Physical Distancing Monitoring with Background Subtraction Methods

Hendra Adinanta
Research Center for Physics,
Indonesian Institute of Sciences
Tangerang Selatan, Indonesia
hendra.adinanta@lipi.go.id

Edi Kurniawan
Research Center for Physics,
Indonesian Institute of Sciences
Tangerang Selatan, Indonesia
edi.kurniawan@lipi.go.id

Suryadi
Research Center for Physics,
Indonesian Institute of Sciences
Tangerang Selatan, Indonesia
suryadi@lipi.go.id

Jalu A. Prakosa
Research Center for Physics,
Indonesian Institute of Sciences
Tangerang Selatan, Indonesia
jalu.ahmad.prakosa@lipi.go.id

***Abstract*—World Health Organization (WHO) has confirmed that the spreading of coronavirus disease 2019 (COVID-19) could be avoided by keeping the physical distance at least 3 feet (1 meter). Then, we have a motivation to employ computer vision techniques to monitor social distancing violations. The principle of the works are to detect persons, then to assess the physical distancing violation from their distance. Most of the researchers have tried to utilize object detection methods such as faster RCNN, Yolo, and SSD to detect persons from the frame. Those methods rely on, the support of Graphics Processing Unit (GPU) to execute their heavy computation. In this works, we propose social distancing monitoring by applying background subtraction methods based on Gaussian Mixture Models (GMM) i.e. Geometric Multigrid (GMG), k-Nearest Neighbor (KNN), Mixture of Gaussian (MOG), and Mixture of Gaussian 2 (MOG2). These methods have been used to filter persons from the frame with computational process. Some parameters evaluation measures have been determined to check the best method suitable for this works. In terms of performance, better methods are ranked as KNN, MOG, MOG2, and GMG.**

***Index Terms*—social distancing, COVID-19, background subtraction, monitoring**

I. Introduction

COVID-19 outbreak has been reported at the end of 2019. And, World Health Organization (WHO) per July 18, 2020 confirmed that the COVID-19 pandemic has spread to over 216 countries with 13,876,441 positive cases including 593,087 deaths [1]. To this date, many health institutions across the globe put their efforts to develop medicines and vaccines to combat this virus. One of the alternatives to decrease the transmission of this infectious virus is by implementing social distancing on public places.

Social distancing can reduce the infection rate and delay the peak of cases if it is appropriately implemented at the early of this pandemic. This will reduce the burden of the healthcare systems and also lower the death rate [2]. Fig. 1 (a) [3] figures the effects of social distancing measures in reducing the plague peak and matching the health care system capacity. The effect of social distancing can reduce the number of infected cases and the earlier implementation as shown in Fig. 1(b) [3].

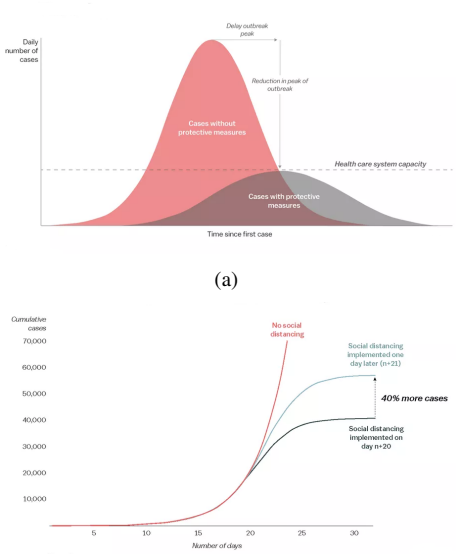

Fig. 1. Illustration of benefits on controlling social distancing in the outbreak.

WHO announces that maintain at least 1 meter (3 feet) distance between one person to another to avoid virus transmission via droplet [4]. This regulation is supported by affected countries to mitigate the coronavirus pandemic with minimum economic loss.

Many governments have been implemented social distancing during the COVID-19 pandemic, they make policies to restrict some aspect as travel, territorial, and public area. In this case, large-scale measures are difficult to implement, because not all public places can be closed, persons still need to go outside to fulfil their daily needs. In this context, technologies

are needed in facilitating social distancing monitoring. To measure social distancing, there are several technologies which can be applied, e.g. AI [5], thermal [6], computer vision with deep learning [7], ultrasound [8], and visible light [9].

