In: 2022 IEEE International Conference on Multisensor Fusion and Integration (MFI 2022),
20-22 September 2022, Cranfield University, UK

Reducing Viral Transmission through AI-based Crowd Monitoring and Social Distancing Analysis

Benjamin Fraser, Brendan Copp, Gurpreet Singh, Orhan Keyvan, Tongfei Bian, Valentin Sonntag, Yang Xing, Weisi Guo, Antonios Tsourdos
School of Aerospace, Transport and Manufacturing, Cranfield University
Cranfield, United Kingdom
Email: {B.Fraser, Brendan.Copp.309, Gurpreet.Singh.388, Orhan.Keyvan.428, Tongfei.Bian.040, Valentin.Sonntag.053, Yang.X, Weisi.Guo, a.tsourdos}@cranfield.ac.uk

***Abstract*— This paper explores multi-person pose estimation for reducing the risk of airborne pathogens. The recent COVID-19 pandemic highlights these risks in a globally connected world. We developed several techniques which analyse CCTV inputs for crowd analysis. The framework utilised automated homography from pose feature positions to determine interpersonal distance. It also incorporates mask detection by using pose features for an image classification pipeline. A further model predicts the behaviour of each person by using their estimated pose features. We combine the models to assess transmission risk based on recent scientific literature. A custom dashboard displays a risk density heat-map in real time. This system could improve public space management and reduce transmission in future pandemics. This context agnostic system and has many applications for other crowd monitoring problems.**

***Keywords*— Social Risk Analysis, Pose Estimation, Distance Estimation, Mask Detection, Behaviour Classification**

I. Introduction

The importance of public space management during future pandemics like COVID-19 is paramount. Crowd monitoring and social distancing analysis can ensure adequate safety measures. This work proposes a computer vision-based system that provides multi-functional crowd monitoring. Risk metrics based on scientific literature are output using a heatmap. Social distance, behaviour, and mask usage are the main factors impacting transmission [1]. These metrics form the basis for the proposed risk model.

Social distancing is a common topic of study following the recent COVID-19 pandemic. Work in [2] applies OpenCV and Faster-RNN to transform an input scene into an aerial view, which is used to estimate distance and identify dangerous social distancing behaviour. However, a disadvantage with this technique is the requirement to manually configure camera parameters. Work by [3] used OpenPose to perform human recognition and obtained higher accuracy than conventional object detection approaches, since pose features allow more precision and context than raw bounding boxes. Nawaz et al. [4] also proposed a similar method and added population density analysis. [5] and [6] extend on this by using the assumption that the height of a human is approximately the same, which enables localised homographic matrices and the least squares method to perform automatic configuration and estimate inter-personal distance. Such techniques could prove valuable for an integrated crowd analysis platform, as proposed in this work.

Several recent studies have focussed on social distance analysis, behaviour analysis and mask detection, such as [7] and [8]. These focus on object detection based systems with using simple scenes of people with and without masks. A criticism is that the data used is too simple compared to real world applications of social distancing analysis, which would generally involve significantly more complex and crowded scenes. They also lack application of homography techniques to accurately estimate the positions of people in scenes, which arguably is crucial for effective distancing analysis.

This work overcomes these limitations by development of a pose-based system to analyse social distancing risk using three unique downstream models. The model results are used to estimate various social distancing risk factors. The final outputs provide valuable crowd monitoring capabilities and assessment of risk using a density heatmap. This is displayed alongside the original scene, which helps pin-point key regions of risk over time for different areas. Thus, the framework provides an intuitive display of social risks and crowd monitoring analytics that can improve public space management.

The core contributions of this work include:

1) Integration of crowd-pose features to perform social distance analysis and crowd-monitoring on real-world camera-surveillance scenes.
2) Crowd inter-personal distance estimation using automated homography techniques, without the need for manual camera calibration or referencing.
3) Classification of person behaviour status using extracted pose features.
4) Face-mask classification using an image classification pipeline based on pose-feature extracted regions.
5) Development of a combined social risk metric, which can be used to gauge crowd safety during pandemics or general purpose crowd analysis.

All code and modelling is available on GitHub at the following link: https://github.com/BenjaminFraser/Social-Distancing-Pose-Platform.

