

# Crowd Behavior Analysis based on Convolutional Neural Network: Social Distancing Control COVID-19

Fatma Bouhlel<sup>1</sup> <sup>a</sup>, Hazar Mliki<sup>2</sup> <sup>b</sup> and Mohamed Hammami<sup>3</sup> <sup>c</sup>

<sup>1</sup>MIRACL-FSEG, University of Sfax, Faculty of Economics and Management of Sfax, Road Airport Km 4, 3018 Sfax, Tunisia

<sup>2</sup>MIRACL-ENET'COM, University of Sfax, National School of Electronics and Telecommunications of Sfax, Road Tunis City El Ons, 3018 Sfax, Tunisia

<sup>3</sup>MIRACL-FS, University of Sfax, Faculty of Sciences of Sfax, Road Sokra Km 3, 3018 Sfax, Tunisia

**Keywords:** COVID-19, Crowd Behavior, Social Distancing, Crowd Density Estimation, Human Detection, Convolutional Neural Network, UAV.

**Abstract:** The outbreak of the COVID-19 and the lack of pharmaceutical intervention increase the spread of COVID-19. Since no vaccine or treatment are yet available, social distancing represents a good strategy to control the propagation of this pandemic and learn to live with it. In this context, we introduce a new approach for crowd behavior analysis from UAV-captured video sequences in order to monitor social distancing. The proposed approach involves two methods: a macroscopic method and a microscopic method. The macroscopic method aims to estimate the crowd density by classifying the aerial frame patches into four categories: Dense, Sparse, Medium and None. However, the microscopic method allows to detect and track humans and then compute the distance between them. The quantitative and qualitative results validate the performance of our methods compared to the state-of-the-art references.


## 1 INTRODUCTION


In December 2019, Wuhan city which is the capital of Hubei became the propagation source of a pneumonia outbreak, known as coronavirus disease 2019 (COVID-19) (Lewnard and Lo, 2020; Zhang et al., 2020; Zhou et al., 2020). According to the number of confirmed cases and the number of caused death, the COVID-19 has declared by the World Health Organization (WHO) as though a pandemic virus (WHO, 2020). Indeed, this pandemic was propagated promptly (Punn et al., 2020) to more than 216 countries. In fact, about 26.121.999 confirmed cases are noted along with 864.618 deaths in the world on September 4, 2020. The lack of active therapeutic agents and the dearth of immunity raise the spread of COVID-19 (Punn et al., 2020; Singh and Adhikari, 2020; Wilder-Smith and Freedman, 2020). As no vaccines are currently available, social distancing constitutes an effective strategy allowing the standstill of this pandemic spread (Punn et al., 2020; Singh and Adhikari, 2020; Park et al., 2020; Cristani et al., 2020). In fact, social distancing in pandemic time has


internal as well as external benefits helping humans to learn living with this virus (Zhou et al., 2020). The internal benefits consist of being less probably to catch the virus in the case of socially distant. As for external benefit, it helps reducing the propagation of the virus to other people, precisely other strangers. In the context of COVID-19, it is recommended to maintain a distance of two meters from other humans (Repici et al., 2020).

For this purpose, we introduce a new approach, that tends to help the social inspection force by alerting them in the case of non-compliance with social distancing measures. The proposed approach aims to analyze the crowd behavior from UAV-captured video sequences in order to monitor social distancing. Our approach consists of two methods: a macroscopic method and a microscopic method. These methods are based on the use of convolutional neural networks (CNN) and transfer learning. The macroscopic method estimates the crowd density by classifying the aerial frames patches within four categories: Dense, Sparse, Medium and None. The microscopic method detects and tracks humans, to compute the distance between them. The main contributions of this paper are outlined as follows:

- The proposed strategy of dividing the input frame

<sup>a</sup>  <https://orcid.org/0000-0001-9979-9729>

<sup>b</sup>  <https://orcid.org/0000-0002-0285-0944>

<sup>c</sup>  <https://orcid.org/0000-0003-3580-0473>

into N-sub-frames and estimating their density in parallel process allows to reduce the time computation. Such contribution makes our approach suitable for real-time applications.