In this paper, we propose computer vision-based social distancing monitoring by using background subtraction method. This method has potency to measure and detect persons position and measure the distance of each other. This method also offers low computational process so the need of an additional hardware such as GPU is unnecessary. In this works, we compared background several subtraction method such as Geometric Multigrid (GMG), k-Nearest Neighbor (KNN), Mixture of Gaussian (MOG), and Mixture of Gaussian 2 (MOG2).

II. Background Subtraction Methods

The background subtraction method is a simple technique to get a region of interest, in this case a moving object, by subtracting an image with an object from the background image. The BS method comprises of five main steps i.e (1) model the background, (2) extract the foreground, (3) change the detection, (4) detect the foreground, and (5) detect the motion [10]. Study [11] explain that a pixel position (x,y) in the present frame I_t as foreground if

$$|I_t(x,y) - B_t(x,y)| > T \tag{1}$$

where T is a determined threshold and B_t is a background image. The background image B_t is updated by using an Infinite Impulse Response (IIR) filter as follow

$$B_{t+1} = \alpha I_t + (1-\alpha) B_t \tag{2}$$

where α is the adaptation rate between the current model with the observation.

Background subtraction using Gaussian Mixture Model (GMM) is a widely used approach for foreground detection by an average, a covariance matrix, and a prior probability of each K Gaussian characterization. Every pixel X_t and K Gaussian distributions correspond with a certain pixel $\{x_0, y_0\}$ probability, at time t, where its history describe by [12].

$$\{X_1, \cdots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \tag{3}$$

which I is the image sequence.

The present history of every pixel, $\{X_1, \cdots, X_t\}$ is modeled with K Gaussian distributions mixture. The possibility of observing a pixel of intensity X_t in the present image can be calculated with [12]

$$P_x(X_t) = \sum_{i=1}^{K} w_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \tag{4}$$

where K is distribution amount, $w_{i,t}$ is weight estimation, $\mu_{i,t}$ is the mean value, $\Sigma_{i,t}$ is a covariance matrix of the i-th Gaussian mixture at time t respectively, and η is the density of Gaussian probability.

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma (X_t - \mu_t)} \tag{5}$$

K is decided by the computing power assumed with K = 4 in [13] and K = 3 to 5 in [12], the covariance matrix to be

$$\Sigma_{k,t} = \sigma_{k,t}^2 I \tag{6}$$

The background is regenerated to be

$$\omega_{k,t} = (1-\alpha)\,\omega_{k,t-1} + \alpha\,(M_{k,t}) \tag{7}$$

$$\mu_t = (1-\rho)\,U_{t-1} + \rho X_t \tag{8}$$

$$\sigma_t^2 = (1-\rho)\,\sigma_{t-1}^2 + \rho\,(X_t - \mu_{i,t})^T\,(X_t - \mu_t) \tag{9}$$

$$\rho = \alpha\eta\,(X_t | \mu_k, \sigma_k) \tag{10}$$

where $\omega_{k,t}$ is the prior weights of the K distributions at time t, α is the learning rate, $M_{k,t}$ is 1 for the matching models and 0 for the remaining models. The μ and σ are parameters for mismatched distribution abide the same and updated as μ_t and σ_t. The second learning ρ is an equally effective type with causal of the low-pass filter.

Several methods of GMM are GMG, KNN, MOG, and MOG2 [14]. In this research, those 4 methods will be compared by using a computer vision library in OpenCV.

A. Geometric Multigrid (GMG)

This algorithm combines statistical background image estimation and per-pixel Bayesian segmentation that introduced by [15]. They employ a probabilistic foreground segmentation algorithm that identifies possible foreground objects using Bayesian inference.

B. k-Nearest Neighbor (KNN)

In study [16], the KNN method was presented by used recursive equations to constantly update the Gaussian mixture model parameters and simultaneously selects the appropriate number of components for each pixel so that the implementation of the KNN method more optimal.

C. The Mixture of Gaussian (MOG)

In the works [17], In the works [17], MOG was presented with a method that optimizes the adaptive background mixture model by reinvestigating the renew equations and exploit distinct equations at distinct phases.

D. The Mixture of Gaussian 2 (MOG2)

Zivkovic et al. [18] presented an adaptive algorithm by using the Gaussian mixture model probability density. They used recursive equations to invariably regenerate the parameters to pick the appropriate number of components for every pixel.