II. Proposed Method

The system applies multi-person pose detection on all the people in each scene. The extracted features include the predicted positions and associated confidences of 17 human skeleton key-points. The chosen pose-estimation model was AlphaPose [9], due to its high-performance on a variety of complex and crowded scenes. This outperformed OpenPose significantly in comparison [10].

AlphaPose is used to obtain pose features for all persons in a scene. These features include the x and y coordinates for each key point, along with associated confidences and human bounding box coordinates. These features are then fed into a range of downstream models, which produce various predictions relating to social-risk and crowd monitoring. The collective results are combined and analysed using a custom density-heatmap, which shows areas with higher transmission risk averaged. Each downstream modelling technique is summarised in the following sections.

Algorithm 1: Camera parameters and positions estimation algorithm

Input: $x_1^h, y_1^h, ..., x_n^h, y_n^h$: Head positions of the persons in the frame
$x_1^f, y_1^f, ..., x_n^f, y_n^f$: Feet positions of the persons in the frame
ε: Error on estimated focal length
α: Interval reduction factor ($0 < \alpha < 0.5$)
h: Average height of a person
$frame_{width}$: Width of the frame
$[FOV_{min}, FOV_{max}]$: Confidence interval of the camera FOV

Output: θ, H and f: Camera parameters
$X_1, Y_1, ..., X_n, Y_n$: Positions of the persons in the scene

Initialize interval I = $\left[\frac{frame_{width}}{2\,tan(\frac{FOV_{max}}{2})}, \frac{frame_{width}}{2\,tan(\frac{FOV_{min}}{2})}\right]$

while $\frac{\max I - \min I}{2} > \varepsilon$ **do**
  $f_1 \leftarrow \min I \; + \alpha\,(\max I - \min I)$
  $f_2 \leftarrow \max I - \alpha\,(\max I - \min I)$
  for $i \in [1, 2]$ **do**
    for $p \in [1, ..., n]$ **do**
      if $x_p^h \neq 0$ *and* $x_p^f \neq 0$ **then**

$$X_p \leftarrow \frac{x_p^h\, h}{\sqrt{\left(f_i\left(1-\frac{x_p^h}{x_p^f}\right)\right)^2+\left(y_p^h-y_p^f\frac{x_p^h}{x_p^f}\right)^2}}$$

$$Y_p \leftarrow \frac{X_p\, y_p^f}{x_p^f\sqrt{1-\left(\frac{X_p}{h}\left(\frac{y_p^h}{x_p^h}-\frac{y_p^f}{x_p^f}\right)\right)^2}}$$

$$\theta_p \leftarrow \arccos\left(\frac{X_p}{h}\left(\frac{y_p^h}{x_p^h}-\frac{y_p^f}{x_p^f}\right)\right)$$

$$H_p \leftarrow \frac{X_p}{x_p^f}\left(f_i\sqrt{1-\left(\frac{X_p}{h}\left(\frac{y_p^h}{x_p^h}-\frac{y_p^f}{x_p^f}\right)\right)^2}-y_p^f\cos(\theta_p)\right)$$

      end
    end
    $\theta_{f_i} \leftarrow \underset{p\in[1,...,n]}{median}\,\theta_p$
    $H_{f_i} \leftarrow \underset{p\in[1,...,n]}{median}\,H_p$

$$\Delta_{f_i} \leftarrow \underset{p\in[1,...,n]}{median}\left(\left(f_i X_p\frac{\sin\theta_{f_i}}{H_{f_i}+Y_p\cos\theta_{f_i}\sin\theta_{f_i}}-x_p^f\right)^2+\left(f_i Y_p\frac{\sin^2\theta_{f_i}}{H_{f_i}+Y_p\cos\theta_{f_i}\sin\theta_{f_i}}-y_p^f\right)^2\right)$$

  end
  if $\Delta_{f_1} > \Delta_{f_2}$ **then**
    $\min I \leftarrow f_1$
  else
    $\max I \leftarrow f_2$
  end
end
$f \leftarrow \frac{\max I - \min I}{2}$
for $p \in [1, ..., n]$ **do**
  if $x_p^h \neq 0$ *and* $x_p^f \neq 0$ **then**

