Sparse: A Reservation and Computer Vision-Based Room Occupancy System for Malayan Colleges Laguna's Center for Learning and Information Resources

Job Lipat*, Charmaine Eunice Rabano, Mark Anthony Mamauag, Madeline Isabel Galang, and Jennifer Contreras.

College of Computer and Information Science, Malayan Colleges Laguna, Laguna, Philippines
*Corresponding author, E-mail: lipatjj@live.mcl.edu.ph

Abstract

The Center for Learning and Information Resources (CLIR) or the library in Malayan Colleges Laguna (MCL) faces some problems regarding its usage. First, students get frustrated when they go to the CLIR just to find out that are no more available seats. Second, its operations are affected by safety protocols amidst the pandemic. To help alleviate these problems, the researchers proposed a system called Sparse. It is a computer vision-based room occupancy detection and seat reservation system. To develop the system, the researchers performed three key activities. First, the researchers trained and benchmarked three object-detection models, namely the Faster R-CNN, RetinaNet, and SSD models. These models were tested using the SCUT Head dataset. The metrics used for comparison were the models' average processing time and root mean squared error. Second, the researchers collaborated with the visitor and librarian resource persons to identify the necessary features for the system. Lastly, the researchers surveyed 20 prospective users, including MCL students, librarians, and faculty members, to evaluate the usefulness of the identified features.

From these activities, the researchers compared the Faster R-CNN (10.321 Pre-trained RMSE, 18.139 Trained RMSE, 0.117s average processing time), RetinaNet (12.660 Pre-trained RMSE, 19.026 Trained RMSE, 0.111s average processing time), and SSD (15.351 Pre-trained RMSE, 21.900 Trained RMSE, 0.018s average processing time). Among them, the Faster R-CNN was selected because it resulted in the least amount of Root Mean Square Error while having an acceptable average processing time. Then, the features for the website that were identified for visitors are the Room Occupancy Chart, Interactive Suggestions, Library Information, and Seat Reservation. For librarians, the features that were identified are the Room Occupancy Information, Report Generation, Summary of Reserved Spots, Change Effective Capacity, Library Information Management, and Account Management, wherein the last three features are only accessible to the head librarian accounts. Third, the results of the survey show that the identified features were useful to users. Additionally, the survey has also gathered suggestions that could be implemented in the future.

Keywords: *room occupancy, reservation, people counting, computer vision, object detection*

1. Introduction

The Center for Learning and Information Resources (CLIR) or the library in Malayan Colleges Laguna (MCL) is one of the central areas of education in the college. It is a place where visitors may spend their time studying, waiting for their next class, or reading reference materials. However, there are two problems with using the CLIR: 1) MCL's students, staff, and faculty members go to the CLIR but immediately leave because of no vacant seats. This is a common experience among visitors of the CLIR and is not only frustrating but also a waste of time and energy. 2) With the current COVID-19 pandemic, it became imperative to avoid crowded places to minimize the spread of the virus. For this reason, the government has set policies to control the operational capacities of certain places (Department of Health, 2021). These policies include a reduction in the operational capacities of libraries. With limited seats available, it becomes harder for visitors to secure seats.

To solve these problems, the researchers would like visitors to know the room occupancy of the CLIR without having to go there. Given this, the researchers propose a system named Sparse. Its primary functions include 1) recording of room occupancy of the CLIR at regular intervals; 2) viewing of current room occupancy and other useful information for visitors; 3) CLIR seat reservation for visitors.

For recording the room occupancy of the CLIR, the researchers implemented a computer vision-based people counting algorithm. To determine which algorithm is the most appropriate for the system, the researchers compared three object-detection models: the Faster R-CNN model (Ren et al., 2017), the SSD model (Liu et al., 2016), and the RetinaNet model (Lin et al., 2017). These models were chosen due to their high accuracy. Further discussion is provided in the Related Literature section of this paper.

This study would benefit both the visitors and the librarians. For visitors, the system would allow them to view the CLIR's Room Occupancy Information beforehand, which will enable them to decide whether to visit the CLIR and at what time. The visitors, therefore, will be allowed to have an improved experience when using the CLIR. For the librarians, the proposed system would allow them to maintain the CLIR's room occupancy in safe numbers, as set by government policies, by adjusting the room's effective capacity. Also, the librarians will get access to useful information about the usage of the CLIR.

