industrial internet CONSORTIUM

Physical Distancing and Crowd Density Monitoring Using Computer Vision

Authors:

Saurabh Mishra
Manager, Industry Consulting
Internet of Things
SAS Institute Inc.
Saurabh.Mishra@sas.com

Hamza Ghadyali
AI Specialist
AI and ML Ctr. of Excellence
SAS Institute Inc.
Hamza.Ghadyali@sas.com

Kedar Prabhudesai
AI Specialist
AI and ML Ctr. of Excellence
SAS Institute Inc.
Kedar.Prabhudesai@sas.com

Varunraj Valsaraj
Sr. Manager
AI and ML Ctr. of Excellence
SAS Institute Inc.
Varunraj.Valsaraj@sas.com

Hardi Desai
Sr. Associate ML Developer
AI and ML Ctr. of Excellence
SAS Institute Inc.
Hardi.Desai@sas.com

Ravi Shankar Subramanian
Software Developer, IoT
SAS Institute Inc.
RaviShankar.Subramanian@sas.com

INTRODUCTION

As the world welcomed 2020, the city of Wuhan in China was fighting with newly discovered coronavirus "COVID-19." Unfortunately, the timing of the emergence of this virus coincided with holiday season, with millions travelling around the world. The novel coronavirus turned out to be highly contagious and by early March, World Health Organization (WHO) declared the COVID-19 coronavirus as a pandemic[1]. Countries all around the world responded by imposing strict lockdowns restricting human interaction. While scientists are still working for development of immunization to this new virus, the world is trying to adjust to the new normal. A big part of COVID-19 response centers around safe return of employees and customers in public spaces such as retail shops, manufacturing plants, airports, cafeterias etc. This requires an ongoing need to monitor compliance with physical distancing and crowd density guidelines. To support business operations in these uncertain times, we propose a technology solution based on computer vision, streaming analytics and edge computing.

The objective of this project is to monitor physical distance compliance from video footage using deep learning and streaming analytics in a non-disruptive and anonymized manner targeted towards well defined spaces such as cafeterias, plant floors, retail checkout, airport security etc. Our solution surfaces insights via a real-time dashboard, an alerting mechanism, and a post-facto dashboard.

- The real-time component can be deployed on-premise, configured to suit the target organization's guidelines and provide immediate feedback via a real-time dashboard and push notifications to alert for urgent situations.
- The post-facto dashboard is meant to be deployed in a back-end environment such as the cloud and provides summarized feedback across multiple dimensions such as location, day of week and time of day. The dashboard supports interactivity to allow for what-if analysis and is a great tool to understand the efficacy of measures such as signage, one-way traffic etc. over a small period.

Few potential applications of the proposed solution are:

- **Monitoring public spaces:** Many public spaces already employ security cameras for surveillance. Utilizing the same setup, this solution can monitor physical distancing compliance in these public spaces.

[1] https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

- **Real-time alerts for urgent situations:** The real-time component supports the ability to generate immediate alerts by sending out e-mail and text messages for scenarios such as violation of the threshold for maximum number of people in a given space or not maintaining physical distancing for a long duration. Authorities can then take immediate steps to mitigate these urgent situations.
- **Detect hotspot "patterns":** Hotspots are the locations and time periods where people might find it difficult to practice physical distancing. The post-facto dashboard can help to quickly identify and stop the hotspot patterns with respect to time of the day, day of the week and by location. Identifying these hotspots can help authorities take appropriate steps to prevent them in future.
- **Cleaning Crew Guidance:** Identifying areas of maximum activity helps cleaning crews prioritize the area to focus while cleaning at the end of the day. Heatmap analysis can help to quickly classify the location by activity for the cleaning crew.

METHODOLOGY

In this work, we have designed a real-time physical distance monitoring system using computer vision and streaming analytics. Figure 1 provides a high-level overview of the system. At its core, the system detects people in each video frame and tracks their trajectory across frames while anonymizing their identity. Distances between people are then calculated to determine if they are maintaining a safe physical distance. The system sends real-time notifications to alert for violation of physical distancing and generates interactive visualizations to better monitor physical distancing. For real-time execution, the main components of the system are deployed on an edge server which is running the SAS® Event Stream Processing® (ESP) engine. This solution was developed using a combination of open-source software[2] and propriety SAS software for a surveillance camera located in a cafeteria on SAS campus, referred to as SAS café from here on.

Person Detection and Trajectory Tracking

Object detection is one of the fundamental research challenges in the computer vision community. This problem is observing continuous interest from the deep learning researchers to build accurate and faster models. A major breakthrough was observed with the You Only Look Once (YOLO) object detection algorithm[3] which achieved state-of-art accuracy while significantly improving inference speed, thereby allowing for real-time deployment.

[2] Bradski G. The OpenCV Library. *Dr Dobb's J Softw Tools*. 2000.

[3] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2016:779-788. doi:10.1021/je00029a022.

Our solution uses the Tiny-Yolo-V2[4] deep learning model to perform the task of person detection. We chose the Tiny- Yolo-V2 model because it has fewer layers and thus fewer parameters compared to Yolo-V2, allowing for fast inference on the edge server which is important for real-time deployment. For training the model, we annotated 5,000 video frames with a bounding box around each person.