- The complementary combination of macroscopic and microscopic methods: The filtering level process in the macroscopic method allows selecting the suspicious crowd regions and focuses on the most critical situation. In fact, the detection of dense or medium crowd regions will trigger an alert to the inspection forces (police). However, the detection of a sparse crowd regions, will activate the microscopic process to check the maintain of the social distancing within these regions. This filtering process provides a particular focus on the most suspicious crowd regions and reduces the complexity cost and hence the time computation.
- Our approach stands out from the literature (Punn et al., 2020) in the use of UAV sensor which diverts the fixed coverage areas problem and assure open spaces surveillance (Hazar et al., 2019; Mliki et al., 2020). In fact, the UAV provides more flexibility enabling it to overcome the occlusion problem and monitor open space areas.

The remaining sections of the paper are organized as follows. Section 2 identified the related works on macroscopic and microscopic crowd behavior analysis. In section 3, the outlines of the proposed approach are reported. The experimental results are discussed in section 4. The last section reviews the proposed approach and offered some future research perspectives.

## 2 RELATED WORKS

The crowd behavior analysis is based on two types of complementary analysis methods: macroscopic method and microscopic method. In the following sections, the related works of these two methods in aerial views were explored.

### 2.1 Macroscopic Methods

The macroscopic methods handle visually indistinguishable crowded (VIC) aerial frames and analyze the crowd as a global entity. In fact, these methods neglect the local information and focus on the treatment of global information such as crowd density, crowd counting and crowd flow. In this state-of-the-art study, we are interested in crowd density estimation methods in aerial views. In this context, Meynberg and Kuschk (Meynberg and Kuschk, 2013) pro-

posed to encode crowd patches through the Gabor filter bank. Next, the resulted Gabor features are fed to the support vector machine (SVM) classifier. Finally, they classify crowd patches into two categories: dense and none. For the purpose of improving their previous work, the selfsame authors (Meynberg et al., 2016) proposed to make use of an LBP descriptor instead of the Gabor descriptor. They managed to estimate the crowd density in a fixed ground sampling distance (GSD) aerial images which ranges from 10 to 13 cm. Within the same axis, Mliki et al. (Hazar et al., 2019), extracted the points of interest from each crowd patch through the scale-invariant feature transform descriptor (SIFT). Afterward, they extracted from each interest centered region the texture features using the multiblock local binary pattern (MB-LBP) descriptor. Thereafter, they generated a global MB-LBP feature vector through the concatenated of each centered region MB-LBP vector features. Finally, the obtained global MB-LBP features are fed to the SVM classifier.

### 2.2 Microscopic Methods

The microscopic methods analyze crowd behavior by detecting each person in the crowd and analyzing its behavior. These methods describe efficiently the unusual events in an aerial view. Nonetheless, they can only handle the sparse crowds, as they are not able to separate humans in dense or medium crowds. Therefore, we analyzed person detection methods in aerial videos. AIDahoul et al. (AIDahoul et al., 2018) proposed a real-time human detection method from the UAV-captured video sequences. As a pretreatment, they used an optical flow technique to detect potential motion regions. Next, they classified these potential motion regions using a pretrained CNN AlexNet (Krizhevsky et al., 2012). Nonetheless, this method doesn't deal with the close objects constraint. To handle this constraint, Mliki et al. (Mliki et al., 2020) proposed to integrate a module of regions of interest generation and selection which helps adapting the classical CNN, devoted to the classification problem, to detect humans.