$$X_p \leftarrow \frac{x_p^h\, h}{\sqrt{\left(f\left(1-\frac{x_p^h}{x_p^f}\right)\right)^2+\left(y_p^h-y_p^f\frac{x_p^h}{x_p^f}\right)^2}}$$

$$\theta_p \leftarrow \arccos\left(\frac{X_p}{h}\left(\frac{y_p^h}{x_p^h}-\frac{y_p^f}{x_p^f}\right)\right)$$

$$H_p \leftarrow \frac{X_p}{x_p^f}\left(f\sqrt{1-\left(\frac{X_p}{h}\left(\frac{y_p^h}{x_p^h}-\frac{y_p^f}{x_p^f}\right)\right)^2}-y_p^f\cos(\theta_p)\right)$$

  end
end
$\theta \leftarrow \underset{p\in[1,...,n]}{median}\,\theta_p$
$H \leftarrow \underset{p\in[1,...,n]}{median}\,H_p$
for $p \in [1, ..., n]$ **do**

$$X_p \leftarrow \frac{H\, x_p^f}{f\,\sin\theta - y_p^f\,\cos\theta}$$

$$Y_p \leftarrow \frac{H\, y_p^f}{f\,\sin^2\theta - y_p^f\,\cos\theta\sin\theta}$$

end

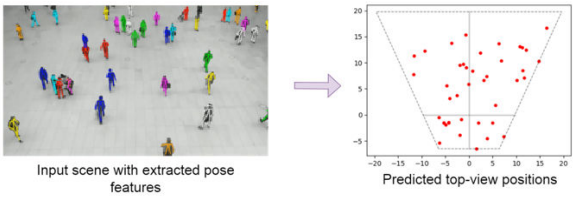

Fig. 1. Homography of a scene.

A. Distance Estimation within 2D Scenes

Whether people are violating social distancing rules or not can be predicted by estimating the distance between them. We propose a novel approach where the extracted pose features are used to generate and tune the camera parameters needed for homography. Homography allows transitioning from the camera view to the top-down view of each scene (Fig. 1), which facilitates effective distance estimation and group clustering.

The advantage of this method is that it does not require manual calibration and measurement of camera parameters. Additionally, it provides robustness for the model in overcoming drift errors and allows real-time adaptation to dynamic contexts such as a moving camera on a drone. This method is more accurate than using bounding boxes to determine the distances between people because they are more sensitive to pose-feature estimation errors. This method finally uses inverted-homography to project the risk heatmap on the camera display to better identify high risk areas.

The distance estimation is based on the homographic relations between the image position (x, y) and the 3D position (X, Y, Z). The central point of the 3D position is defined as the origin, this is then projected on the centre of the 2D image using a homography function. The function parameters are the tilt angle θ, the height H and the focal length f of the camera.

It is possible to estimate the camera parameters and the positions of the detected persons in the 3D scene given their head and feet positions in the frame. The process applied in this work is detailed in Algorithm 1. Estimates are first computed for the camera tilt angle and height to determine the focal length using the head and feet positions for each person in the frame. The median value for each parameter is then computed across all estimates, which helps remove noisy or spurious values. Once the focal length has been determined, the two estimations of camera tilt angle and height are computed, and their median values are taken as the final values. These camera parameters are then used to calculate the positions of the detected persons in the 3D scene.

B. Pose Behaviour Classification

According to simulation data on diffusion of coughing particles from Muthusamy [1], particles expelled from the mouth can travel a greater distance while in a sitting position compared to a standing position. This means we need to differentiate people in different poses so that the model can apply different risk weights. Human pose features offer a valuable means of understanding the behaviour and intentions of persons in a scene. This is exploited by feeding the pose features from AlphaPose into a neural network. This then predicts the most likely current behaviour status out of one of five possible defined classes (Fig. 2).

The pose classification model was developed using a custom dataset created from the Oxford Streets dataset. The pose features were extracted from each frame using AlphaPose and each person labelled according to their current behaviour. This included standing, sitting, walking, lying and other pose categories. Sparse categorical cross-entropy was used as the model training loss. ReLU activation was applied throughout all Deep Neural Network (DNN) layers, with exception to the final layer, which was softmax. The model inputs included the key-point coordinates and confidences

Fig. 2. Examples of standing, walking, sitting and lying status classes.