The original architecture of Tiny-Yolo-V2 uses a combination of convolution and max-pooling layers to generate object detection feature maps of size 13 × 13 from an input image of size 416 × 416. We modified the original architecture to generate feature maps over a finer grid of size 26 × 26, to allow the model to detect people which appear small in the video frame. After training the model for several epochs and validating the person detection performance over a holdout dataset, we deployed the trained model on the edge server to detect people in each frame of the input video stream.

[4] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017*. 2017;2017-Janua:6517-6525. doi:10.1109/CVPR.2017.690.

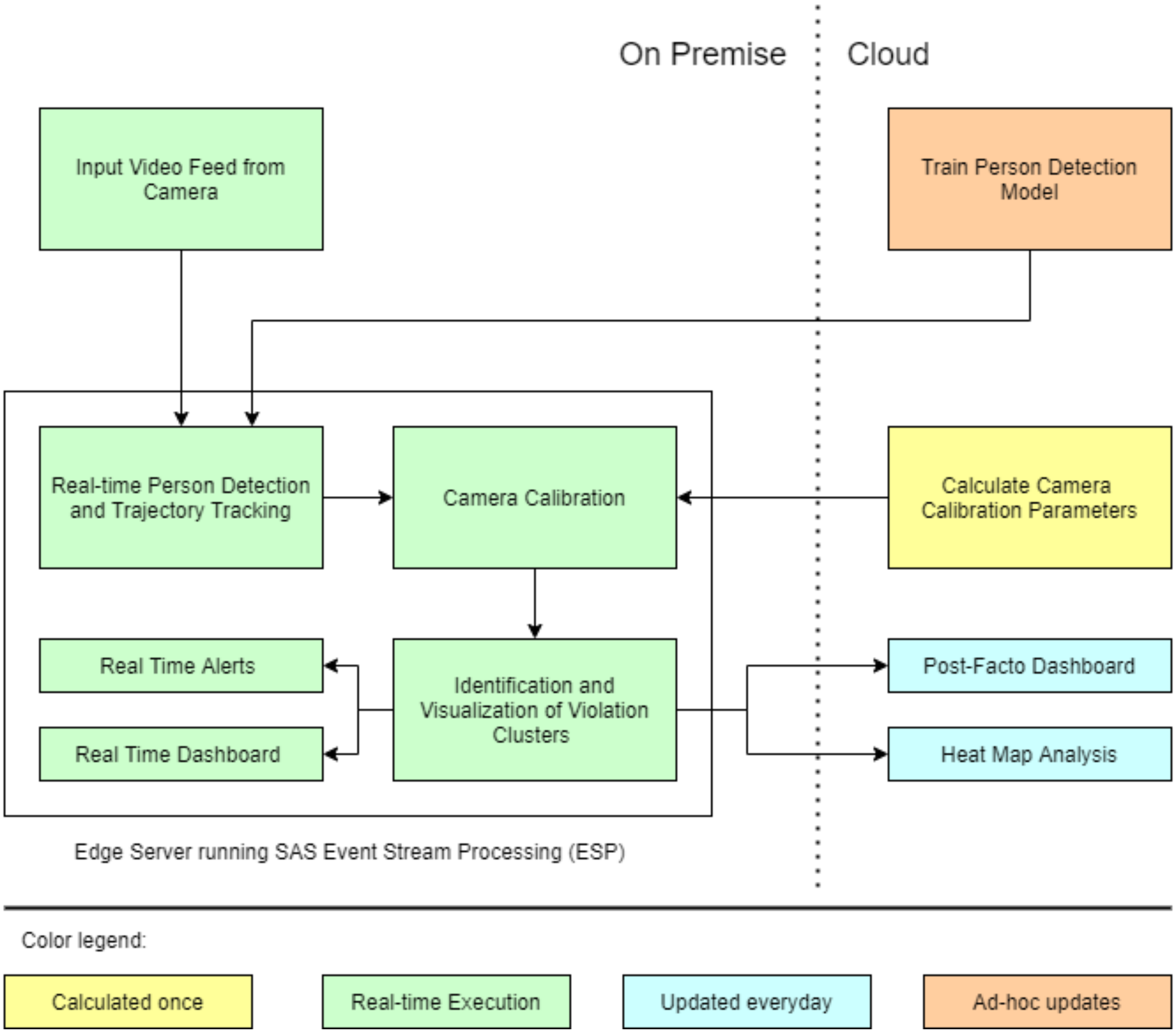

On Premise
Cloud
Input Video Feed from Camera
Train Person Detection Model
Real-time Person Detection and Trajectory Tracking
Camera Calibration
Calculate Camera Calibration Parameters
Real Time Alerts
Real Time Dashboard
Identification and Visualization of Violation Clusters
Post-Facto Dashboard
Heat Map Analysis
Edge Server running SAS Event Stream Processing (ESP)
Color legend:
Calculated once
Real-time Execution
Updated everyday
Ad-hoc updates

Fig. 1: A high-level overview of the physical distance monitoring system using computer vision and streaming analytics. The color legend indicates the frequency with which each component is updated. Camera calibration denoted by the yellow box is needed only once for each input video. Components in green boxes are updated continuously in real-time, components in blue boxes are updated daily, whereas the person detection model needs to be updated as per requirement.