2. Objectives

1) To identify the most appropriate computer vision-based people counting algorithm for the system
2) To identify and implement features that would be useful for visitors and librarians
3) To evaluate the usefulness of each feature of the system

3. Related Literature

3.1 People Counting Systems

People counting is one of the central tasks that the proposed system is going to perform. Raghavachari et al. (2015) defined people counting as the task of determining the number of people in an area. These systems are applied in a wide variety of situations, such as in retail (Lacanlale et al., 2021), in practicing social distancing (Punn et al., 2021), and on university campuses (Waitz Inc, 2020).

3.2 Techniques for People Counting

The techniques available for counting people have evolved in the past few decades. One of the earliest ways to automatically count people is by using infrared beam counters (Beymer, 2000). These counters use infrared lasers to detect when someone passes through a certain area. Another early technique to determine room occupancy is to use thermal-based counters as demonstrated by Amin et al. (2008). These systems determine the room occupancy by using heat signatures of people in an area.

A more recent way to perform people counting is by using network traffic-based techniques. To determine how many people there are, the system detects network traffic such as Wi-Fi probe requests or Bluetooth signals in an area (Waitz Inc, 2020). These are then used to derive the number of people in the room.

Another modern approach is by using computer vision-based techniques. It is the state of the art in people counting. Computer vision-based techniques start by capturing an image of the area. After capturing an image, the system counts the number of people in the image by applying image processing techniques and methods such as object detection (Nixon & Aguado, 2019).

3.3 Object Detection and People Counting

In computer vision, object detection is the task of determining the identity of objects in an image given a set of known labels. By filtering the objects such that only people are detected, the number of people in an image could be detected. Traditionally, object detection is done through techniques such as HOG (Dalal & Triggs, 2005), SIFT (Lowe, 2004), and the Viola-Jones algorithm (Viola & Jones, 2001). However, recent developments in deep learning and convolutional neural networks (CNNs) have led to high accuracy models such as Faster R-CNN, Single Shot Detector, and RetinaNet.

The Faster R-CNN model was proposed by Ren et al. (2017). In their paper, they have identified that the region proposal algorithm was bottlenecking the older Fast R-CNN model. To solve this issue, they introduced a region proposal network (RPN). Compared to the old region proposal algorithm, the RPN is a fully convolutional network. These allow the model to be trained end to end. By introducing this method, they were able to achieve an average precision (AP) of 21.9 on the Microsoft Common Objects in Context (COCO) test-dev dataset with an average running time of 198ms per image.

The Single Shot Detector (SSD), proposed by Liu et al. (2016), is a single-stage detector, meaning that the SSD model detects objects in one stage, as compared to the R-CNN which has a separate stage for region proposal and classification. The SSD performs object detection first by dividing the image into S-by-S boxes. Then bounding boxes and confidence scores are generated for each grid cell. Lastly, non-maximum suppression is applied to each detection to select the best bounding box on repeated detections. On top of this, SSD also uses multi-scale feature maps to improve detection. These feature maps allow object detection at multiple scales. By implementing this, they were able to achieve an AP of 26.8 and an average FPS of 59.

RetinaNet proposed by Lin et al. (2017) is also another single-stage detector. In their paper, they have identified that the vast number of background examples in training causes poor model accuracy. To address this problem, they have introduced the focal loss factor to the standard cross-entropy loss. The focal loss factor decreases the loss for easy negatives like background while giving a high loss for hard negatives, which allows the model to focus on training on misclassifications. By applying this, RetinaNet was able to be the first one-stage detector to surpass the average procession of two-stage detectors at that time.

4. Materials and Methods

4.1 Model Selection

To identify the most appropriate computer vision-based people counting algorithm for the system, as mentioned in objective 1, the researchers first selected a dataset. Then, the models for people counting were identified together with the correct metrics to be used in comparing the identified models. The next two sections will respectively discuss the dataset and metrics used to compare each people counting model.

4.1.1 Dataset

The SCUT Head dataset by Peng et al. (2018) was used to train and test the model. The dataset contains 1443 images for training, 462 images for validation, and 500 images for testing. The distribution of the number of people per image is shown in Figure 1.

The SCUT Head dataset was chosen over other full-body datasets for two reasons. First, it is expected that most images that would be taken from the CLIR will only contain the upper body as the lower body will be obstructed by tables. Second, the SCUT Head database mostly contains images taken from classrooms and libraries. These reasons make the dataset more representative of the environment where the model will be deployed.