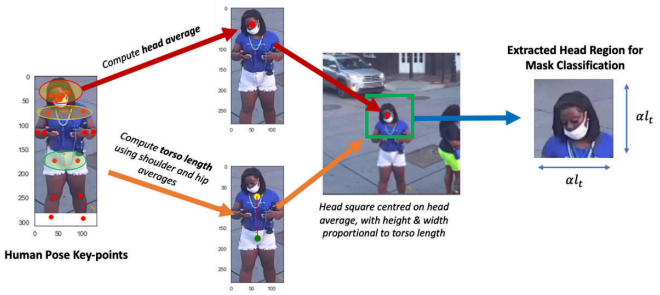

Fig. 3. Overview of the head-region extraction process using pose features.

obtained for each person from AlphaPose. During pre-processing the coordinates were centred and normalised between 0 and 1. This was achieved using the human bounding box coordinates as reference points.

C. Mask Detection

The purpose of mask detection is to predict the presence of a mask on all people in a scene. Existing works have tackled this problem through applying object detection over an entire scene to detect masked faces. This work proposes a novel approach where the extracted pose features are used to obtain suitable head regions. These are then fed into an image classifier to predict the likelihood of wearing a mask. The classifier was developed using transfer-learning and fine-tuning with a BiT-M R50x1 architecture pre-trained on the ImageNet-21K dataset. The Big Transfer fine-tuning and optimising process proposed by [11] was adopted.

This process consists of computing the average head, shoulder and hip coordinates from the pose features. The torso length is approximated using the Euclidean distance between the average shoulder and hip co-ordinates. This is multiplied by a scaling factor to give the length of extracted square region, which is centred on the average head co-ordinates. This small region can then be efficiently processed by an image classifier to predict mask likelihood for each person (Fig 3).

This process is advantageous over object detection since classification is performed efficiently on small regions. This provides faster inference on large scenes. It also integrates more naturally into the proposed framework, since the pose features are directly leveraged for obtaining the head regions. Finally, the classification performance metrics are simple to evaluate and optimise compared to object detection metrics.

D. Risk Density Heatmap

The downstream model outputs are combined to estimate the risk profile for each scene. A density heat-map was generated to represent this using Kernel Density Estimation (KDE) with risk-based weights (Equation 1). A gaussian based kernel was chosen. Each weight was computed based on the distance, behaviour and mask status of the associated person in the scene. This represents

Fig. 4. Example of the risk density heatmap on the top-down perspective of a scene.

the temporal social-distancing risk for each scene, with a 2D risk matrix developed for each timestep (Fig. 4).

$$p(x|D) = \frac{1}{N}\sum_{n=1}^{N} \alpha_{risk}\kappa_h(x - x_n) \tag{1}$$

where x_n is the input feature vector for the n^{th} sample, N is the total number of samples in the dataset D, κ_h is the chosen density kernel function, and α_{risk} is the risk−proportional weight.

The data samples, D, are the 2D coordinates of each person in the top-down representation of each scene. The sample weights were computed for each sample, based on whether that person was violating distance, wearing a mask, and their current pose status (Equation 2).

$$\begin{aligned}\alpha_{risk}(x) = I_{vio}(x)\big[\tau_{dist} + I_{no\_mask}(x)\tau_{mask} \\ + I_{pose\_risk}(x)\tau_{pose}\big]\end{aligned} \tag{2}$$

where I_{vio} is an indicator function (0 or 1) for a person violating distance, I_{no_mask} is an indicator function for a person with no mask, I_{pose_risk} is an indicator function for a higher−risk pose, τ_{dist} is a chosen distancing factor, τ_{mask} is a mask−usage factor and τ_{pose} is a pose status factor.

For a given scene, all persons that are well-distanced from others by a defined distance are assigned a weight of zero. The violation distance used in this work was 1.6 m, based on research from [1]. At any instance in time, the risk is zero for that location and no contribution is made towards the risk density heatmap if no distance-violation is detected. Conversely, those persons violating proximity will be assigned a risk-based weight based on whether that person is wearing a mask (I_{no_mask}) and whether they are assuming a risky behaviour (I_{pose_risk}). The influence of each social distancing factor on the final weight is determined by the specific factors chosen for each: τ_{dist}, τ_{mask}, and τ_{pose}, as determined from [1].

