Shopping Mall Digitalization and Customer Analysis Visualization

Fei Xiong*
Harvard University

Zidong Huang†
Harvard University

Zihui Zhang‡
Harvard University

Figure 1: Computer vision based solution to extract and analyze the customer flow and features

ABSTRACT

In recent years, the offline shopping malls have suffered from the traditional empiricism analysis and management methodology, because it has become increasingly more challenging to deal with the fast-changing landscape of offline consumption merely adopting the empiricism. *Spatial Intelligence* is a digitalization and visualization tool to help commercial real estate digitalize its physical spaces and manage them in a data-driven way. It utilizes computer vision technology to label customer features and record their moving tracks, then processes the raw data with customized algorithm into KPIs (Key Performance Index) reflecting the mall's performance, and finally visualizes these information in selected types of dashboards with efficient interactivity using tools like D3.js [2], Three.js [4] and Xeogl [8].

1 INTRODUCTION

The physical shopping malls have been suffering from the inadequate and outdated empiricism analysis. The lack of digitalization leads to a redundant structure of the marketing and operation department. It was common that hundreds of people were hired to collect the customer data, analyze the customer trend by talking to stores one by one with low efficiency. What's more, the whole decision process was mostly based on empiricism. Different operation managers might have completely different decisions on rents, brand selection, etc. Faced with the fast changing consumption landscape where the power of empiricism gets weaker, and the fierce competition with online shopping, the offline shopping malls are in urgent need of digitalization and data-driven management ecosystem.

The *Spatial Intelligence* was developed as a solution to help the offline shopping mall go through the digitalization process. It will not only empower the mall with a digital analysis system similar to the online shopping tool, but will also help it utilize the location-based information unique to offline spaces to generate more insights. In this project, we utilized 400 existing surveillance cameras in the HopsonOne Mall in Beijing, China, to label customer features and record their moving tracks. We then transformed these data into KPIs both of the whole mall and of each store, like the count entering / leaving, female / male, youth / senior, etc., on a daily basis. Based on these KPIs, we designed three dashboards to progressively visualize the customer activity patterns to systematically and qualitatively help managers answer critical questions, e.g., What brands are the main source attracting customers into the mall? Which floor has the most female customers? Which exit has the most youth generation? What type of things are people most likely to purchase after having their dinners?

*e-mail: feixiong@gsd.harvard.edu
†e-mail: zidonghuang@gsd.harvard.edu
‡e-mail:zzhang2@gsd.harvard.edu

2 RELATED WORK

2.1 Digital Twin

The digital twin is a concept that builds digital models and structured data of physical spaces and buildings. Its focus has been shifting from academia to industry in recent years, as GIS, BIM, and CIM technology became mature. The digital twin concept is now widely used in smart cities, healthcare, and manufacturing. We review some papers, including [6] and adopt the concept, especially on our "model" page.

2.2 Offline Shopping Reports

Offline shopping reports provide insights into our system design. The Mckinsey [3] brought up the business quadrant of the retail, which is based on the customer number and business volume. That data is needed to classify "popular but not selling" and "lonely but selling much" types [7] suggests some exciting ideas to promote the relationship between customers and stores, based on the corresponding data. [5]It is indicated that the most significant issue of offline shopping is the lack of sophisticated analysis tools.

2.3 Visualization Techniques

The visualization techniques are critical to our project. [1] introduced some basic concepts and techniques of building interactive 3D visualization in web browser. [9] provided a novel approach to visualize the traffic flow in the shopping mall. [2], [4], [8] provided detailed documentation for our implementation, respectively.

3 METHODS

3.1 Data Collection and Categorization

The raw data were generated by the 400 cameras in the HopsonOne mall. We used computer vision techniques to categorize customers based on gender, age, and other appearances. Then, we

cleaned, grouped, and converted the data into the JSON format:

{"store_name": "AAPE", "date": "2021-2-1", "type": "fashions", "enter_number": 179, "enter_rate": 0.0460036, "female_number": 49, "female_rate": 0.27476, ...}

For privacy protection, facial recognition is not used in our project. We collected the data on a customer group level instead of individual level. Also, the raw data is deleted after higher-dimensional data is produced. Other desensitization approaches are taken according to local law and policy.

It is worth mentioning that with the limitation of the camera number, only part of the area are analyzed. Therefore, in the model page, around one fourth of the stores could be explored. Nevertheless, our methodology could be easily extended to a larger scale. Also, for testing purposes, we only included the dataset of February 2021.

3.2 Visual Encoding and Interaction Design

The overall principle of the visual encoding was to present the complex information creatively by reasonable reorganization to help the management team understand the current situation quickly and generate detailed insights with further interaction. Therefore, it should be able to provide both an overview and enough details. We designed three separate but internally connected parts to achieve the goal: a Sankey particle chart, a bubble chart, and a 3D model.

