south part of corridor are people who entering the shopping mall, by deploy a fixed robot there, we can confirm almost all people entering the shopping mall from east corridor can receive the flayer which help intrigue their shopping interest.

5. Still Customer Behavior Analysis

Compare to the walking pedestrian, still customers tend to spend more time in the shopping mall and have larger change of consuming in the mall. By analysis their behavior, we can future filter out the popular areas and detect their interests.

In this part, we also filter out the still customer by velocity. If the velocity of walk is less than 0.15m/s, we mark the person as still. After filtering out the still customer, we plot the still customer distribution, by the method we used in part 2 of this section, as shown in Figure 12. In this plot, the



bar shows the number of person-time shows in the grid for the whole day.

Comparing with Figure 12, the number of still customer decreases even more significantly on Wednesday comparing to Sunday. On Wednesday, there is nearly no person still in the shopping mall, which verify the conclusion we made from Figure 13 that on Wednesday, most people are just walking through the shopping mall.

For the Sunday still customer distribution, we found there are some relative popular areas (color in red or orange), and we tried to figure out why these regions tends to be more popular, and where should the shopping mall add function to make better use of every area.

As shown in Figure 19 (the still customer distribution on Sunday, 6th January, 2013), there are in total 8 popular regions in the area (marked in black circle).

We analysis the layout of the shopping mall on the day, and come up with a matching about the possible reason of popular areas as shown in Figure 19. To verify our match, we take the facing angle into consideration. As shown in Figure 20, we calculate the average facing angle for still customers in a grid for the whole day. The result of Figure 20 helps us verifing some matches. The pamphlets and TV area is popular is indeed because of the pamphlets and TV, because the still people there are looking to the wall direction which has pamphlets and TV on it. Similarly, the reason why information boards area, shop front area, information desk area and bench area are popular is indeed because of these functions. Because the facing angles are indeed facing to the information boards, shop front, information desk and sit on the right direction of bench. While, for the shopping event match, although we cannot verify it by facing angle, the analysis on the next section will help us to verify its correctness.

5. Children Distribution Analysis

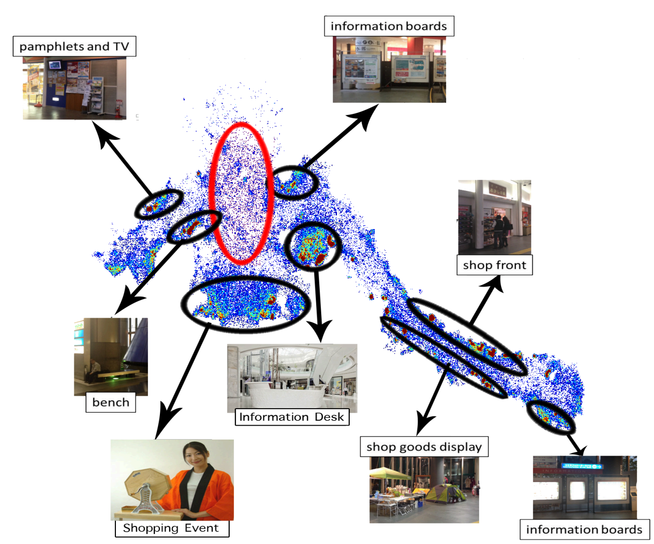


Figure 19 Popular area in Mall

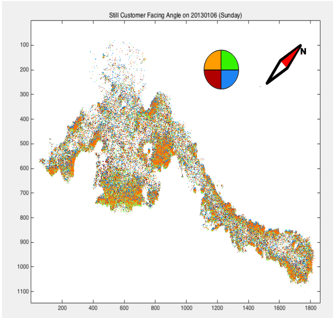


Figure 20 Still Customer Facing

Now, let's take z axis into consideration. We take people under 1.2 meter tall (in z axis) as children and filter them out to research their distribution similar to what we did in Figure 12. As shown in Figure 21, the distribution change of children tends to be larger than all pedestrians. In other word, on Wednesday, there are nearly no children in the shopping mall, while on weekend, the number of children gains significantly. This actually quite suit our common sense. We will further analysis the detailed difference for the number of children in the next section.