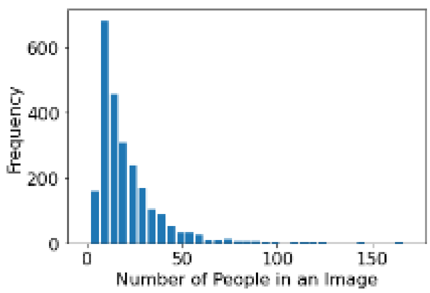

Figure 1 The frequency of the number of people in an image

4.1.2 Model Evaluation

The object detection models that were evaluated are the Faster R-CNN, SSD, and RetinaNet. These models used the implementations from the Torchvision library. The Faster-RCNN and RetinaNet model has a ResNet FPN backbone and were pre-trained on the COCO dataset. On the other hand, The SSD model uses a Vgg16 backbone and has also been pre-trained on the COCO database.

To determine which model is the most appropriate for the people counting system, the three mentioned models processed the test dataset of the SCUT Head dataset on Google Colaboratory. While processing the dataset, two metrics were collected and were used to compare the models.

The first metric that the researchers considered is the processing time. Processing time is important to consider so that the system can deliver information promptly. As shown in Figure 2, the processing time is calculated by getting the average time it takes to process each data item in the dataset. Models with lower processing times will be considered better. The researchers arbitrarily selected 5 seconds as the threshold for acceptable processing time. If the model's processing time is more than 5 seconds, the model will be deemed unacceptable.

$$\mathrm{T} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Figure 2 The formula for Processing Time

Error is the measure of how far the predicted values of the model were from the actual values. This factor is important to consider because having low error would help users get better information about the room occupancy of the CLIR, and therefore allow them to make better decisions. To determine the error of a model, the researchers are going to use the root mean square error metric. The formula can be seen in Figure 3. The Root Mean Square error is used to determine the average prediction error over all data values in the test set (Salkind, 2010). The lower the root mean squared error, the better the model has performed.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(i_i - y_i)^2}{N}}$$

Figure 3 The Root Mean Square Error Formula

To determine the most appropriate computer vision model, the researchers compared the average processing time and mean square error of each model. The model that provides the lowest error while being sufficiently fast was selected.

4.2 System Requirements Gathering

To determine the requirements needed for the system, in line with objective 2, the researchers have consulted one resource person each for visitors and librarians. For visitors, the researchers have consulted a third-year IT student who frequently uses the CLIR. For librarians, the researchers consulted the Library Officer of the CLIR.

At the end of each sprint, the researchers demonstrated the system to the resource persons. After demonstrating, the researchers gathered feedback on the current features and suggestions on what additional features could be implemented. Then, these features were implemented in the next sprint.

4.3 System Evaluation

To determine if the system is useful for users, in line with objective 3, the researchers surveyed 18 visitors and 2 librarians. The survey consists of three parts. The first part consists of profiling questions. In the second part, the users evaluated the usefulness of each feature through a 5-point Likert scale. Lastly, the users were asked what additional features they would like to be implemented in the future.

Additionally, the researchers have created survey guidelines for the respondents. The survey guidelines contain all the necessary information the respondents need to know before answering the survey. It includes the objective of the survey, videos explaining each feature of the system, contact information of the researchers, and answers to frequently asked questions.

5. Results and Discussion

5.1 Model Selection

The next two subsections will discuss how the model was trained and compared based on the metrics mentioned in the previous chapter. Using these metrics, the most appropriate computer vision-based people counting algorithm for the system was identified and, therefore, accomplishing the first objective.

5.1.1 Model Training

As seen in Figure 4, each model started with a high validation loss but gradually decreased as the epoch increased. The validation losses of the R-CNN are the most varied among the models. It could be seen that the loss increased at epochs 3 to 4 and epochs 6 to 7, which was because the learning rate of the model was set too high at those points. Because of this, the researchers had to lower the learning rate in the next epoch to allow the validation loss to continue to decrease. The final validation loss of the R-CNN model is 5.135.

The learning rate of the RetinaNet model plateaued in the fifth epoch, but, after adjusting the learning rate, the loss significantly decreased. The final loss of the RetinaNet model is 4.077. Meanwhile, the validation loss of the SSD model drops sharply in the first few epochs but plateaued in the latter epochs. The final loss of the SSD model is 19.022.