The person detection model makes predictions to detect people in each frame independent of previous and future frames. This does not allow to track the trajectory of the movement of people across frames, which is crucial to monitor the movement of people. We use a high-speed tracking-by-detection approach[55] which is based on tracking people detections across frames using the intersection over union between detections in consecutive frames. Thus, the person detection results for each frame are provided as inputs to the tracking model and the tracker outputs a trajectory of movement for each person. The tracker provides a unique ID associated with each person's trajectory. These IDs are randomly generated by the tracking model and are not associated with the identity of the person such employee name or employee ID. Further, to ensure that the identity information is not compromised, our solution blurs the pixels in the bounding box of person detection before visualization. Thus, our solution detects and tracks people anonymously since it does not have any information about the identity of the people.

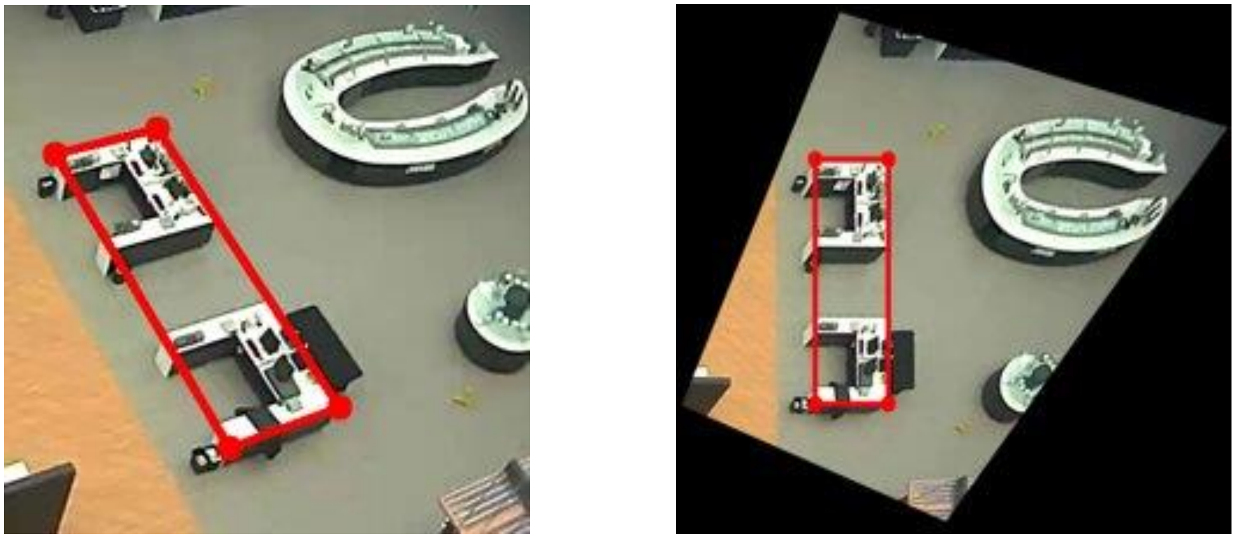
Distance Calculation via Camera Calibration

The deep learning model outputs the coordinates of a detected person relative to the input image and not relative to the physical space. To calculate distances between pairs of people we must first account for radial and tangential optical distortion[6] due to the camera and apply a transformation to the coordinates. In this example from our SAS café camera, the input image is taken as a crop from an area towards the center of the field-of-view of the camera and we observe no radial distortion. However, the ground is not parallel to the imaging plane, so we must correct for perspective distortion causing objects further away to shrink. We correct for this using a perspective transform[7] aligned with some reference markers in our physical space (See Figure 2). After applying the transform, we can approximate true physical distances between every pair of points.

[5] Bochinski E, Eiselein V, Sikora T. High-Speed tracking-by-detection without using image information. *2017 14th IEEE Int Conf Adv Video Signal Based Surveillance, AVSS 2017*. 2017;(August). doi:10.1109/AVSS.2017.8078516.

[6] Wolf L, Shashua A. On Projection Matrices Pk→P2,k=3,...,6, and their Applications in Computer Vision. *Int J Comput Vis*. 2002;48:53-67.

[7] Haralick RM. Using perspective transformations in scene analysis. *Comput Graph Image Process*. 1980;13(3):191-221.



(a)

(b)

Fig. 2: Camera calibration for distance calculation. a) The original image with four reference markers drawn as the red polygon around the cashier stations. b) The perspective transform applied relative to these markers rectifies the image, correcting for perspective distortion. This allows us to map pixel-distances to true physical distances.