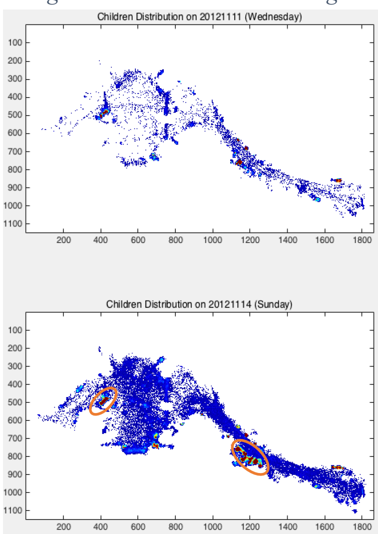


Figure 21 Children Distribution

From Figure 21, apart from the larger fluctuation between weekdays and weekends, the children distribution data also shows high concentration on Weekend. Many children spend lots of time on bench and shop good display area (circled red in Figure 21). We suggest the shopping mall to assign more security guard to protect these children and do children topic promotion in these two area.

iii. Pedestrian Flow Rate Analysis

1. Analysis of Pedestrian tracking data

A. *Weekdays and Holidays flow rate comparison*

The raw data obtained from origin mobile social robots research contains columns of information including time, id numbers, x_position, y_position, height, velocity, motion angle, and facing angle. Note an id number is assigned to each person when they first entered the 3D sensor sensed region and continue tracking until the person exit the area [1]. Also, The time was measured in million seconds. In order to simplify the plotting load, process the data files

```
1351423491.460,20243400,33585,-19081,1748.046,1037.173,2.932,3.017,40
1351423491.460,20243500,-23657,1172,1406.207,252.536,-0.873,-0.841,40
1351423491.460,20244600,-33535,301,1927.094,863.277,0.851,0.958,40
1351423491.460,20244601,39785,-21196,1805.129,1653.765,2.766,2.768,40
1351423491.506,15423701,-20843,-8084,998.553,471.153,-2.562,-2.594,40
1351423491.506,18474301,45020,-28837,967.363,612.359,-2.356,-2.362,40
1351423491.506,20145900,40794,-16693,1023.806,856.481,2.937,2.846,40
```

Figure 22 Dataset example with time indentifier after python code process

where processed by python to add an extra column identifier for time slot (Figure 22), this divide each day into forty time slots.

Then, we use apache pig to process the data to get the pedestrian flow rate of single day with respect to times. The apache pig is a platform for analyzing large data sets. It produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist, from which we get high efficiency in processing data [11]. The procedures are as follow:

1) Load data to apache pig source
2) Group the data by identifier times
3) Count ids in each group of identifier times
4) Generate data for identifier times vs. number of ids

Therefore, obtain the time vs. number of persons of the whole space in mall. Repeat the pig procedures for a package of datasets of Wednesdays, and then for Sundays for comparison.

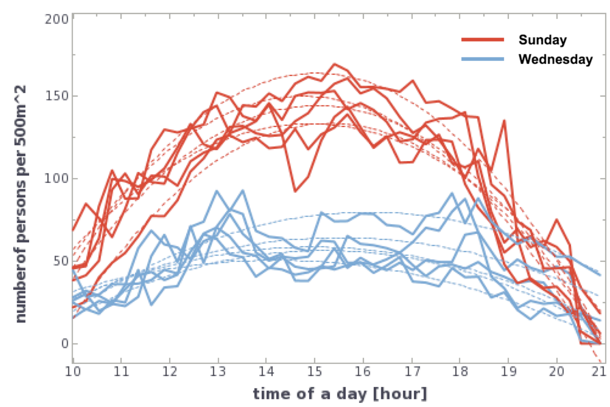


Figure 23 Comparison of pedestrian flow rate on Wednesday and Sunday; Data includes: Wednesdays (20121024; 20121031; 20121107; 20121114; 20121121); Sundays (20121028; 20121104; 20121111; 20121118; 20121125)