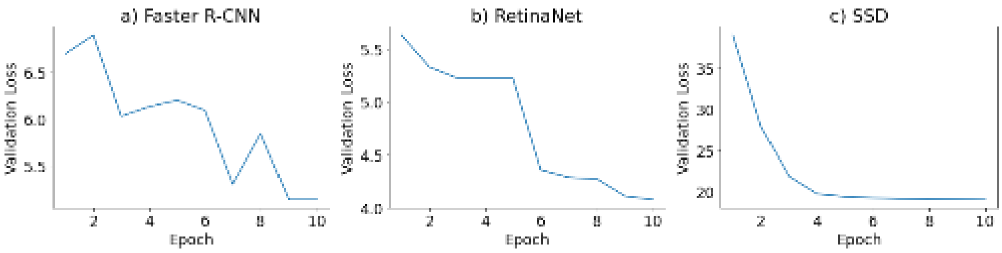

Figure 4 The validation loss of each model

As shown in Figure 5, training the model for 10 epochs takes a significant amount of time, therefore, the researchers could only train the model for a few epochs. The Faster R-CNN took 5.33 hours to train, while RetinaNet took 8 hours and SSD took 4 hours. Google Colaboratory dynamically allocates resources depending on how many people are using the platform, which may have affected the models' training time.

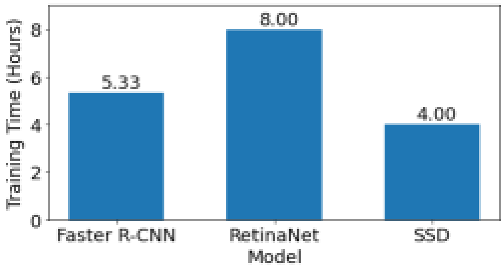

Figure 5 The training time of each model

5.1.2 Model Comparison

As seen in Figure 6a, all the average processing times of the trained models fall within the acceptable range of processing time. The RetinaNet and Faster R-CNN models had almost similar processing times of 0.111 seconds and 0.117 seconds respectively. However, the processing time of the SSD model is significantly lower than the other models having an average processing time of 0.018 seconds. The SSD model may be more suitable if the model needs to be deployed in a computer with limited processing power.

As could be seen in Figure 6b, each model had a different root mean squared error. The models with the blue bars were trained by the researchers, and the models with the orange bars are the pre-trained models of PyTorch. The results show that the models trained by the researchers have unfortunately performed worse than the pre-trained models, which may be because the models were only trained for a few epochs resulting in the model underfitting the test dataset.

The model that performed the best was the pre-trained Faster R-CNN model with a root mean square error of 10.31. Because of this, the Faster R-CNN model was integrated into the system.

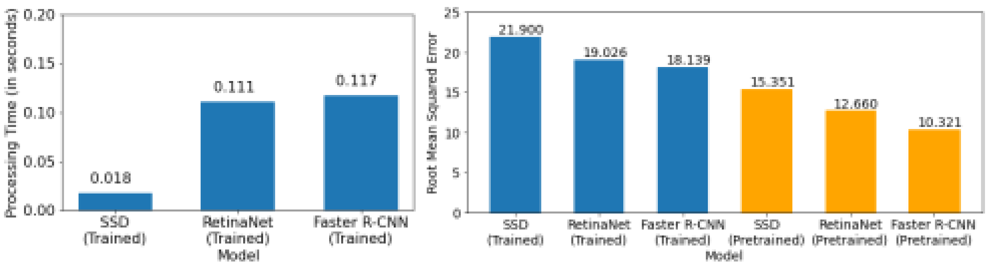

Figure 6 The (left) average processing time and (right) RMSE of each model

5.2 Feature Identification and Implementation

For the second objective of the paper, the researchers designed and developed the Sparse system based on the identified features. The next two subsections will respectively discuss the system architecture and the features that were implemented.

5.2.1 System Architecture

As seen in Figure 7, the system starts with the camera inside the CLIR. The video captured by the camera is sent to the computer vision model. From the video feed of the camera, the computer vision model calculates the current room occupancy and stores the calculated value inside the database. The web server queries the database to render the student or librarian interface with the requested data. This interface is then viewed by the student or librarian through a web browser.