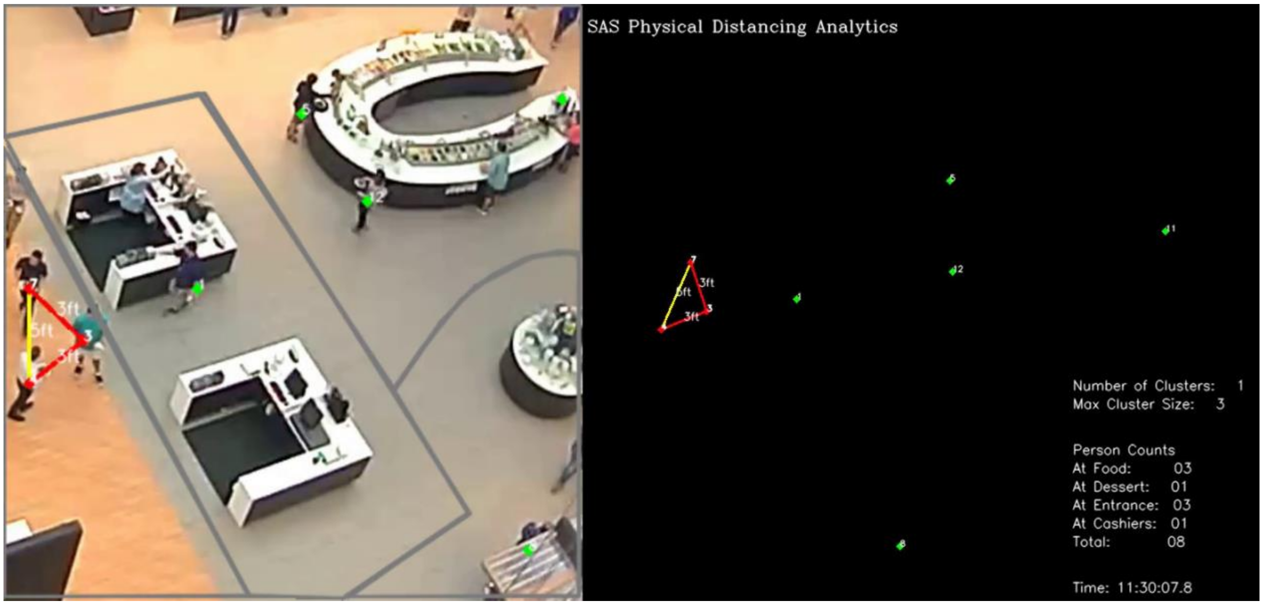
Clustering

After we perform person detection on a per-frame by basis and person-tracking across frames, we can monitor the formation of groups and clusters and track them over time. Combined with the video calibration to convert pixel distances within the image to true physical distances in the physical space, we can detect when any number of people are involved in a physical distancing violation. In this project, we have used computer vision algorithms to perform person detection and tracking from video-input, but the methodology for monitoring groups and clusters only depends on the positions of each person over time. If other input sensors are used to detect positions of people in the space over time, the subsequent analytics for physical distancing still hold.

At each frame, we use the position of every detected person and compute all pairwise distances. These distances are true physical distances. We then apply single-linkage clustering[8] to connect any two people that are within a threshold, such as 6 feet or 2 meters. The visualization on Figure 3 shows a dot for each detected person as well as edges between any pair of people who are within 6 feet of each other. We also use color to show when people are close or dangerously

[8] Gower JC, Ross GJ. Minimum spanning trees and single linkage cluster analysis. *J R Stat Soc Ser C (Applied Stat*. 1969;18(1):54-64.

close (yellow or red, respectively) versus people that are isolated (green dots). This results in a graph where we can consider the connected components of the graphs to be clusters that need to be monitored and tracked over time. We can also keep track of the number of clusters and the size of each cluster over time. As part of our summary analysis we can focus our attention on clusters of a given size that persist for longer periods of time between the same people as shown in Figure 3 below.



SAS Physical Distancing Analytics

```
Number of Clusters:   1
Max Cluster Size:   3

Person Counts
At Food:        03
At Dessert:    01
At Entrance:   03
At Cashiers:   01
Total:          08

Time: 11:30:07.8
```

Fig. 3: Real-time visualization of person-detection, tracking, and geofence counting, and clustering. On the left, we have the original camera footage with gray lines depicting user-defined geofences. In this example, we have four geofences corresponding to the food stations (top-right), the dessert area (lower-right region), the cashiers (rectangular region in the middle), and the entry way (lower-left). A dot is placed on each detected person. After applying a perspective transform for camera calibration, we show on the right the location of people and the estimated distances between two people when they are closer than 6 feet apart and draw an edge between them. The edge is colored red if their distance is less than 4 feet, and colored yellow if the distance is less than 6 feet. On the bottom right, we summarize information about the clusters and the density of people in each geofence.

Density Analysis with Geofences

We also perform a density analysis with user-defined geofences. The user defines regions of interest along with capacity limits (optionally) for each region. Our solution counts the number of people detected in each region at every time-point. On the left half of Figure 3, we show our implementation in the example of our SAS café camera. The gray curves show the user-defined geofences. In this example, we have four geofences corresponding to the food stations (top-right), the dessert area (lower-right region), the cashiers (rectangular region in the middle), and the entry way (lower-left). We also send out real-time alerts when capacity limits have been exceeded using ESP which is discussed in the next section.