These concepts were extended to video sequences by using a 3D tensor, which contained a 2D risk-matrix for each timestep. This generates a risk profile heatmap based on an average risk over a designated time-period rather than individual frames. Furthermore, this was also projected to the original scene, rather than from a top-down perspective, using reverse-homography techniques.

III. Experiments and Results

The overall system modelling process consists of one large loop, which takes an input video feed and processes the image frames at 5 frames-per-second. Pose features are extracted from each frame and applied to each downstream model. The results are then integrated into a set of outputs suitable for analysis and visualisation on a dashboard application.

TABLE I

DATASETS USED FOR SYSTEM DEVELOPMENT

Dataset	Summary	Used for
Moxa3k [12]	3,000 facemask classification images.	Custom mask classifier. dataset.
Real-World Webcam Mask Dataset [13]	2,311 facemask classification images.	Custom mask classifier evaluating pose estimation.
Face Mask Detection Video Dataset [4]	4,357 facemask classification video frames.	Evaluating mask classifier, final test dataset.
CityUHK-X-BEV [14]	3,191 images from CCTV cameras with their parameters.	Evaluating pose estimation, developing distance estimation, final test dataset.
Oxford Street Dataset [15]	5 minutes video of public street with camera parameters.	Developing distance estimation, density heatmap, test dataset.
Human-centric Video Analysis in Complex Events (HiEve) [16]	32 airport video sequences with persons' activity.	Final test video scenes.

A. Datasets

In the process of system modelling, some data sets are used for the development of downstream models and testing at different stages. These are: Moxa3k [12], Real World Webcam Mask Dataset [13], Face Mask Detection Video Dataset [4], CityUHK-V-BEV [14], Oxford Street Dataset [15], and HiEve [16]. These datasets contain large amounts of images and videos of crowds and public spaces (Table I).

B. Distance estimation

An example of the distance estimation accuracy computed across a range of different values of θ and H is given in Fig. 5. The distance estimation accuracy is dependent on AlphaPose's accuracy. In general, AlphaPose becomes less accurate with higher distances from the camera and larger crowds. This is due to the lower resolution images of each person and occasional overlapping body parts. The distance estimation has a global 10% confidence interval. For typical tilt angle of CCTV camera, the confidence interval drops below 5%, which is less than a 10 cm range for a 2 m social distance limit for example.

Cluster analysis was performed using the DBSCAN algorithm because it's effective for identifying the movement of social groups in public places. This density-based clustering model can be adapted to current social distancing policy by adjusting the threshold distance of the algorithm.

C. Pose behaviour classification

The behaviour classifier is a fully connected DNN that outputs the behaviour class of each detected person. Its input contains coordinates and confidence scores of 17 key-points and information on bounding box size. Information on bounding box size includes aspect ratio and relative width and height to the frame. The formula used for aspect ratio is:

$$Aspect\ Ratio = \frac{Width_{bbox}}{Height_{bbox}} \tag{3}$$

The formulas for relative width and height to the frame are:

$$Relative\ Width = \frac{Width_{bbox}}{Width_{frame}} \tag{4}$$

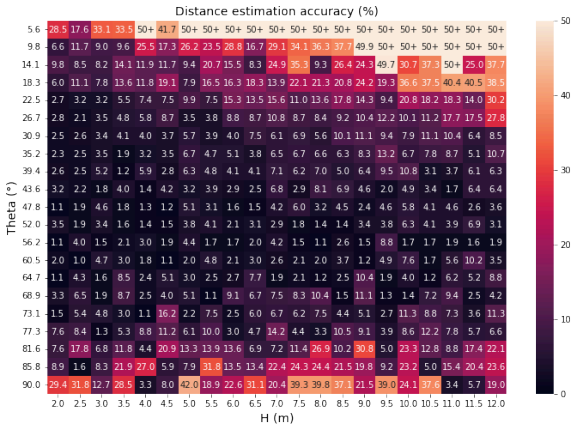

Fig. 5. Distance estimation accuracy for different camera parameters.