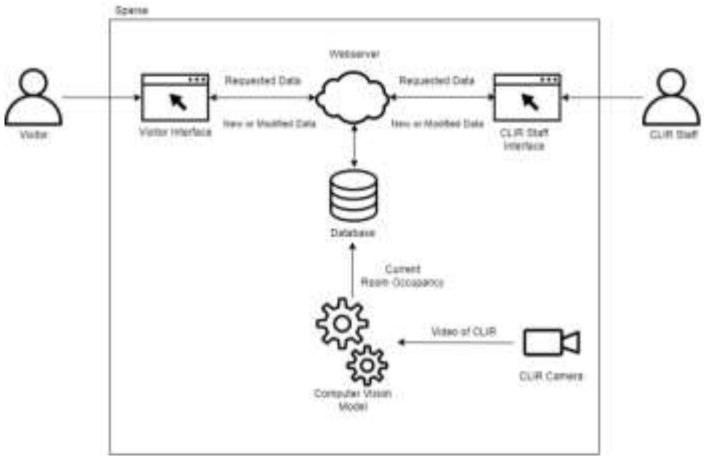

Figure 7 The system architecture of Sparse

5.2.2 System Features

There are two main interfaces in the system, the visitor interface, and the librarian interface. The features in the visitor interface include library information, room occupancy graph, interactive suggestions, and reservation. The available features in the librarian interface are room occupancy information, change effective capacity, report generation, a summary of reservations, library information management, and librarian account management.

As could be seen in Figure 8a, the library information panel in the visitor interface displays when the CLIR is open. The displayed schedule can help visitors decide at what time they will go. Next to the information panel, visitors could see the Room Occupancy Graph feature which is shown in Figure 8b. The computed occupancy shown in the center of the graph is computed by adding the occupancy detected by the computer vision model, and the number of reserved spots. The reserved spots and the model occupancy are differentiated in the chart for the information of the visitor. Additionally, the status of the current room occupancy will also change based on the current room occupancy.

Figure 8 The interface of the (a) library information and (b) room occupancy graph features

The panels shown in Figure 9 are the Interactive Suggestions feature. The suggestions can help visitors decide if they want to go to the CLIR and at what time. The panels in Figure 9a will display action suggestions based on the current room occupancy status that was shown in Figure 8b. The panels in Figure 9b display a message about the comparison of the current day's room occupancy to the previous day's room occupancy. The panels in Figure 9c display the peak hours from past room occupancy records.

(a) (b) (c)

Figure 9 The interface of the interactive suggestions feature

Figure 10a shows the Reserved Spots per Hour panel which displays the number of reserved spots per hour for the future hours and days. Checking this will allow the visitors to make better decisions before reserving. This feature is also displayed in the librarian interface.

The reservation button is found below the Reserved Spots per Hour panel. After clicking the button, the user will be redirected to the reservation page shown in Figure 10b if the user has already logged in. If not, the visitor will be redirected to the login page.

On the reservation page, visitors can pick their desired date and time and can also see how many other visitors have reserved at each time. After reserving, the user will be redirected to the Successfully Reserved page shown in Figure 10c. On this page, the reservation details are displayed, and the user is informed that the user must claim the spot ten minutes after the reserved time. Once the user has arrived in the CLIR, the user must click the "I'm Here" button, to claim the reserved spot. If the spot has been claimed, the user will be redirected to the Successfully Claimed Spot page shown in Figure 10d.

(a) (b)