The flow density of the whole space (Figure 23) on Wednesday increases during morning and remains at a lower level for most day time. In comparison, on Sundays, the trend follows a parabolic pattern, the number of persons in mall increases more rapidly during the morning, peaks during the afternoon from 14:00 to 16:00. The result shows in general more pedestrian were tracked on Sundays than on Wednesdays, as also confirmed in our own experience. In addition, we observed the difference in pedestrian density between Monday and Wednesday is more apparent in the afternoon period, roughly 12:00-18:00. This provides clue for shops in mall to expect excess customers during the period of times. For example, suggest the shops schedule more afternoon shift service crews in compensation to expected increases in demands.

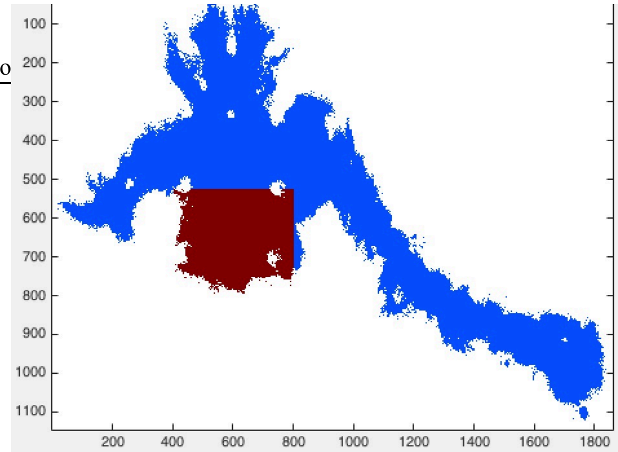


Figure 25 Southern stairs area of study

From the same set of data, we also plotted the average number of tracked pedestrian in mall on Wednesday and Sunday (Figure 24). From which we found the average pedestrian flow rate on Sunday is about twice as much that on Wednesday.

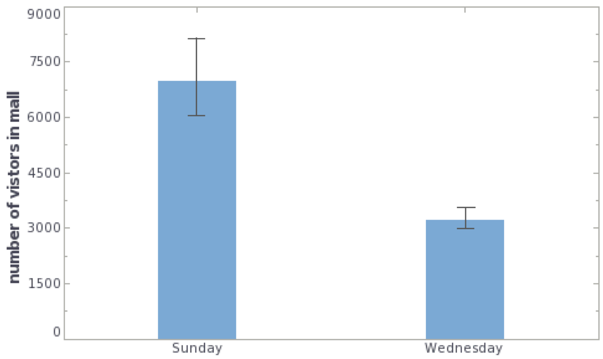


Figure 24 average number of visitors in mall

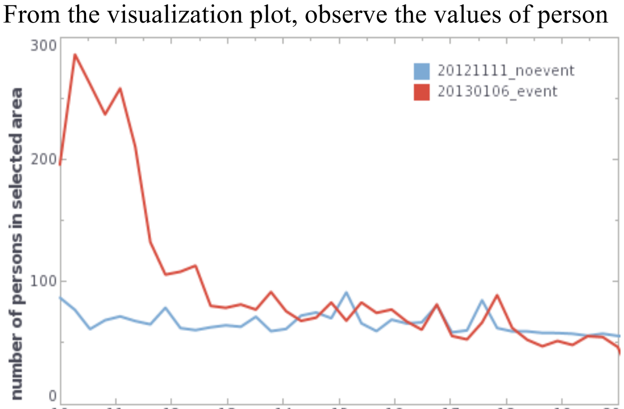
B. Event day and Nonevent day flow rate comparison

Another approach was made by analysis of the pedestrian flow density of a corner area for event days and nonevent days. From previous research done by D. Brscic, et al. [1], the use of southern stairs shows distinct differences on event and nonevent days. Excess use of stairs could be a risky scenario for public safety that both manager of shopping mall and customer should have concerned about. In order to evaluate how bad, the use of stairs situation can be for event days in comparison to regular days, and provide suggestion for event planer, we perform following analysis.