The Sankey particle chart (Figure 2) was designed to provide the most straightforward information about the overall customer flow on a designated day. The larger the volume, the wider the link, and the faster and denser the particles. Users can choose the categorization of customers (there are two choices in the current version: Gender and Age) and any date they are interested in to control the information displayed. The exact number of a specific link or node will be displayed to provide more details by hovering on the link or the node. This chart provides answers to general questions about the mall, e.g., "where did men mostly go?" or "what customer type was the majority visiting fashion stores, young people or seniors?" which will help the management team grasp an overview of the current situation. We intentionally kept the graph abstracted to help the user focus. More details will be provided in the following graphs.

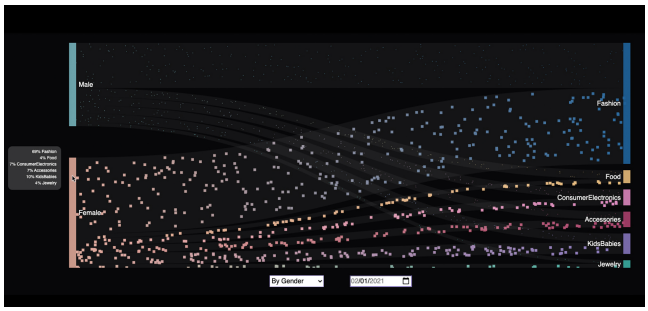

Figure 2: Sankey Chart

The bubble chart (Figure 3) further provides information about the trend of each store's customer flow through a period of time. Users can designate a specific customer type, and then the chart will display the flow data of this type of each store in the form of discrete bubbles. By hovering on a specific bubble, the trendline of the specific store will be displayed. Furthermore, by clicking on it, users are able to keep the trendline stay for comparison. The design intention here was to allow users to control the visibility of crucial information thus reducing distractions. This chart will further help the management team understand the trend of a specific customer type at the individual store level, generating more detailed insights for them.

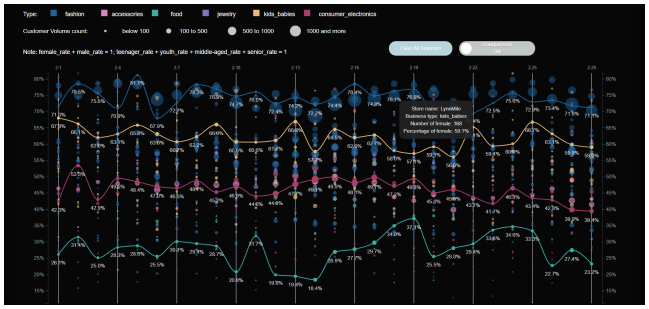

Figure 3: Bubble Chart

The last 3D model (Figure 4) was designed to bring the management team into the most fundamental level of a single store to examine its data thoroughly. Unlike the previous two containing overviews, this one was designed to only focus on the designated object. The reason to adopt a 3D model was to provide unique location information that no other chart can provide, which is critical when analyzing a specific store.

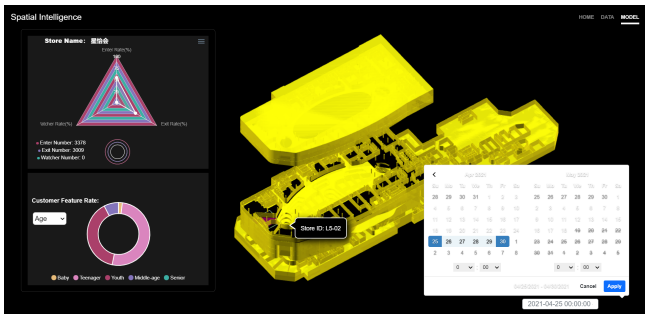

Figure 4: 3D Model

The three charts together help the management team better comprehend the customer flow from different but progressive levels.

4 Results

The final deliverable of our project is a website consisting of three interactive and informative dashboards with information on the offline customers' activities in the mall. We have conducted a few user experience surveys on the user's interpretation of all three pages. There are some minor confusions from their feedback, and we have tried our best to solve them. In general, we believe that our final delivery has made contributions to this area. What is more, it has revealed some fun facts. We have collected some as follows:

- The Sankey Chart shows that 'female rate' particles are running far faster than the 'male' ones to the fashion shopping categories on Feb 1 2021, which means there were more female entering the fashion stores than men on this day.
- The Bubble Chart shows that the movie theater CGV has averagely the highest entering rate through the February, which indicates customers mostly went there with a clear purpose to watch movies other than just wandering around or passing by.
- Store Haircode has the most senior customers.
- Two adjacent stores can have a dramatic disparity of customer enter counts.

However, there are some limitations to our work. The limitations are lying in both the dataset and the representation. Since the dataset

is collecting from the security cameras without facial recognition, the tagging might not be 100 percent accurate when trying to mark customers with tags that are not conclusive, just entirely based on human figures, like age or gender. Moreover, the dataset is incomplete, so it is assertive to make any decisive conclusions. For the Sankey, we have not humanized each particle to have it contained the tags accordingly. Moreover, our color palette is not color-blind-friendly, which can be difficult for them. Furthermore, the model page takes too long to load, which should be improved.