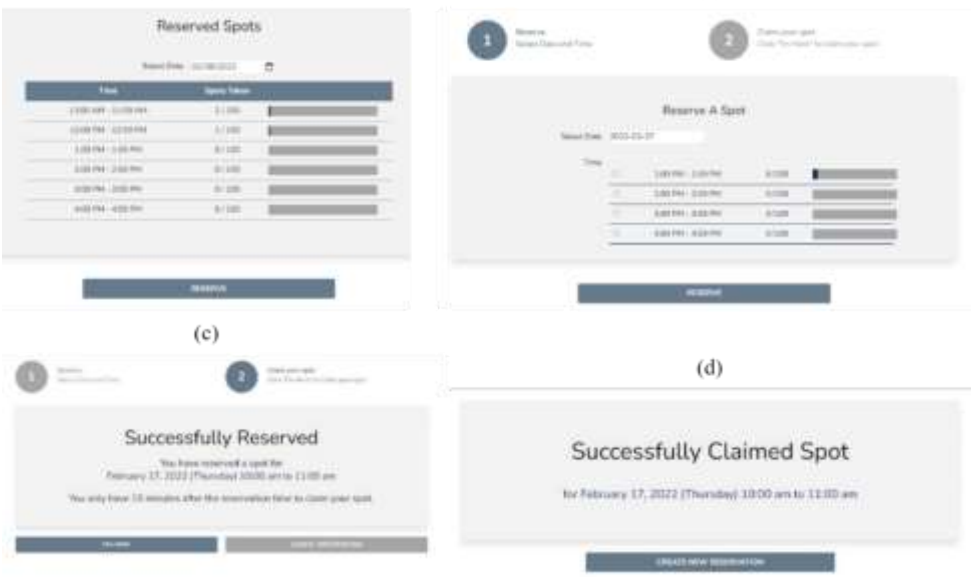

Figure 10 The interface of the reservation featureIn the librarian interface, the room occupancy information is

located on the left side of the first panel shown in Figure 11a. It is similar to the room occupancy chart shown to visitors. This feature is important for librarians so that they will be aware of the current room situation and will be able to take any necessary actions.

The effective capacity is located on the right side of the first panel. Librarians may only view the current capacity, while head librarians may change the current effective capacity of the CLIR as shown in Figure 11b. It is an important feature during the current COVID-19 situation where the government implements a policy that reduces the operational capacities of certain places, including libraries (Department of Health, 2021). This feature makes it possible to change the maximum number of visitors at a time according to changes in protocols.

(a) (b)

Figure 11 The interface of the room occupancy and change effective capacity feature

Another feature in the librarian interface is the reports feature. This feature was suggested by the librarians as they mentioned that it would be useful for their daily operations. They have also suggested that the reports should be downloadable in Microsoft Excel format. The reports include visualizations of the Average Room Occupancy, Room Occupancy for the Past Hour, and Student Visitor Reports shown in Figure 12. The excel format of these reports could be downloaded by clicking the “Download report” button.

Figure 12 The interface of the reports feature

The features in Figure 13 are only accessible to head librarians. The library management feature, shown in figure 13a, allows the head librarian to change the opening and closing schedule of the CLIR as needed. On the other hand, the librarian account management feature, shown in Figure 13b, allows head librarians to create, modify, delete and view other librarian accounts.

Figure 13 The library information management and account management features

5.3 Evaluation of the Usefulness of Each Feature

In line with objective 3, a survey was conducted to evaluate the usefulness of the system features. The next three sections will respectively discuss the respondents' profile, the evaluated usefulness of each feature, and some comments and suggestions from the respondents.

5.3.1 Survey Respondents

Figure 14a shows that the researchers have surveyed 15 students, 3 faculty members, 1 librarian, and 1 head librarian. The researchers have used convenience sampling and selected a small sample size because of the limited time, the difficulty of finding students who have used the CLIR, and because there are no face-to-face classes. As seen in Figure 14b, most of the students surveyed were in their third year (66.7%) and a smaller proportion was in their second year (33.3%). Meanwhile, all faculty members that were surveyed were part of the College of Computer and Information Science (CCIS) as seen in Figure 14c.

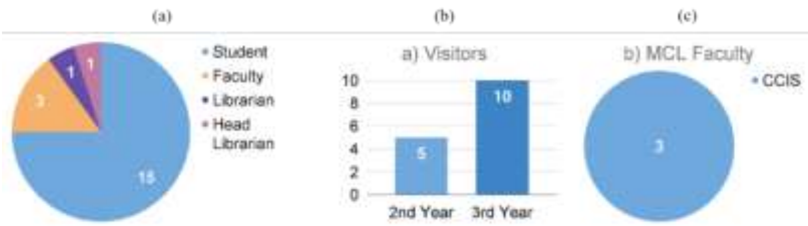

Figure 14 The profile of survey respondents

Respondents varied in the number of times they used the CLIR before lockdown as seen in Figure 15. The largest proportion, comprising 35.3% of the visitors, used the CLIR more than thrice a week. The second-largest proportion, comprising 23.5%, used the CLIR less than once a week. The number of respondents who use the CLIR twice and thrice a week each comprises 17.6% of the samples. The smallest proportion, comprising 5.9% percent of the participants, uses the CLIR only once a week.