TABLE II

PERFORMANCE OF DIFFERENT INPUTS

Input	Accuracy	Loss
Only Coordinates	78.52%	0.531
Add Confidence Scores	87.13%	0.331
Add Bounding Box Information	93.04%	0.213

$$Relative\ Height = \frac{Height_{bbox}}{Height_{frame}} \tag{5}$$

To accelerate convergence and improve performance, the x and y coordinates of the key-points were centred and normalized as follows:

$$X = \frac{X_{keypoint} - X_{bbox}}{Width_{bbox}} \tag{6}$$

$$Y = \frac{Y_{keypoint} - Y_{bbox}}{Height_{bbox}} \tag{7}$$

The final behaviour classifier used all 54 dimensions of data from the extracted pose features because it improved the model to obtain the highest classification performance. The final model used seven layers with 64 units per layer.

After 120 epochs, the accuracy on the test set reached 93.04%. On inspection, the model appears more prone to error when distinguishing between standing and walking (Fig. 6). Since the DNN is simple, it does not require costly computation, and thus the prediction speed was measured as more than 32 people per millisecond on CPU.

D. Mask Detection Modelling

For assessing mask detection performance, a combination of accuracy, precision, recall, F1-score, and the Receiver Operator Characteristic (ROC) were assessed. False positives (predicting a mask) were deemed more important than false negatives (predicting no mask), and therefore precision was optimised through adjustment of the prediction threshold to 70% confidence during evaluation.

To obtain the best compromise of precision and recall, the mask prediction probability threshold was optimised using the ROC curve, which gave the confusion matrix results in Fig. 7. The model prioritised reducing the number of false positives. This came at the expense of allowing higher numbers of false negatives where masked people were classified as non-masked. This adjustment facilitates a cautious risk model, which accounts for the relatively high consequences of transmission.

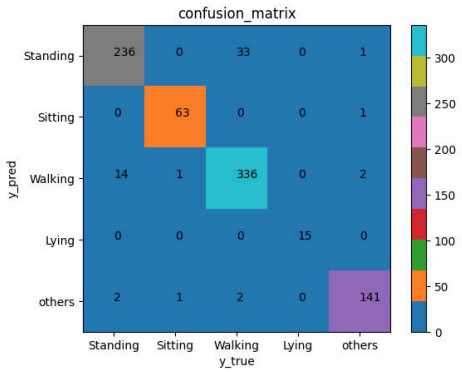

Fig. 6. Confusion matrix of pose behaviour classification model.

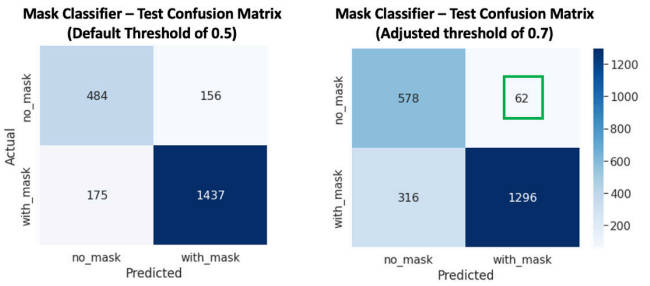

Fig. 7. BiT mask classifier final confusion matrix on the held-out test set.

IV. Final System Results and Discussions

A. Combined System Results and Risk Density Heatmap

The final system was tested on four distinct areas containing public crowds in different social settings, which were selected from a combination of the datasets presented in Table I. The modelling pipeline was applied on video sequences from each area at a frequency of 5 frames per second. Examples of this are shown for an airport in Fig. 8. Overall, the system effectively demonstrated the potential of a multi-person pose feature framework. The models worked well across a variety of different scenes and camera viewpoints, highlighting the benefits that the automated homography approach provides.

The heatmap generated for each area using all the downstream modelling results can provide valuable visualisations for public area management. This is particularly valuable when scenes are analysed for social-risk and crowd-features over designated time-periods (Fig. 8). This strategy allows analytics to be generated during key times throughout the day, and even specially chosen dates throughout the year. This could be helpful to identify key trends that can be used to implement additional safety barriers or make infrastructure adjustments to improve daily flows of people. This also facilitates the gathering of useful statistics regarding social groups and movement of people, including person counts, mask usage, group sizes, and pose behaviours.