### 2.3 Discussion

In the literature, two types of complementary crowd behavior analysis methods are identified: macroscopic method and microscopic method. Although the good performance achieved by handcrafted methods (Hazar et al., 2019; Meynberg and Kuschk, 2013; Meynberg et al., 2016), they overlook the semantic information resulting from extremely abstract deep

features. Furthermore, they depend on the choice of descriptor (Mliki et al., 2020). Referring to this state-of-the-art study, we propose a new macroscopic method for crowd density estimation based on pre-trained CNN. Regarding the microscopic methods, we adopted the Mliki et al. (Mliki et al., 2020) method since it deals with the close objects constraint and adapts the classical CNN, devoted to the classification problem, to detect humans. Nonetheless, we proposed to add the human tracking step, in order to generate continuous humans location and then computing distance between them.

### 3 PROPOSED APPROACH

In order to monitor COVID-19 social distancing, we proposed a new approach for crowd behavior analysis from UAV-captured video sequences. The proposed approach consists of two methods as shown in Fig. 1: a macroscopic method and a microscopic method. The macroscopic method aims to estimate the crowd density by classifying the aerial frame patches into four categories: Dense, Sparse, Medium and None. The microscopic method allows to detect and track humans in order to compute the distance between them. In the following sections, each of these methods was detailed.

#### 3.1 Macroscopic Method

Since the proposed approach is designed for real time applications, it's important to improve the computation time and the algorithm complexity. Therefore, we parallelize the crowd density estimation process on multiple CPU to reduce time computation. In fact, each aerial frame is divided into  $N$  equal sub-frames depending on the number of the CPUs Cores. The crowd density estimation proposed method is based on a pretrained CNN. Various pretrained CNN have been proposed such as: AlexNet (Krizhevsky et al., 2012), ZFNet (Zeiler and Fergus, 2014), VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), ResNet (He et al., 2016). Basing on a comparative study performed by Kaiming et al. (He et al., 2016) which assesses the pre-trained CNN effectiveness in terms of the number of layers and classification error rate, we adopted the pre-trained CNN 'AlexNet'. AlexNet

Thereby, we substitute the classification layer by a novel softmax layer to classify the crowd patches into four categories: Dense, Sparse, Medium and None.

Thereafter, we fine-tuned the obtained model to our context of study in order to generate an adequate