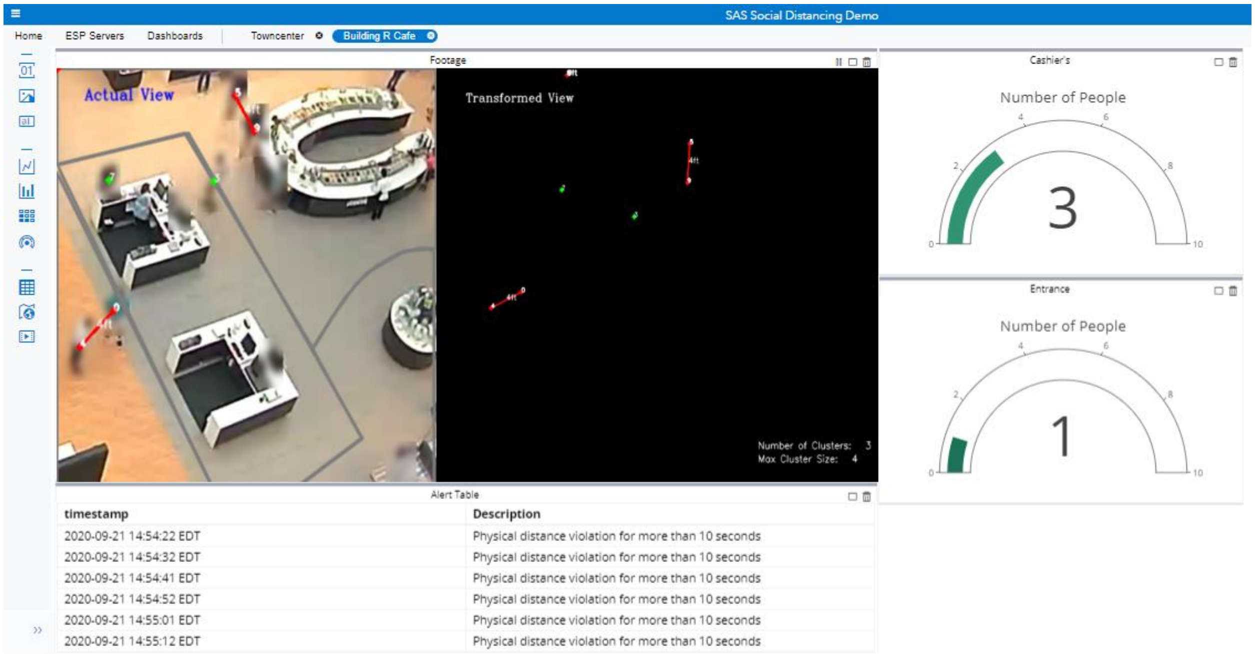
Event Stream Processing

SAS® Event Stream Processing® (ESP) enables real-time execution to analyze many continuously flowing events. ESP can be deployed on edge servers, on-premise servers or in the cloud. In this project, ESP was used for the following:

- Read the streaming images from the IP camera
- Pre-process the input image streams
- Detect person using the pre-trained computer vision model
- Track the detected people across frames
- Perform post-processing to calculate distances between the detected persons
- Define geofences and track people in them
- Send alert notifications via email or SMS to the admin in case of any violation
- Feed the data for both real-time streaming dashboard and post-facto dashboard

Real Time Analytics

Figure 4 displays the contents of the real time streaming dashboard which is populated by the data coming from the ESP engine. The dashboard has various components such as streaming video from the ESP server with markers indicating the distance between two individuals. The image viewer also contains the transformed view which is the image after applying perspective transform. With streaming running in real time, we also have live analytics such as the total number of individuals in each geofence at every point of time. The alert table in the dashboard has the list of violation alerts generated by ESP server.



timestamp	Description
2020-09-21 14:54:22 EDT	Physical distance violation for more than 10 seconds
2020-09-21 14:54:32 EDT	Physical distance violation for more than 10 seconds
2020-09-21 14:54:41 EDT	Physical distance violation for more than 10 seconds
2020-09-21 14:54:52 EDT	Physical distance violation for more than 10 seconds
2020-09-21 14:55:01 EDT	Physical distance violation for more than 10 seconds
2020-09-21 14:55:12 EDT	Physical distance violation for more than 10 seconds

Fig. 4: ESP Dashboard with real time visualizations of physical distance violation clusters displaying the results in an anonymous fashion by blurring the individuals in live streaming of camera feed. In the middle, we view the violations clusters with respect to its mapping to true physical world. On the right side, there are gauges showing the count of people in each geofence. The bottom part of the dashboard has the alert table which has the list of alerts along with the timestamp of the alert origination and the alert description.

VISUALIZATIONS

Heatmap Visualizations

We use heat maps superimposed on the original images to answer a few different questions. The first heat map is generated by aggregating the positions of every_detected person over a given time interval. In our examples we use a time interval of 5 minutes which allows us to study how the heatmaps change over time. This heat map on Figure 5 allows us to answer questions about overall traffic flow and can be used as an indication of where cleaning crews might need to spend additional time due to heavier traffic.