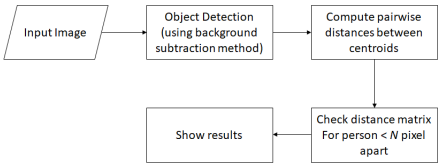

Fig. 2. The steps of social distancing monitoring using background subtraction methods.

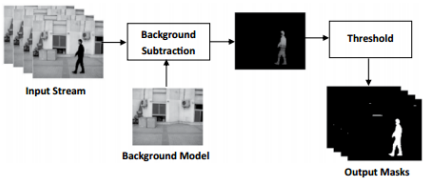

Fig. 3. Background subtraction process [23].

III. Experimental Setup

In this article, we use BS technique based on the Gaussian mixture for detecting pedestrians to develop social distancing monitoring system. We compare 4 methods that are GMG, KNN, MOG, and MOG2. The major benefit of this method is all processes are done by taking gray-scale frames so that no need for more high-performance computing hardware because the transmission of gray-scale video stores immense bandwidth and time. Thus, these methods are used for object detection and tracking models for social distancing detector implementation. The steps and BS process of social distancing monitoring are shown in Fig. 2 and 3 respectively.

A. Detail Process

As shown in Fig. 2, the necessary steps social distancing monitoring are explained as follows.

1) Convert each image sequence including a walking figure from the input video stream into an associated temporal sequence of distance signals at the preprocessing stage. This is done training and recognition from background subtraction process shown in Fig. 3.
2) Determine the stationary background image as a background model of shifting objects from the distinction between the current frame and reference frame.
3) Detect shifting regions by reducing the present image pixel-by-pixel from a reference background image where the distinction over the threshold are classified as foreground.
4) Perform erosion, dilation, and closing to reduce the noise, the reference background is updated to adapt dynamic scene changes.
5) Identify the scattered foreground regions by blob extraction in detected foreground mask. This blob extraction is marked by rectangle with the centroid at (c_x, c_y)

$$c_x = (x + (w/2)) \tag{11}$$

$$c_y = (y + (h/2)) \tag{12}$$

where c_x is center position of pixel coordinate of x added by width of rectangle w divided by 2, while c_y is center position of pixel coordinate of y added by height of rectangle h divided by 2.

6) Identify the pairwise distance in pixels between two centroids [19], where the actual distance can be determined from the focal length of camera F [20]

$$F = \frac{P \times D}{W} \tag{13}$$

where P is apparent width of taken picture in pixel, D is actual distance of camera to the object, W is width of the object. After getting the focal length of the camera, we can apply the triangle similarity to determine the distance of the object to the camera D' with

$$D' = \frac{W \times F}{P} \tag{14}$$

7) Define D' as the minimum distance (in pixels) parameter that person must stay from each other in order to follow social distancing protocols.
8) Compute the Euclidean distances between all pairs of the centroids with

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{15}$$

where p, q two points in Euclidean n-space, p_i, q_i Euclidean vectors, n is n-space.

9) Mark each person with bounding box and centroid. The persons who are in a safe distance will have green bounding box, while those who are in unsafe distance will have red bounding box.

B. Evaluation Measures

Most of machine learning techniques and statistical classification use a confusion matrix to assess the thematic accuracy of a land-cover map and numerous accuracy measures have been proposed for summarizing the information contained in this error matrix [22]. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class or vice versa [23]. It is a special kind of contingency table with two dimensions, actual and predicted. Identical sets of classes in both dimensions, each combination of dimension and class is a variable in the contingency table.

In this research, we wrap up the confusion matrix function with the input as the list of elements with binary elements 0 and 1 as shown in Fig. 4.

The after processing BS algorithms using MOG, MOG2, GMG, and KNN in video capture are compared with the ground truth which are manually generated. The ground truth is required to validate the accuracy and to validate the other metrics to see how well the images are segmented. The ground truth is available for all videos constituting the database allowing the evaluation of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) numbers.

		Ground truth	
		Foreground (255, vessels)	**Background (0, tissue)**
Predicted	**Foreground (255, vessels)**	True Positive (TP)	False Positive (FP)
	Background (0, tissue)	False Negative (FN)	True Negative (TN)

Fig. 4. Confusion matrix of image segmentation subtraction method.