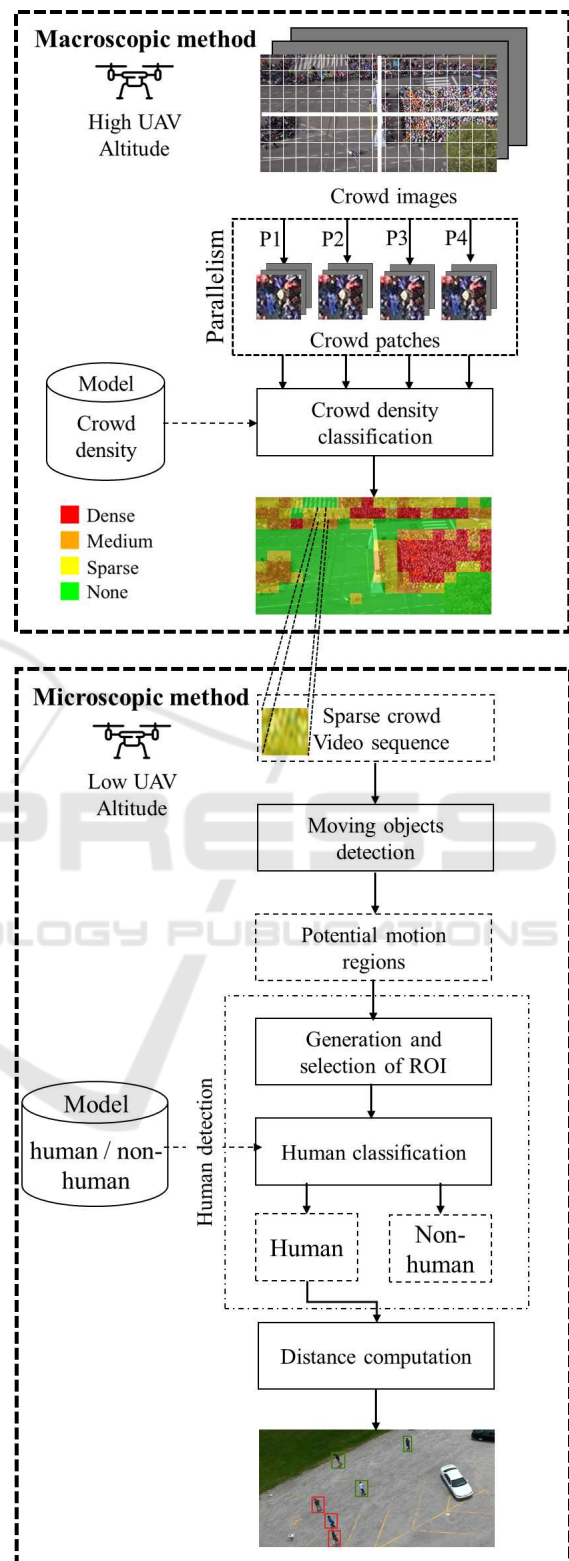


Figure 1: Proposed approach for crowd behavior analysis to monitor social distancing.

crowd density model able to filter crowd regions and focuses on the most critical ones. Hence, three scenarios can be addressed as illustrated in Fig. 2:

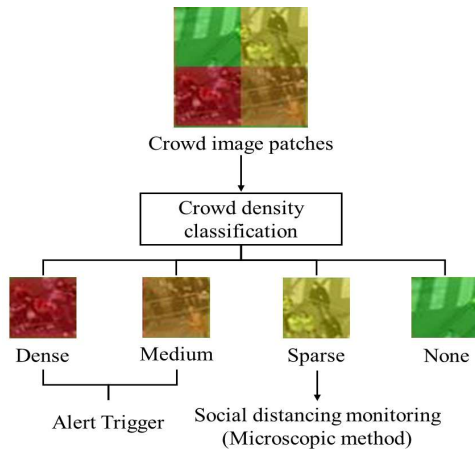


Figure 2: Filtering level process.

- Crowd patch is classified as ‘none’: this patch is empty.
- Crowd patch is classified as ‘dense’ or ‘medium’: this patch presents non-conformity with social distancing problem. Therefore, an alert is triggered to the social inspection forces with GPS location information.
- Crowd patch is classified as ‘sparse’: this patch is suspicious and needs more focus to check the conformity of social distancing using a microscopic method.

## 3.2 Microscopic Method

This method aims to focus on the ‘sparse’ crowd region where people are distinguishable. The proposed method consists of 3 steps: (1) Moving objects detection, (2) Human detection and (3) Social distance computation.

### 3.2.1 Moving Objects Detection

This step allows eliminating the acquisition sensor ego-motion and detecting the potential motion regions. In fact, the estimation of each pixel motion provides a set of motion vectors resulting from both of the potential motion regions and the UAV displacement in the scene. To distinguish these motions, the speed of the motion included from UAV is assumed to be lower than the speed of the potential motion regions (Mliki et al., 2020). Therefore, the motion related to the UAV displacement, which affects dynamic as well as static objects in the scene, is not taken into

consideration. As for the motion related to the objects, it is detected by computing the optical flow using the Lucas-Kande algorithm (Thota et al., 2013) thanks to its speed, simplicity and accuracy (Mliki et al., 2020).