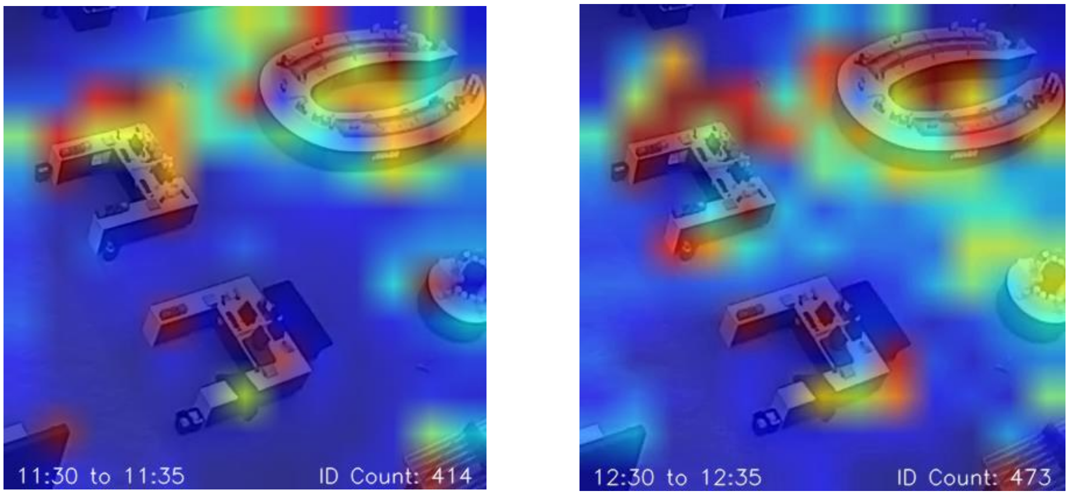
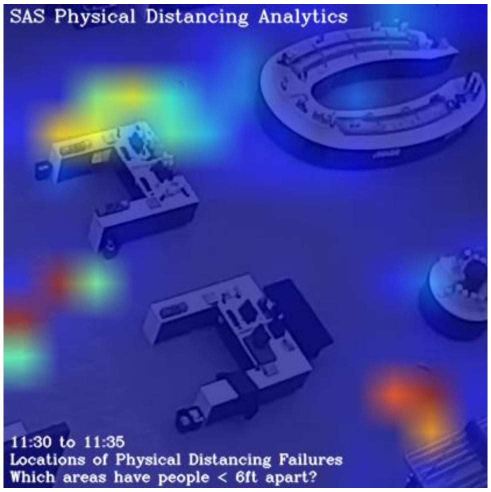


Fig. 5: This first set of heatmaps summarizes the locations of every detected person aggregated over 5-minute intervals, answering questions about the areas of highest traffic at different times. On the left, we aggregate over 11:30am to 11:35am whereas on the right, we aggregate over 12:30pm to 12:35pm. We observe different spatial patterns of activity between the two-time intervals, such as the increased traffic around the cashiers near the bottom of the image. We also observe similarities such as the high traffic around the horse-shoe shaped salad bar area or the primary cashier station at the top-left. In practice, we build heatmaps over all time-intervals of interest and animate to show their continuity.

Sometimes, despite our best efforts, there can be areas in an environment where it is nearly impossible to distance, such as a narrow hallway, or a blocked sidewalk, and so on. A well-designed space can allow high traffic while maintaining physical distancing. But at the other end of the spectrum, a poor-design can thwart our best efforts to physical-distance. The first heat map only studies the overall traffic volume and cannot answer where physical distancing is failing.


SAS Physical Distancing Analytics
11:30 to 11:35
Locations of Physical Distancing Failures
Which areas have people < 6ft apart?

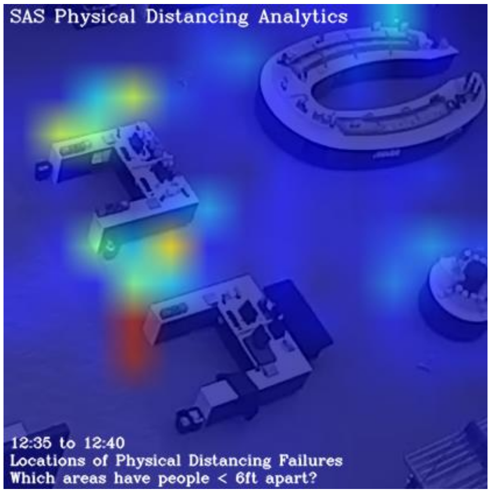

SAS Physical Distancing Analytics
12:35 to 12:40
Locations of Physical Distancing Failures
Which areas have people < 6ft apart?

Fig. 6: This second set of heatmaps summarizes the locations of physical distancing failure, also aggregated over 5-minute intervals. This helps answer which areas have people involved in a physical distancing violation. On the left, we aggregate over 11:30am to 11:35am whereas on the right, we aggregate over 12:35pm to 12:40pm. We observe different spatial patterns between the two-time intervals, such as the increased volume of violations in the area between the two cashiers. We also observe similarities such as the consistent violations near the primary cashier station (located at the top-left). In practice, we build heatmaps over all time-intervals of interest and animate to show their continuity.

In contrast, our heatmap on Figure 6 is specifically designed to answer the question “What are the locations of physical distancing failures?” (i.e. Which areas have people less than 6 feet apart?) This is computed by restricting attention to only the positions of people that are involved in a physical distancing failure during the failure. This question is framed to understand whether there are features in the environment that are leading to physical distancing failures.

This analysis is performed with the highest level of privacy since the focus is not on any given individual but rather on the physical environment instead. This summary can be used to redesign the space, clear blockages, or remove obstacles to make it easier for everyone to keep their distance. In the second figure, we can see hot spots were physical distancing is frequently failing such as near the checkout areas or the area in the bottom-left corresponding to a food station. One advantage of studying these heat maps over time is that we can examine both where physical distancing failures occur and when. Our solution generates a heat map for each time interval and stitches them together into an animation video for qualitative analysis and human review.