From mentioned parameters above, we calculate the following measures [21]:

1) Sensitivity, recall, hit rate, or true positive rate (TPR)

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

2) Specificity, selectivity or true negative rate (TNR)

$$Specificity = \frac{TN}{TN + FP} \tag{17}$$

3) Miss rate or false negative rate (FNR)

$$FNR = \frac{FN}{FN + FP} = 1 - TPR \tag{18}$$

4) Fall-out or false positive rate (FPR)

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{19}$$

5) Precision or positive predictive value (PPV)

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

6) Matthews correlation coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{21}$$

7) Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

8) Percentage of Wrong Classifications (PWC)

$$PWC = \frac{FP + FN}{TP + FN + FP + TN} \times 100 \tag{23}$$

9) F- Measure

$$F - measure = \frac{2 \times Recall \times Precission}{Recall + Precission} \tag{24}$$

The overall algorithms were executed using computer with core processor i5-4300U CPU at 1.90 GHz, 8 GB of RAM. The video in this experiment uses a CCTV video of pedestrians in Oxford [22]. The video was recorded with a frame size of 700 x 392 pixels in mp4 format with OpenCV and python tools.

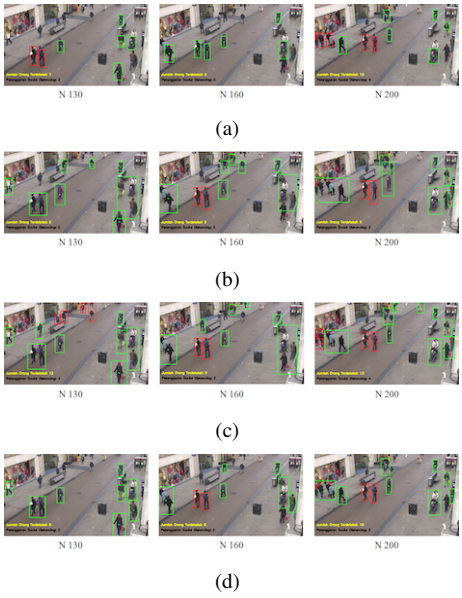

Fig. 5. The output of physical distancing monitoring using (a) MOG, (b) MOG2, (c) GMG, (d) KNN .

IV. Result and Discussion

Based on the above experimental scheme, several main parameters for the demonstration of physical distancing monitoring program have been set including a minimum pairwise distance of 50 pixels and a minimum threshold detected area of 500 pixels. These parameters were applied with the same conditions for proposed methods, namely MOG, MOG2, GMM, and KNN.

The output frame is shown in Fig. 5 shows that the output of physical distancing monitoring using proposed methods of MOG, MOG2, GMG, and KNN in the sampling frame of 130, 160, and 200. The output shows the detection frame that identified the person with rectangle boundary and centroids. It is seen that the person detection of each proposed method was different. Some rectangular boundary has not detected the person and sometimes detect several person in one rectangular boundary.

Detail information about the number of detected persons and violations can be seen in Fig. 6. From this figure, the differences in recognition results seem shown from each method. The highest average person detection can be shown by the KNN, MOG2, GMG, and MOG method with 9.00, 8.67, 7.49, and 5.74 respectively. Then, the average detected violations are KNN, GMG, MOG2, and MOG with 2.37, 1.72, 1.54, and 1.49 respectively. It indicates that the best detection accuracy was handled by the KNN method.

In the analysis, we use the ground truth to quantify the performance of a segmentation algorithm. We begin with a ground truth data set, which has been manually segmented using image editing software Adobe Photoshop. We compare ground truth with the predicted binary segmentation, showing accuracy alongside more effective metrics. Table I shows the comparison of used methods with the ground truth. This

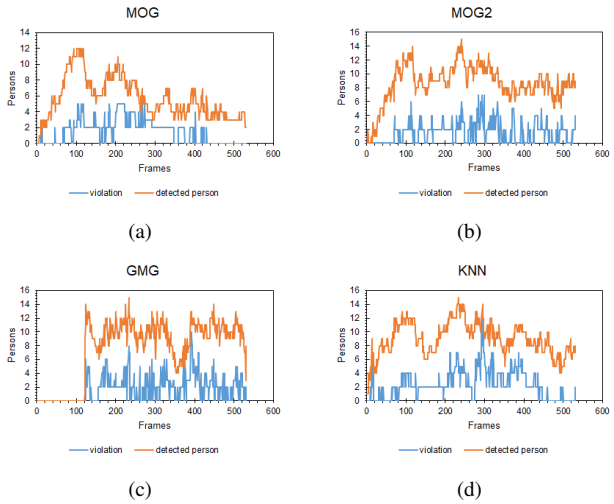

Fig. 6. The output of detected person and violation using (a) MOG, (b) MOG2, (c) GMG, (d) KNN.