5 Discussion

The initiation of this project has been an experiment on the managing and operating system of modern offline shopping malls. It evokes the potentials of unattended digital information and its applications for spatial improvements. In this case, we are targeting the commercial real estate spatial structures. This project explores multiple design techniques and engaging methods to showcase the potentials of digitalizing physical figures and their movements while dehumanizing data for preserving privacy.

Figure 5: Capturing physical human features into tags in stores.

Our whole system is beneficial by the concept of tagging, which is about subjective description of basic observation. The process is more like to symbolize a real person into a generic persona with no background story. However, the tagging can be subjective, but the cataloging can be objective and assertively conclusive. Therefore, our project can serve as an experiment to test out the accuracy of tagging and cataloging.

On the other hand, due to the commercial value of this experiment, the potential is to establish a database in the physical world while respecting the privacy of individuals. Recently, we got to apply this technique to help the managing and operating system grasp more potentials in storefront design, promotion events, or visual influences for spatial management.

As in 3d model of the Hopson One mall, the digital twin also provides users a sense of spatial experience rather than just gazing at the 2d elements like lines or shapes. Under this circumstance, we also have to consider the four-dimensional representation as in shifts in time and space. We hope that this visualization could impact the interaction and engagement development of data visualization, enabling a more user-friendly experience with data.

6 Future Work

This project has taken a sip from the future application of digitalizing physical data. Nevertheless, as discussed in the previous sections, some remaining problems have not been solved, and applying pitfalls still stand to what we have now.

In the visualizing process, attempts were made to revivificate the customers' tracks in the mall. However, due to the deficiency of security cameras and the lack of techniques to link the cameras together to identify the same customer among different cameras, the

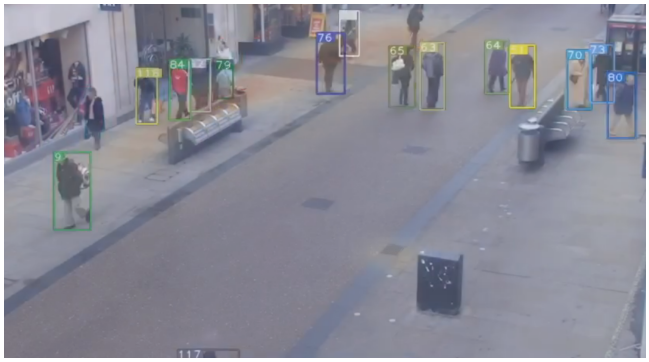

Figure 6: Potential application for fighting for the offline world

attempts failed. Therefore, we believe that there should be more tags with more detailed descriptions to pinpoint an individual while protecting his/her privacies. During the developing process of this experiment, we hope to arouse the imagination in applications with this potential database and the integrity to revere the experimenting participants and suppress the potential abuse of the datasets.

As for the potential improvement in the visualizations themselves, the three pages are somewhat still a little bit alienated from each other, so that we hope to enhance the linkage among those pages. For example, clicking on the particles or rectangles from the Sankey page can lead the users to the other pages for more detailed info like the store's location in the 3d model or the entered rate trend of that specific store the footprints of that specific customer.

7 Conclusion

Digitalization is a primary step for the Internet of Things, and Spatial Intelligence utilizes the transformation from images to numeric and time-stamped data to provide service for digitalized spatial management. These visualizations provide opportunities to imagine the possible application for imaging tagging and visualizing. In the meanwhile, our visualizations are not only informative but also lucrative for commercial usage. Serving as an assessment and analysis system can provide subjective decision-making suggestions on rent, business type deployment in malls, and customer flow monitoring. Our vision is to let the viewers be more aware of the unattended resources near us and work towards a better offline world with the database we are about to establish in the near future.

Acknowledgments

This work was supported by PlaceInt.

References

[1] L. Bergman. Creating an appealing 3d-visualization tool for baseboards in the web browser, 2020.

[2] M. Bostock and d3 authors. d3:data-driven documents, 2020.

[3] H. Briedis, A. Kronschnabl, A. Rodriguez, and K. Ungerma. Adapting to the next normal in retail: The customer experience imperative, 2020.

[4] R. Cabello and threejs authors. threejs.org, 2010.

[5] D. Chaturvedi. Research on online market vs offline market. 02 2019.

[6] A. Fuller, Z. Fan, C. Day, and C. Barlow. Digital twin: Enabling technologies, challenges and open research. *IEEE Access*, 8:108952–108971, 2020. doi: 10.1109/access.2020.2998358

[7] J. H. III, J. S. Brown, T. Samoylova, K. M. Lobaugh, and N. Goel. The retail transformation cultivating choice, experience, and trust.

[8] x. Kay. xeogl:data-driven webgl from xeolabs, 2016.

[9] J. Opolka, P. Riehmann, H. Peter, and B. Froehlich. Inside mall: Visual analytics of customer behavior and activities. pp. 183–190, 01 2019. doi: 10.5220/0007284901830190