B. Further System Discussions and Improvements

A huge advantage of the system is the ability for the homography and distance modelling to calibrate automatically to any scene, regardless of camera position and angle. Thus, the system can be integrated into new applications with minimal effort, providing there is available video surveillance. It even supports accurate use with moving cameras operated from a dynamic platform such as a drone or aircraft. Thus, the concepts could be extended to any outdoor public space for crowd monitoring regardless of pre-existing infrastructure.

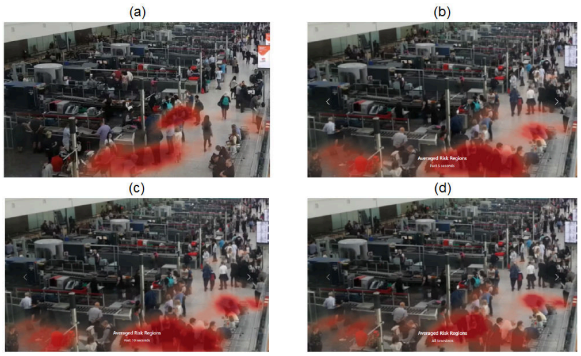

Fig. 8. Dashboard displaying the scene at (a) 1 second, averaged risk-regions at (b) 5 seconds, (c) 10 seconds and (d) all timesteps.

Beyond social distancing the system has enormous potential. Although this paper focused on social distancing, the same concepts could be applied to other domains. An example might be the adaptation to health and safety, where the workforce areas could be monitored for correct working practices. This might include personal protective equipment (safety helmets and high-visibility vests) and ensuring they are not lone-working for safety-critical tasks. Another example is self-driving or driver-assisted vehicles, where the crowd-monitoring framework could help analyse the risk profile of nearby pedestrians. A further application could be monitoring social compliance and safety within prisons for security personnel and inmates.

An observed limitation of the system was the accurate counting and monitoring of people in complex and crowded scenes with large number of people (typically greater than 30). In these complex environments it was common for people to be missed by the object detector used during pose estimation. This highlights the heavy reliance the system has on the performance of the initial human detection phase. A YOLOv3 model was used within this framework, but improvements could be made using a more modern and improved architecture more suited for small objects.

A further concern was the difficulty in assigning risk factors for distance violation, mask-usage and pose behaviour for the risk-density heatmap. Despite being informed from research, the values chosen are subjectivity and depend on many contextual factors. This includes the type of epidemic, the virus, and its variants (contagiousness, survival rate, etc). This subjectivity presents a challenge and would require refinement through further research and work with specialist organisations.

A valuable improvement would be the use of person tracking. This could provide advanced behaviour modelling for individuals and groups. For example, clusters of small groups of people could be analysed over time to assess whether they are travelling together. This would help provide more realistic and usable results for the system risk heatmap. Sequence tracking would also improve the reliability and performance of mask-usage and behavioural classification. In this case, predictions could be averaged for individuals over time, helping to overcome noise, anomalies and obscurities.

V. Conclusion

The aim of this paper was to present an integrated machine-learning based framework for crowd monitoring and social distance analysis in public environments. The final system combines many methods and technologies into a unified framework and displays social risk distribution using a heatmap. The results were promising and demonstrate the flexibility and adaptability of the framework. The modelling parameters are informed by academic research however a challenge is that they remain subjective. This is due to the uncertainty and disagreement that may arise when assigning

risk weights for distance violation, mask-usage and behaviour status. Moreover, the system was focussed predominantly on the COVID-19 virus. Nevertheless, the system has great potential beyond analysing social risks. The concepts naturally extend to other applications, including autonomous vehicles, health and safety monitoring, social compliance in prisons, shopping centres, public transport and more. The system is also highly flexible, since the automated homography allows convenient adaptability to new environments or even moving cameras.

ACKNOWLEDGEMENT

This research was supported by School of Aerospace, Transport and Manufacturing and Centre of Autonomous and Cyber Physical Systems of Cranfield University.

REFERENCES

[1] J. Muthusamy, S. Haq, S. Akhtar, M. A. Alzoubi, T. Shamim, and J. Alvarado, 'Implication of coughing dynamics on safe social distancing in an indoor environment—A numerical perspective', Building and Environment, vol. 206. Elsevier BV, p. 108280, Dec. 2021. doi: 10.1016/j.buildenv.2021.108280.