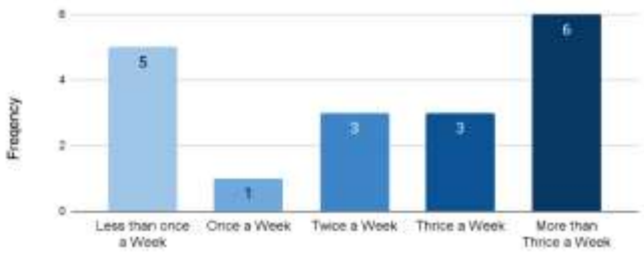

Figure 15 Number of Times per Week Visitors Use the CLIR Before Lockdown

5.3.2 Usefulness of Features

In the survey, the visitors evaluated the usefulness of the visitor features. As shown in Figure 16, all features have an average rating above 4. The feature with the highest score is the Room Occupancy Chart having an average score of 4.72. Next is the Seat Reservation feature with 4.61 and the Library Information feature with 4.39. The feature with the lowest score is the Interactive Suggestions feature with an average score of 4.28 because some visitors found the information presented was too wordy or repetitive.

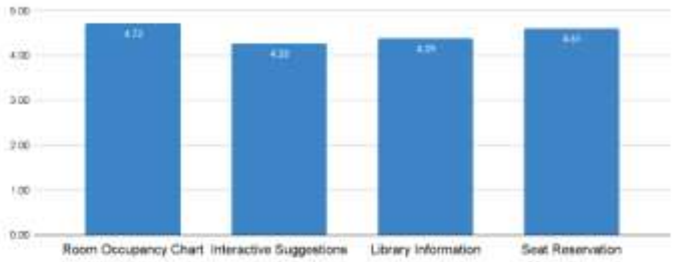

Figure 16 The average rating of each visitor feature

On the other hand, the librarians evaluated the usefulness of the Room Occupancy Information, Report Generation, Summary of Reserved Spots, Change Effective Capacity, Library Information Management, and Account Management features. All of these have an average rating above 4, shown in Figure 17. The head librarian features all received a rating of 5 because the system was said to complied with all their suggestions. The other librarian rated all the features 4 except for the report generation which was rated 5. The respondent found this feature useful because the respondent oversees reports and data generation.

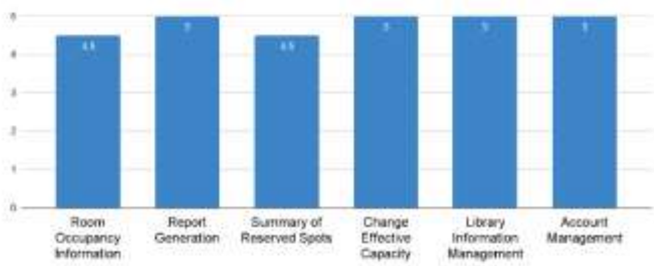

Figure 17 The average rating of each librarian feature

5.3.3 Comments and Suggestions

Most suggestions of the respondents were about how to improve the reservation feature of the system. The most common suggestion on the reservation feature is to implement a seat map. Using this feature, visitors may visually select which specific seat they want to reserve. Some have also suggested allowing visitors to reserve library computers. There was also a suggestion to allow visitors to reserve in groups so that visitors can use the CLIR for group studies.

There were also comments and suggestions about the interactive suggestions feature. One commented that the "peak hours" suggestion is very useful for them. However, some respondents thought that some of the messages were redundant and wordy. There was also one respondent who suggested including information and status about the current reservation in the interactive suggestions feature.

Some respondents pointed out how to improve the user interface elements of the website. These include decreasing the size of user interface elements, decreasing the font size, adding more graphics, and using the branding of MCL. Some respondents mentioned ways to improve the user experience. One respondent suggested that they should be able to see all the information on the visitor's home page while making a reservation. Another mentioned moving the reservation button on top of the website where it could be easily found.

6. Conclusion

In this paper, the researchers sought to create a reservation and computer vision-based room occupancy system called Sparse. Given this, the researchers had three objectives that they were able to accomplish. For the first objective, the researchers determined that the Faster R-CNN model is the most appropriate computer vision-based people counting algorithm for the system among Faster R-CNN, RetinaNet, and SSD. Then, in line with the second objective, the researchers identified the features of the system by collaborating with resource persons for visitors and librarians. The features include library information and management, room occupancy graph and information, interactive suggestions, seat reservation, change effective capacity, report generation, a summary of reservations, and librarian account management. For the last objective, the researchers were able to show that the features were useful through a survey.