### 3.2.2 Human Detection

The human detection required the generation of human/non-human model to classify potential motion regions. We generated the human/non-human model using the pretrained CNN AlexNet (AIDahoul et al., 2018). In order to adapt the classic CNN to handle the detection problem, we integrated a module of regions of interest generation and selection. The regions of interest generation is performed using the Edge boxes (Zitnick and Dollár, 2014). Since, the generated regions vary slightly in shape, scale or position, we performed on these regions a selection step using the Non-Maximum Suppression(NMS) (Zitnick and Dollár, 2014) algorithm. Such module allows handling the close-up objects, which often appear in the selfsame potential motion region and are usually classified as a single object. Afterward, the selected regions are classified into human/non-human objects. Next the detected humans are tracked using Kalman filter (Welch and Bishop, 1995), in order to get continuous human location.

### 3.2.3 Social Distance Computation

We computed the distance between the detected humans on each frame. According to the computed distance, we used a green bounding box on the detected human if he is away at least two meters from all humans. Otherwise, we used a red bounding box and an alert is triggered. The compute of social distance is based on the estimation of the ground sampling distance (GSD), which defines an image pixel size on the ground. Fig. 3 describes the Ground Sampling Distance parameters.

The  $GSD$  is computed through the following equations:

$$GSD_h[cm/px] = \frac{Altitude[cm] \times Sensorheight[cm]}{Focallength[cm] \times Frameheight[px]} \quad (1)$$

$$GSD_w[cm/px] = \frac{Altitude[cm] \times Sensorwidth[cm]}{Focallength[cm] \times Framewidth[px]} \quad (2)$$

In the case when UAV have usually square pixels, the  $GSD = GSD_h = GSD_w$ . However, when  $GSD_h \neq GSD_w$ , then  $GSD$  takes the maximum value. Then, we extracted the centroid of each human bounding box and we computed the Euclidean distance between the



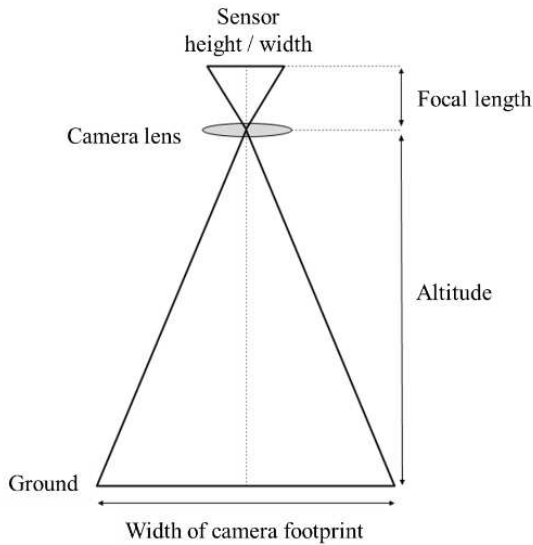


Figure 3: Ground sampling distance parameters.

obtained centroids, as follows:

$$D(p_i, p_j)[px] = \sqrt{(cx_{p_i} - cx_{p_j})^2 + (cy_{p_i} - cy_{p_j})^2} \quad (3)$$

Where,  $(cx_{p_i}, cy_{p_i})$  and  $(cx_{p_j}, cy_{p_j})$  are respectively the coordinates of the centroid of the detected person  $p_i$  and  $p_j$ . In order to convert the distance in centimeters, we multiply it by the *GSD*.

## 4 EXPERIMENTAL STUDY

In this section, we evaluated the performance of the proposed approach for crowd behavior analysis from UAV-captured video sequences in order to monitor social distancing. In the following sections, we described the used datasets and the experimental results.

### 4.1 Datasets Description

The assessment of the proposed macroscopic method was achieved on Mayenberg's dataset (Meynberg and Kusch, 2013; Meynberg et al., 2016) and Mliki's dataset (Hazar et al., 2019), while the evaluation of the microscopic method was carried out on the UCF-ARG dataset (Nagendran et al., 2010). These datasets were captured by UAV in a controlled environment.