[2] O. Karaman, A. Alhudhaif, and K. Polat, 'Development of smart camera systems based on artificial intelligence network for social distance detection to fight against COVID-19', Applied Soft Computing, vol. 110. Elsevier BV, p. 107610, Oct. 2021. doi: 10.1016/j.asoc.2021.107610.

[3] M. Al-Sa'd, S. Kiranyaz, I. Ahmad, C. Sundell, M. Vakkuri, and M. Gabbouj, 'A Social Distance Estimation and Crowd Monitoring System for Surveillance Cameras', Sensors, vol. 22, no. 2. MDPI AG, p. 418, Jan. 06, 2022. doi: 10.3390/s22020418.

[4] F. Nawaz, W. Khan, S. Yasen, and A. Hussain, 'Face Mask Detection Video Dataset'. Mendeley, Nov. 18, 2020. doi: 10.17632/V3KRY8GB59.1.

[5] S. Das et al., 'Computer Vision-based Social Distancing Surveillance with Automated Camera Calibration for Large-scale Deployment', 2021 IEEE 18th India Council International Conference (INDICON). IEEE, Dec. 19, 2021. doi: 10.1109/indicon52576.2021.9691485.

[6] M. Aghaei, M. Bustreo, Y. Wang, G. Bailo, P. Morerio, and A. Del Bue, 'Single Image Human Proxemics Estimation for Visual Social Distancing'. arXiv, 2020. doi: 10.48550/ARXIV.2011.02018.

[7] S. Srinivasan, R. Rujula Singh, R. R. Biradar, and S. Revathi, 'COVID-19 Monitoring System using Social Distancing and Face Mask Detection on Surveillance video datasets', 2021 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, Mar. 05, 2021. doi: 10.1109/esci50559.2021.9396783.

[8] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, 'SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2', Sustainable Cities and Society, vol. 66. Elsevier BV, p. 102692, Mar. 2021. doi: 10.1016/j.scs.2020.102692.

[9] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, 'RMPE: Regional Multi-person Pose Estimation', 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Oct. 2017. doi: 10.1109/iccv.2017.256.

[10] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, 'OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1. Institute of Electrical and Electronics Engineers (IEEE), pp. 172–186, Jan. 01, 2021. doi: 10.1109/tpami.2019.2929257.

[11] A. Kolesnikov et al., "Big Transfer (BiT): General Visual Representation Learning," in ECCV 2020: 16th European Conference, Dec. 2020, pp. 491–507. doi: https://doi.org/10.1007/978-3-030-58558-7-29.

[12] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas, and T. Das, 'MOXA: A Deep Learning Based Unmanned Approach For Real-Time Monitoring of People Wearing Medical Masks', Transactions of the Indian National Academy of Engineering, vol. 5, no. 3. Springer Science and Business Media LLC, pp. 509–518, Jul. 25, 2020. doi: 10.1007/s41403-020-00157-z.

[13] E. Adhikarla and B. D. Davison, 'Face Mask Detection on Real-World Webcam Images', Proceedings of the Conference on Information Technology for Social Good. ACM, Sep. 09, 2021. doi: 10.1145/3462203.3475903.

[14] Z. Dai, Y. Jiang, Y. Li, B. Liu, A. B. Chan, and N. Vasconcelos, 'BEV-Net: Assessing Social Distancing Compliance by Joint People Localization and Geometric Reasoning', 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Oct. 2021. doi: 10.1109/iccv48922.2021.00535.

[15] B. Benfold and I. Reid, 'Stable multi-target tracking in real-time surveillance video', CVPR 2011. IEEE, Jun. 2011. doi: 10.1109/cvpr.2011.5995667.

[16] W. Lin et al., 'Human in Events: A Large-Scale Benchmark for Human-centric Video Analysis in Complex Events'. arXiv, 2020. doi: 10.48550/ARXIV.2005.04490.

Reducing viral transmission through AI-based crowd monitoring and social distancing analysis

Fraser, Benjamin

2022-10-13

Fraser B, Copp B, Singh G, et al., (2022) Reducing viral transmission through AI-based crowd monitoring and social distancing analysis. In: 2022 IEEE International Conference on Multisensor Fusion and Integration (MFI 2022), 20-22 September 2022, Cranfield University, UK
https://doi.org/10.1109/MFI55806.2022.9913843