This study has various implications. The proposed system would improve the experience of using the CLIR. It could be a model or inspiration for other libraries or public places that would like to implement a similar system in the future. They would have a better grasp of the features that would be useful and the challenges they would have in implementing such a system. Moreover, the researchers also compared different computer vision-based people counting algorithms. Other researchers in the computer vision field can use this as a reference, as the results, methodology, or recommendations can guide them.

However, there are still aspects of the system that future researchers could improve on. Future researchers could implement features that were suggested by the respondents of the survey. They may also focus on the user interface and user experience of the system. For the model, the researchers recommend using more powerful hardware for training, which would allow the model to be trained for more epochs and, therefore, reduce error. The researchers also recommend exploring other people counting methods, like other object detection models (YOLO, Mask R-CNN, etc.) or network-based methods.

The researchers also have suggestions for those who are looking to implement the system in the future. First, future implementers must consider how they are going to manage data collected by the system as the daily room occupancy data could occupy a lot of space in the database. Second, they must also consider possible concurrency issues as these may lead to a sluggish system. Lastly, they must consider the synchronization points of the system. If the people counting system were not synchronized properly with the reservation system, the reservation system may erroneously allow a visitor to make a reservation.

7. Acknowledgements

We would like to thank our resource persons for the development of the system. Ma'am Lady Diana Mendiola for sparing her time to give feedback on the system from a librarian's perspective. Also, Patrick Mediodia for giving feedback on our project from a visitor's perspective. Lastly, we would like to express our gratitude to our survey participants for giving insightful feedback on the system. Without them, the project would not be possible.

8. References

Amin, I., Taylor, A., Junejo, F., Al-Habaibeh, A., & Parkin, R. (2008). Automated people-counting by using low-resolution infrared and visual cameras. *Measurement*, 41(6), 589–599. https://doi.org/10.1016/j.measurement.2007.02.010

Beymer, D. (2000). Person counting using stereo. Proceedings Workshop on Human Motion, 127–133. https://doi.org/10.1109/humo.2000.897382

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. 2005 IEEE *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893. https://doi.org/10.1109/cvpr.2005.177

Department of Health. (2021) *What do we need to know under Alert Level 3?: Department of Health Website*. Department of Health Website. Retrieved January 8, 2022, from https://doh.gov.ph/node/33589

Lacanlale, E., Liang, S., Paglicawan, M., & Santos, J. (2021). *Measuring Customer Traffic Using Computer Vision. Philippine Statistics Authority*. Retrieved January 10, 2022, from https://psa.gov.ph/sites/default/files/5.4.1%20Measuring%20Customer%20Traffic%20Using%20Computer%20Vision.pdf

Lin, T., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, 2999–3007. https://doi.org/10.1109/iccv.2017.324

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. & Berg, A. (2016). SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016. ECCV 2016. *Lecture Notes in Computer Science*, 9905, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94

Nixon, M., & Aguado, A. (2019). *Feature Extraction and Image Processing for Computer Vision* (4th ed.). Academic Press. https://books.google.com.ph/books?id=KcW-DwAAQBAJ

Peng, D., Sun, Z., Chen, Z., Cai, Z., Xie, L., & Jin, L. (2018, August). Detecting heads using feature refine net and cascaded multi-scale architecture. In *2018 24th International Conference on Pattern Recognition (ICPR)*. 2528-2533. IEEE. https://doi.org/10.1109/icpr.2018.8545068

Punn, N. S., Sonbhadra, S. K., Agarwal, S., & Rai, G. (2021). Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *arXiv preprint arXiv:2005.01385v4*

Raghavachari, C., Aparna, V., Chithira, S., & Balasubramanian, V. (2015). A Comparative Study of Vision Based Human Detection Techniques in People Counting Applications. *Procedia Computer Science*, 58, 461–469. https://doi.org/10.1016/j.procs.2015.08.064

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

Salkind, N. J. (2010). Encyclopedia of research design (Vols. 1-0). *Thousand Oaks, CA: SAGE Publications, Inc*. doi: 10.4135/9781412961288

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1.* https://doi.org/10.1109/cvpr.2001.990517

Waitz Inc. (2020). [Mobile Application]. App Store & Play Store. Retrieved January 10, 2022, from https://waitz.io/