#### 4.1.1 Mayenberg's Dataset

The Mayenberg's dataset (Meynberg and Kusch, 2013; Meynberg et al., 2016), was captured, through two flight campaigns, in open-air rock Germany festivals. This dataset includes 70.000 patches corresponding to one of the crowd classes: Dense,

Medium, Sparse and None. It shows minor variations in the UAV-captured crowd images.

#### 4.1.2 Mliki's Dataset

This dataset was collected over the net and it represents various contexts as: festival, marathon, manifestation and pilgrimage (Hazar et al., 2019). It is composed of 12,000 UAV-captured patches belong to the crowd classes: Dense, Medium, Sparse and None. The Mliki's dataset shows more complex variations than Mayenberg's dataset.

#### 4.1.3 UCF-ARG Dataset

The UCF-ARG dataset (Nagendran et al., 2010), includes 482 video sequences of human activities and it deals with several constraints such as: the variation of the UAV altitude, the severe ego-motion, the variation in the point of view and the illumination variation conditions.

## 4.2 Experimental Results

We carried out two series of experiments: The first series of experiments aimed to assess the macroscopic method for crowd density estimation, and the second series of experiments evaluated the microscopic method for human detection and social distance computing.

### 4.2.1 First Series of Experiments: Crowd Density Estimation (Macroscopic Method)

Through this series of experiments, we assessed the performance of the proposed macroscopic method for crowd density classification with the state-of-the-art reference methods (Hazar et al., 2019; Meynberg and Kusch, 2013; Meynberg et al., 2016). For the purpose of assuring fair comparison with Mayenberg's (Meynberg and Kusch, 2013; Meynberg et al., 2016) and Mliki's (Hazar et al., 2019) methods, we applied the same experimental protocol. In fact, we used, for each dataset, 110 training crowd patches and 500 testing crowd patches per class. Table 1 reported this evaluation study in terms of accuracy rate.

Referring to this study, we noted that our results surpassed those obtained by (Meynberg and Kusch, 2013; Meynberg et al., 2016) owing to the use of CNN and transfer learning, which deal with the UAV constraints like the variation in: the GSD, the resolution, the illumination and the view angle. Nonetheless, Mliki et al. (Hazar et al., 2019) slightly exceed our method, this is justified by the fact of applying a three-level classification strategy to deal with the

Table 1: Comparison of our macroscopic method for crowd density estimation with the state-of-the-art reference methods in terms of accuracy rate.

Datasets	Methods	Accuracy
Mayenberg	Meynberg et al., 2013	62.3 %
	Mayenberg et al., 2016	74.67 %
	Mliki et al., 2019	89.2 %
	Our method	<b>84.85 %</b>
Mliki	Mayenberg et al., 2016	71.8%
	Mliki et al., 2019	86.8 %
	Our method	<b>82.1 %</b>

crowd density class confusion. Although this strategy enhanced the performance results, it has significantly increased the computational complexity making it not suitable for real-time context.

Table 2 illustrates the run-time gain of our macroscopic method while using N parallel process. Such gain is critical for real-time applications.

Table 2: Evaluation of the parallel process for crowd density estimation on Mliki's dataset (Hazar et al., 2019) in terms of sec/frame.

Material configuration	Execution	Sec/frame
i5 2.4 GHZ CPU (N=2)	Sequential	6.40
8 GB RAM	Parallel	5.60

Figure 4 represents some samples results of our macroscopic method on the Mliki's dataset (Hazar et al., 2019). More qualitative results of our macroscopic method are available on this link: [video1](#).

#### 4.2.2 Second Series of Experiments: Human Detection and Distance Computation (Microscopic Method)

Through this series of experiments, we evaluated the performance of the proposed microscopic method for human detection. For fair comparison with Al-Dahoul et al. (AlDahoul et al., 2018), we used their experimental protocol. Table 3 highlighted this comparative study performed on the UCF-ARG dataset (Nagendran et al., 2010).

Table 3: Comparison of our microscopic method for human detection with (AlDahoul et al., 2018) method in terms of accuracy rate.