Interactive Visual Report using SAS® Visual Analytics®

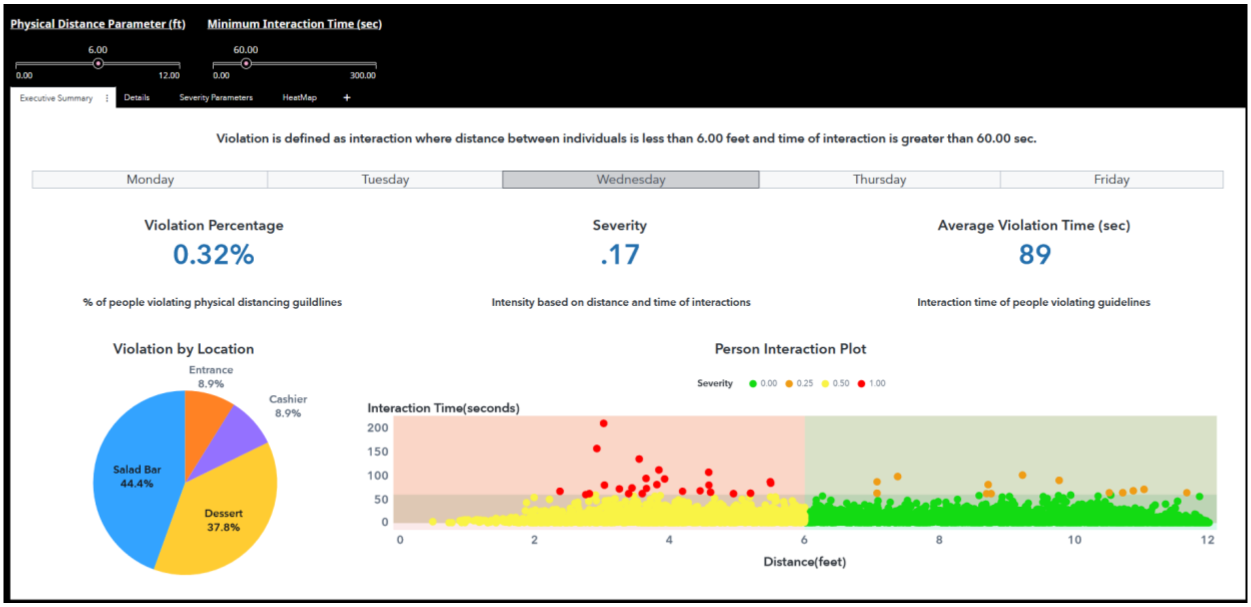
Maintaining physical distancing is one of the most effective methodologies for combating COVID-19. However, physical distance guidelines vary from country to country. For example, in the UK, the current guidelines mandate a physical distance of two meters[9], while in the US, the CDC[10] advises physical distance of six feet (1.8 meters). Australia and Germany have physical distance guideline at 1.5 meters, and France advocates one meter as its official guideline. The World Health Organization recommends people stay at least one meter apart. The risk of transmission halves for every extra meter of distancing up to three meters[11]. Given the variability in guidelines followed by different countries, we have built an interactive dashboard, where the user can quantify the violation by specifying the permissible physical distance as well as interaction time parameters.

The post-facto dashboard shown in Figure 7 provides an executive summary of the physical distancing violations. This dashboard gives a summary of the percentage of people whose interactions result in a violation, the severity of these violations, as well as the average interaction time during a violation based on user specified parameters. The severity of a violation is also computed based on user defined parameters. It is obvious that no interactions are the same. A person having more interaction time and less physical distance is more at risk of transmitting the corona virus than a person spending less interaction time and maintaining enough physical distance. Keeping this in mind, we have built an interaction plot providing a visual understanding of various interaction between people in the video footage.

[9] https://www.gov.uk/coronavirus.

[10] https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html.

[11] https://www.technologyreview.com/2020/06/02/1002493/two-meter-distancing-halves-risk-compared-to-one-meter-suggests-a-new-review/.



Physical Distance Parameter (ft) 6.00

Minimum Interaction Time (sec) 60.00

Violation is defined as interaction where distance between individuals is less than 6.00 feet and time of interaction is greater than 60.00 sec.

Violation Percentage 0.32%

% of people violating physical distancing guildlines

Severity .17

Intensity based on distance and time of interactions

Average Violation Time (sec) 89

Interaction time of people violating guidelines

Fig. 7: A screenshot of the post-facto dashboard providing an executive summary analysis of physical distancing violations. The sliders at the top allow the user to define a violation by selecting distance and interaction time parameters. Based on the selected parameters the violations are summarized using the violation percentage, severity of violation and average violation time. The pie-chart gives a breakdown of the violations for each geofence specifying a different region. Finally, the scatter plot shows an interaction plot for each person which is segmented into four regions based on the threshold parameters set for distance and interaction time.

The dashboard in Figure 8 provides details about the violations in different locations and its distribution over different times of the day. It is important to note, that each of the plot can be filtered using day of the week and location to get in-depth detail insight about the situation. The location-based analysis chart not only provides the user insight about the violation percentages but also its comparison total person visiting the location.

Bar/Line chart on the left indicate the customer arrival (%) and violation (%) by day of the week. There is a positive correlation between number of customer and violation (%). For example, Tuesday, Wednesday & Thursday are high-volume days at SAS café, and we see higher physical distancing violation on these 3 days. On the right, we have a tree map showing the areas of violation. The size of tree map indicates the location where people most of their time in SAS café while colors indicate violation (%). Red color locations have higher violation (%) while the green color locations have lower violation (%). For example, salad bar is the area where people spend most of the time, but dessert is the area with most per-capita violation (%). This makes sense as people are continuously moving in salad area while people tend to gather around desert and usually take longer to decide what they would like to choose. We can also see violation by time of the day at the bottom. There are few counter-intuitive trends here. We would have expected peak violation during peak café times that is between 11:45 am to 12:30 pm. Surprisingly, peak violation (%) is during 1 pm to 1:30 pm time slot.