TABLE I
EXAMPLE OF IMAGE DATA SET

Frames	Ground Truth	MOG	MOG2	GMG	KNN
130					
160					
200					

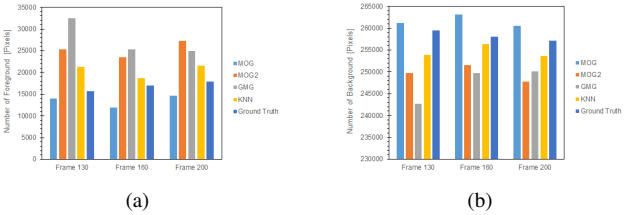

Fig. 7. Comparison of proposed method with the ground truth for (a) number of foreground, (b) number of background

comparison aims to check the number foreground and the background by using pixel-based, see Fig. 7. From this graph, the best result is the MOG which has the nearest value with the ground truth of 3332.33 average number of pixels. For deeper analysis, we calculate the average number of recall and precision, MCC and PWC, also accuracy and F-measure in Fig. 8, Fig. 9, and Fig. 10 respectively. The average score of TPR, TNR, FPR, and FNR more detailed shows in Table II.

V. CONCLUSION

In this paper, we evaluated the performance of the proposed method to achieve mask extraction quality of the foreground for person detection in physical distancing monitoring. Some parameters evaluation measures have been determined to check the best method suitable for this project. From the summary

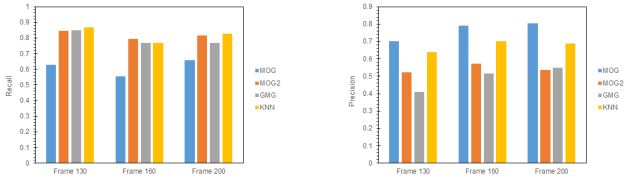

Fig. 8. The results of the proposed method for (a) Recall, (b) Precision

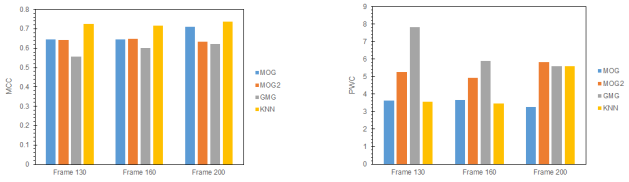

Fig. 9. The results of the proposed method for (a) MCC, (b) PWC

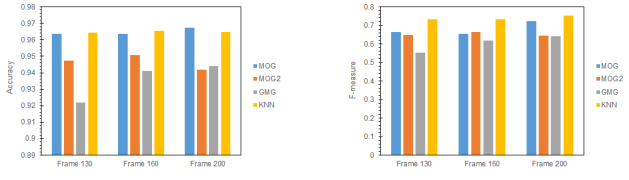

Fig. 10. . The results of the proposed method for (a) Accuracy, (b) F-measure

TABLE II
AVERAGE SCORE OF TPR, TNR, FPR, AND FNR

Methods	Average from Frame 130, 160, and 200			
	TPR	TNR	FPR	FNR
MOG	0.6139	0.9877	0.0123	0.3861
MOG2	0.8187	0.9550	0.0450	0.1813
GMG	0.7957	0.9450	0.0550	0.2043
KNN	0.8215	0.9742	0.0258	0.1785

TABLE III
SUMMARY RESULT OF PROPOSED METHOD BY STAR POINT

Measurement Aspect	Proposed Methods			
	MOG	MOG2	GMG	KNN
TPR	*	***	**	****
FPR	****	**	*	***
Recall	*	***	**	****
Precission	****	**	*	***
MCC	***	**	*	****
PWC	****	**	*	***
Accuracy	****	**	*	***
F-measure	***	**	*	****
Total Score	**24**	**18**	**10**	**28**

result in Table III by using 8 measurement aspects also the number of the highest average person and violation detection, we obtain the rank of the best method by using star points are KNN, MOG, MOG2, and GMG.