Methods	Accuracy
Al-Dahoul et al., 2018	98.0907%
Our method	99.5873 %

Through this experimental study, we underline that our method outperformed the obtained results of the AlDahoul et al. (AlDahoul et al., 2018) in terms of accuracy rate. Such results highlighted the contribution of using the regions of interest generation and selection module, which overcame the problem of close objects classification. Figure 5 shows some results of our social distance computing on the UCF-ARG dataset (Nagendran et al., 2010), where the red boxes indicate the non-conformity of people to the social distancing rule and green boxes refer to the respect of the social distancing rule.

Further qualitative results of our microscopic method are available on this link: [video2](#).

## 5 CONCLUSION

The social distancing represents is the most recommended strategy that allows the mitigating of the COVID-19 pandemic propagation. In this context, we have proposed a new approach that aims to analyze crowd behavior from UAV-captured video sequences to monitor social distancing. Our approach involves two methods: a macroscopic method and a microscopic method. The macroscopic method allows estimating the crowd density by classifying the aerial frames patches into four categories: Dense, Sparse, Medium and None. The microscopic method aims to detect and track humans so as to compute the distance between them. Through the quantitative and qualitative evaluations of the proposed methods, we prove the performance of our approach compared to the existing ones.

As future perspectives, we aim to improve the proposed crowd behavior analysis approach by adding a person re-identification step in order to re-identify persons who are non-compliance with social distancing measures. Furthermore, we are planning to integrate the person temperature measurement by using thermal UAV.