Operationalization

Typical operationalization involves setting up the real-time component in an on-premise environment and the back-end component in cloud:

- **On-premise component:** This includes the SAS® Event Stream Processing® engine configured with a deep learning model and real-time dashboard. Setup includes configuring connectivity to Video cameras generally via a Video Management System (VMS) for inbound data. The target environment is typically an Edge server with GPUs to process the video analytics workload. It is possible for a single edge server to support multiple cameras depending upon the compute available. Additional edge servers can be added to handle a high number of cameras.
- **Back-end component:** This is the SAS Visual Analytics® application deployed in a data center or cloud environment. Since the videos are processed on-premise, the payload and storage requirements are modest and can be easily scaled.

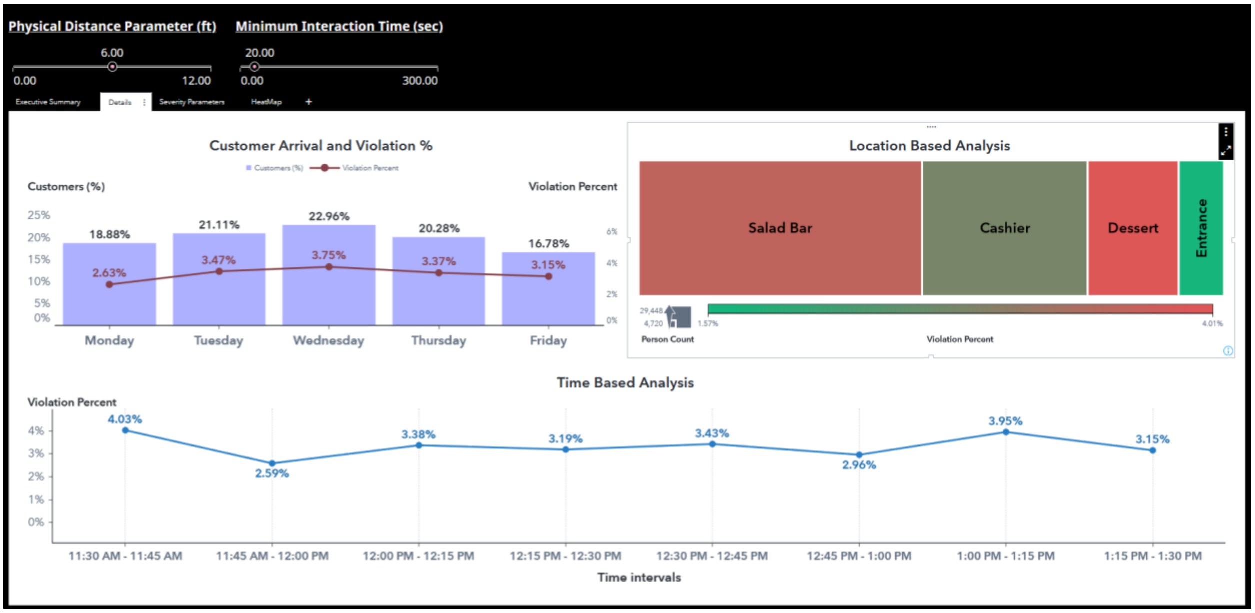

Physical Distance Parameter (ft): 6.00
Minimum Interaction Time (sec): 20.00

Fig. 8: An interactive dashboard for time and location-based analysis. The bar char on the top left provides the distribution of customer arrival and violation (%) based upon the day of the week. It is observed that more customers arrive on middle of the week compared to start or end of the week. Violations follows the same trend. The tree chart on the top right gives insight about number of individuals visiting a give geofence and the observed violations in the location by the size of the section and the color of section respectively. While more individuals have visited the cashier section compared to the dessert section, the violation percentages are higher

in dessert area indicating that more crowd is following the guidelines in cashier section compared to dessert section. Lastly, the line chart in the bottom of the dashboard demonstrates the trend in violations over time grouped in 15 mins intervals during the lunch period. In the dashboard the user can narrow down by selecting day of week and location of interest to observe the trends.

CONCLUSION

Use of computer vision and streaming analytics with an existing camera infrastructure can yield immediate benefits in terms of monitoring physical distancing and crowd density monitoring guidelines. The fact that this solution supports anonymity and does not have any disruptive requirements on individuals (in terms of tracking devices etc.) makes it viable across several industry verticals and associated public spaces. Additionally, it provides a low barrier to entry due the ability to interoperate with existing camera infrastructure (such as security cameras).

Healthy and safe workspaces are critical to our economy now more than ever. Technology aided solutions such as the one described here are an intrinsic part of the strategy to achieve this goal and do it at scale.

➢ Return to IIC Journal of Innovation landing page for more articles and past editions

The views expressed in the *IIC Journal of Innovation* are the contributing authors' views and do not necessarily represent the views of their respective employers nor those of the Industrial Internet Consortium.