In this study, we specified the target area of southern stairs for study (Figure 25). The stairs lead outside to a wide space in front of building, where sometimes used for different events [1]. The enclose southern stairs area is important path to the main event area.

Referring the vitalization flow rate maps that discusses in later session, we first picked dataset for dates that with and without events. 2012 November 11 and 2013 January 6 are studied as an example (Figure 26). First, use pig to program filter out the data lines only for enclose area of stairs [11]. Then repeat the process discussed in comparing pedestrian flow rate on Wednesday and Sunday.

From the visualization plot, observe the values of person



using the stairs is considerably larger on event days only in a specific time slot, for when the event is holding. On 2012 November 11, the persons using the stairs is flatted all day, steady constant flow rate through stairs. On 2013 January 6, there's an abrupt number of person using the stairs in the morning from 10:00 to 12:00, in the period the usage of stair is roughly three times of the regular days. Other than that the flow rate for the rest of days is similar to regular days.

Checking the event schedule in ATC mall for past years [12], we found the Osaka City New Year firefighters' event happened on January 6 in 2013 and the event last two hours from 10 to 12 in the morning. This confirms our prediction of pedestrian flow rate corresponding to event days. Since we got this information, we can give recommendation on adding security around the southern stair area for event period.

3. Applications

We have found from analyzing pedestrian flow rate, shoppers are denser and active on Sundays in comparing on Wednesday and the most differences is in the major afternoon. Therefore, we can suggest the Shopping mall have more afternoon shifts service staff on holiday afternoon, twice as much as than regular days. Moreover, this can provide useful information for shopper who want to avoid peak hours for shopping to shop on weekdays or in the morning and night on holidays.

When planning an event, it's also useful to apply the method we present and discussed above, looking into similar types of events that happened before to predict the pedestrian flow rate in a target area over a spatial time range. From which the event organizer gets ideas about how to plan on public space using, tickets selling, assigning security guard, etc. to guarantee the event's efficiency and safety.

VI. CONCLUSION

Based on the analysis we did in section V. We find some interesting phenomenon.

1. There are more people on Weekends but mainly in the afternoon.
2. Some area is popular only when event happens (which is a waste of space when there is no event)

3. There is in "invisible path" where more people tends to walk on it on weekdays.
4. People follows the public order in narrow areas, and people tends to be politer on weekend.
5. The distribution of children tends to be highly centralized.
6. The ratio of children and group shoppers on weekend is higher than weekdays.

Based on these interesting found, we came up with some suggestions in terms of flyer sending, sale and security strategy, as well as commercial installation.

Flyer Sending & Sales Strategy

1. Weekday flyer senders (which can be robots) should stay beside the "invisible path" and advertise business related service.

2. Weekend promotions can happen mainly in the afternoon at the end of the east corridor or shop good display side of corridor to attract shoppers and avoid congestion.

3. Children sales should mainly focus on weekends and only at shopping good display and bench area.

4. Based on the classification result, the mall should distinguish the group shopper and do suitable sale and send suitable flyer targeting these group shoppers, especially for robot flyer senders.

5. Restaurants in ATC shopping center can set up more single tables on weekdays, and more couple or larger tables on weekends. Note that there is no special need for too many very large table, since most of the groups (both on weekdays and weekends) are of 2 to 3 people.