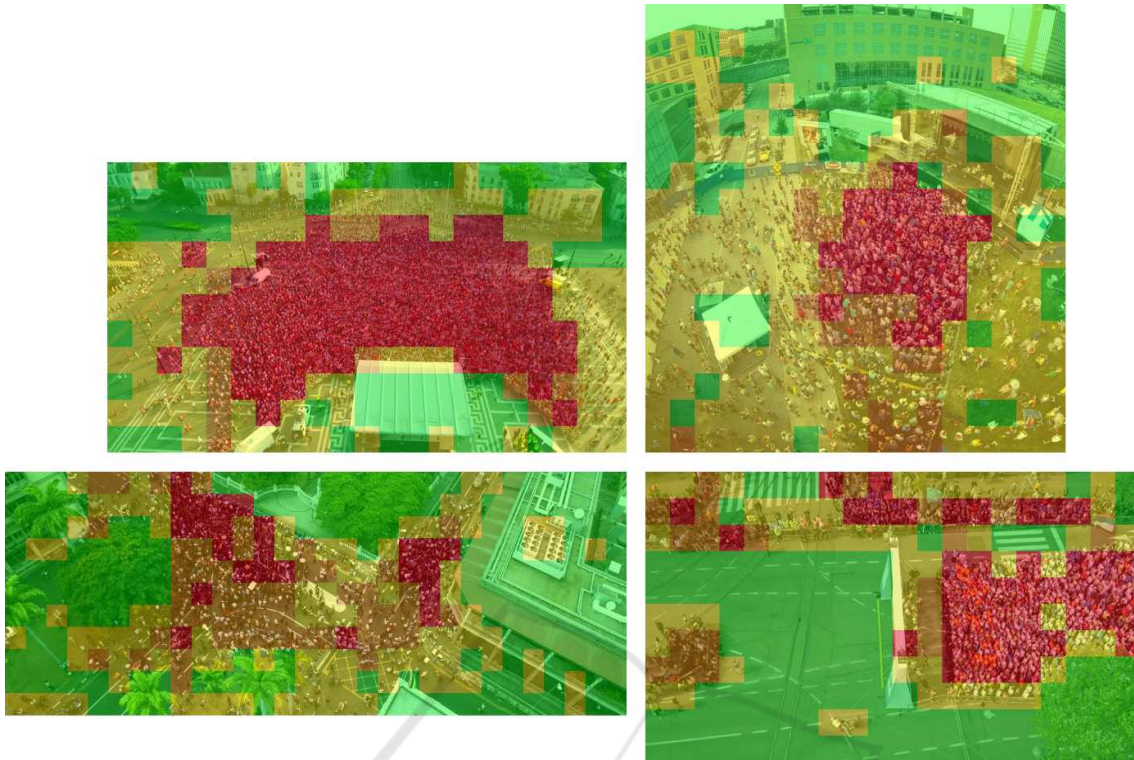


Figure 4: Crowd density estimation samples using the proposed macroscopic method on the Mliki's dataset (Hazar et al., 2019).



Figure 5: Humans detection, tracking and social distance monitoring samples using the proposed macroscopic method on the UCF-ARG dataset (Nagendran et al., 2010).



## REFERENCES

- AlDahoul, N., Md Sabri, A. Q., and Mansoor, A. M. (2018). Real-time human detection for aerial captured video sequences via deep models. *Computational intelligence and neuroscience*, 2018.
- Cristani, M., Del Bue, A., Murino, V., Setti, F., and Vinciarelli, A. (2020). The visual social distancing problem. *arXiv preprint arXiv:2005.04813*.
- Hazar, M., Arous, O., and Hammami, M. (2019). Abnormal crowd density estimation in aerial images. *Journal of Electronic Imaging*, 28(1):013047.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Lewnard, J. A. and Lo, N. C. (2020). Scientific and ethical basis for social-distancing interventions against covid-19. *The Lancet. Infectious diseases*, 20(6):631.
- Meynberg, O., Cui, S., and Reinartz, P. (2016). Detection of high-density crowds in aerial images using texture classification. *Remote Sensing*, 8(6):470.
- Meynberg, O. and Kuschik, G. (2013). Airborne crowd density estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:49–54.
- Mliki, H., Bouhlel, F., and Hammami, M. (2020). Human activity recognition from uav-captured video sequences. *Pattern Recognition*, 100:107140.
- Nagendran, A., Harper, D., and Shah, M. (2010). UCF-ARG dataset, university of central florida. <http://csrc.ucf.edu/data/UCF-ARG.php>.
- Park, S. W., Sun, K., Viboud, C., Grenfell, B. T., and Dushoff, J. (2020). Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (covid-19) in south korea. *medRxiv*.
- Punn, N. S., Sonbhadra, S. K., and Agarwal, S. (2020). Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques. *arXiv preprint arXiv:2005.01385*.
- Repici, A., Maselli, R., Colombo, M., Gabbiadini, R., Spadaccini, M., Anderloni, A., Carrara, S., Fugazza, A., Di Leo, M., and Galtieri, P. A. (2020). Coronavirus (covid-19) outbreak: what the department of endoscopy should know. *Gastrointestinal endoscopy*.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Singh, R. and Adhikari, R. (2020). Age-structured impact of social distancing on the covid-19 epidemic in india. *arXiv preprint arXiv:2003.12055*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- Thota, S. D., Vemulapalli, K. S., Chintalapati, K., and Gudipudi, P. S. S. (2013). Comparison between the optical flow computational techniques. *International Journal of Engineering Trends and Technology*, 4(10).
- Welch, G. and Bishop, G. (1995). An introduction to the kalman filter.
- WHO (Septembre 4, 2020). World health organization corona-viruses (covid-19). <https://covid19.who.int/>.
- Wilder-Smith, A. and Freedman, D. O. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-ncov) outbreak. *Journal of travel medicine*, 27(2):taaa020.
- Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.
- Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., Wu, Q., Merler, S., Viboud, C., and Vespignani, A. (2020). Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in china. *medRxiv*.
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., and Gu, X. (2020). Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet*.
- Zitnick, C. L. and Dollár, P. (2014). Edge boxes: Locating object proposals from edges. In *European conference on computer vision*, pages 391–405. Springer.