REFERENCES

[1] WHO, "WHO Coronavirus Disease (COVID-19) Dashboard", 2020. [Online]. Available: https://covid19.who.int/. [Accessed: 19-Jul-2020].

[2] N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke, "Strategies for mitigating an influenza pandemic", Nature, vol. 442, no. 7101, pp. 448-452, 2006.

[3] U. Irfan, "The math behind why we need social distancing, starting right now", Vox, 2020. [Online]. Available: https://www.vox.com/2020/3/15/21180342/coronavirus-covid-19-us-social-distancing. [Accessed: 20-Jul-2020].

[4] WHO, "Coronavirus disease (COVID-19) advice for the public", 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public. [Accessed: 17-Jul-2020].

[5] W. Naudé, "Artificial intelligence vs COVID-19: limitations, constraints and pitfalls", AI Soc., no. 28 April 2020, pp. 1-5, 2020.

[6] B. Blocken, F. Malizia, T. van Druenen, and T. Marchal, "Towards aerodynamically equivalent COVID19 1.5 m social distancing for walking and running", Preprint, 2020. [Online]. Available: http://www.urbanphysics.net/SocialDistancing 20_White_Paper.pdf.

[7] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques", Preprint, 2020. [Online]. Available: https://arxiv.org/pdf/2005.01385.pdf.

[8] A. Gogna et al., "Diagnostic Ultrasound Services During the Coronavirus Disease (COVID-19) Pandemic", Am. J. Roentgenol., pp. 1-6, 2020.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification", Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1701-1708, 2014.

[10] I. Benraya and N. Benblidia, "Comparison of Background Subtraction methods", Proc. 2018 Int. Conf. Appl. Smart Syst. ICASS 2018, no. November, pp. 1-5, 2019.

[11] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians", Proc. - 2nd IEEE Int. Work. Vis. Surveillance, VS 1999, pp. 74-81, 1999.

[12] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", Proceedings. 1999 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2, pp. 246-252, 1999.

[13] Y. Ivanov, C. Stauffer, A. Bobick, and W. E. L. Grimson, "Video surveillance of interactions", Proc. - 2nd IEEE Int. Work. Vis. Surveillance, VS 1999, no. October, pp. 82-89, 1999.

[14] T. Trnovszk, P. Sýkora, and R. Hudec, "Comparison of background subtraction methods on near Infra-Red spectrum video sequences", Procedia Eng., vol. 192, pp. 887-892, 2017.

[15] A. B. Godbehere, A. Matsukawa, and K. Goldberg, "Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation", American Control Conference (ACC), pp. 4305-4312, 2012.

[16] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction", Pattern Recognit. Lett., vol. 27, pp. 773-780, 2006.

[17] P. Kaewtrakulpong and R. Bowden, "Chapter 11 An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection", Proc. Eur. Work. Adv. Video Based Surveill. Syst., pp. 136-144, 2002.

[18] Z. Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction", Proc. 17th Int. Conf. Pattern Recognit. (ICPR 2004), vol. 2, pp. 28-31, 2004.

[19] A. Rosebrock, "OpenCV Social Distancing Detector". [Online]. Available: https://www.pyimagesearch.com/2020/06/01/opencv-social-distancing-detector/. [Accessed: 17-Jun-2020].

[20] A. Rosebrock, "Find distance from camera to object/marker using Python and OpenCV", 2015. [Online]. Available: https://www.pyimagesearch.com/2015/01/19/find-distance-camera-objectmarker-using-python-opencv/. [Accessed: 17-Jul-2020].

[21] D. M. W. Powers and Ailab, "Evaluation: From Precision , Recall and F-Factor to ROC , Informedness , Markedness and Correlation", J. Mach. Learn. Technol., vol. 2, pp. 37-63, 2007.

[22] A. Harvey and J. LaPlace, "MegaPixels: Origins, Ethics, and Privacy Implications of Publicly Available Face Recognition Image Datasets", Megapixels, 2019. [Online]. Available: https://megapixels.cc. [Accessed: 07-Aug-2020].

[23] S. H. Shaikh, K. Saeed, and N. Chaki, "Moving Object Detection Using Background Subtraction", SpringerBriefs in Computer Science, pp. 15-23, 2014.