6. Promotions targeting couples and families can be done more on weekends then weekdays.

Security Strategy

1. Setting security at centers of clusters.

2. Do not really need more security staff on weekend.

3. Pay attention to the goods display and bench area to protect children.

4. Our Mahout classifier can be used for criminal detection, that policeman can filter out individual pedestrians to solve group crime case in ATC shopping center.

Commercial Installation

1. Open more stores, while the number of shopping goods display can be reduced.

2. Open some store or do some sale activity in the triangle area circled by the "invisible path".

3. The mall need more staff at information desk especially on event days and weekend.

4. The mal should come up with more activities to make better use of the "event area" on non-event days.

The analysis is very important because it helps the mall to find some invisible fact concealed in boring pedestrian flow data and come up with useful and targeting suggestions and strategies.

The program we designed can also be used for other sensor tracking scenario such as traffic tracking, body movement, etc.

Work Distribution: We are a well-operated team, and everyone make their best effort on the project. Thus everyone has the same work distribution.

Acknowledgment

We would like to thank our professor, Ching-Yung Lin for the support and inspiration for the project. We would also like to thank him for the knowledge and skills he taught throughout the semester. Besides, we appreciate the help given by our TA, Yongchen Jiang, He offered a lot of inspiration and help during the process of the project. We would also like to thank other TAs, especially Tian Han and Siyuan Zhang for their support.

APPENDIX

VII. REFERENCES

[1] D. Brscic, T. Kanda, T. Ikeda and T. Miyashita, "Person position and body direction tracking in large public spaces using 3D range sensors," *IEEE Transactions on Human-Machine Systems,* vol. 43, no. 6, pp. 522-534, 2013.

[2] F. Zanlungo, D. Brscic and T. Kanda, "Pedestrian Group Behaviour Analysis under Different Density Conditions," *Conference on Pedestrian and Evacuation Dynamics PED 2014,* vol. 2, pp. 149-158, 2014.

[3] A. Sabelli and T. Kanda, "Robovie as a Mascot: A Qualitative Study for Long-Term Presence of Robots in a Shopping Mall," *International Journal of Social Robotics (IJSR).*

[4] H. Kidokoro, T. Kanda, D. Brscic and M. Shiomi, "Simulation-based Behavior Planning to Prevent Congestion of Pedestrians Around a Robot," *IEEE Transaction on Robotics.*

[5] K. Hayashi, M. Shiomi, T. Kanda and N. Hagita, "Are Robots Appropriate for Troublesome and Communicative Tasks in a City Environment?," *IEEE Trans. on Autonomous Mental Development,* vol. 4, no. 2, pp. 150-160, 2012.

[6] F. Zanlungo, T. Ikeda and T. Kanda, "Potential for the dynamics of pedestrians in a socially interacting group," *Physical Review E,* vol. 89, no. 1, 2014.

[7] M. L. Knapp, "Nonverbal communication in human interaction," *Cengage Learning,* 2012.

[8] K. C. L., "Gaze and eye contact: a research review," *Psychological bulletin,* vol. 100, no. (1), p. 78, 1986.

[9] D. Brscic and T. Kanda, "Changes in Usage of an Indoor Public Space: Analysis of One Year of Person Tracking," *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 2, pp. 228-237, 2015.

[10] "CREST Research content, Street flyer sending robots' design and strategies deployment," [Online]. Available: http://www.irc.atr.jp/crest2010_HRI/research.html. [Accessed 14 December 2015].

[11] R. D. Schneider, "Hadoop for Dummies," John Wiley & Sons, Inc, 2012. [Online]. Available: www.wiley.com.

[12] "2016 Osaka City Fire New Year firefighters' event," 6 1 2013. [Online]. Available: http://www.atc-co.com/event/000117/. [Accessed 12 12 2015].

[13] D. Brscic, T. Kanda, T. Ikeda and T. Miyashita, "Person position and body direction tracking in large public spaces using 3D range sensors," *IEEE Transactions on Human-Machine Systems,* vol. 43, no. 6, pp. 522-534, 2013.