Implementation of Object Detection Method for Intelligent Surveillance Systems at the Faculty of Engineering, Universitas Sebelas Maret (UNS) Surakarta

1st Aris Maulana Fauzan
Dept. Electrical Engineering
Sebelas Maret University
Surakarta, Indonesia
06.fauzan@gmail.com

2nd Sutrisno Ibrahim*)
Dept. Electrical Engineering
Sebelas Maret University
Surakarta, Indonesia
sutrisno@staff.uns.ac.id

3rd Meiyanto Eko Sulistyo
Dept. Electrical Engineering
Sebelas Maret University
Surakarta, Indonesia
mekosulistyo@staff.uns.ac.id

)Corespoding Author

***Abstract*—The number of positive Covid-19 cases in Indonesia continue to increase. This increase influenced by the behavior of Indonesian citizens in dealing with the pandemic, one of which is rarely wearing masks. In this study, we implemented an object detection method for intelligent surveillance systems (ISS) at the Faculty of Engineering, Universitas Sebelas Maret (UNS), Surakarta. By implementing face detection and mask detection, the surveillance system can recognize whether a person in a CCTV video frame is wearing a mask or not. In addition, deep metric learning and histogram of gradient (HOG) are applied to recognize faces of unmasked people in images. The test results show that the surveillance system can recognize the use of masks with 75%-87% accuracy rate. Furthermore, the accuracy rate for facial recognition on images ranges from 69% -100% for each person.**

Keywords— intelligent surveillance systems, object detecting, covid, CCTV, HoG, image processing

I. Introduction

Buildings or infrastructures that are important and visited frequently by people such as hospitals, airports, banks, schools, and jewelry stores usually have a surveillance system. This surveillance system is not only for security purposes, it can also be used as evidence if a crime occurs in the building. However, currently most security systems are monitored manually by human operators. According to Sulman et al. [1], monitoring manually by human operators is an inefficient solution, even impractical because apart from expensive human resources, humans have limited capabilities [1].

One way to increase efficiency in surveillance systems is to use Intelligent Surveillance Systems (ISS). ISS is a surveillance system that can automatically analyze images, video, audio or other types of surveillance data with as little or no human intervention as possible [2]. ISS is an important component for developing smart cities, smart traffic systems, or smart buildings [3] especially for the security and safety side. This camera-based surveillance system includes background-foreground segmentation, object detection and classification, tracking, and behavior analysis [2].

To support the iss to work optimally, a good supporting sensor is needed. The most widely used sensors for surveillance in indonesia are cctv cameras. Even though high-resolution cctv is available and zoom and tilt capabilities are available, cctv only performs action and records video. The available surveillance system has not considered further actions such as detecting the faces of people recorded on cctv.

The ability of surveillance systems to detect faces is increasingly important for the spread of the Covid-19 virus. As of October 8, 2020, the number of positive cases in Indonesia increased by 4,850 to a total of 320,564 cases [4]. If you look at the track record on the Ministry of Health website, positive cases of Covid-19 continue to increase. This is influenced by the pattern of behavior of Indonesian citizens in dealing with the pandemic. Based on the results of a survey by the Central Statistics Agency on September 7-14, 2020 [5], as many as 6% of respondents out of a total of 90.967 respondents rarely wear masks with 2.02% of respondents never even wearing masks. In fact, wearing a mask is one of the efforts to reduce the rate of transmission of the Covid-19 virus. Even in the area of the Faculty of Engineering, UNS, there are still some members of the community, staff, and certain officers who do not use.

II. METHODS

An intelligent surveillance system (ISS) is a surveillance system that has the ability to automatically analyze surveillance data, and take follow-up actions, such as turning on an alarm. ISS has a close relationship with computer vision, pattern recognition, artificial intelligence, machine learning, and communication. As shown in Figure 1, there are five basic steps in a ISS system: data acquisition, foreground-background segmentation [6-7], object detection and classification [8-9], object tracking [10-11] and behavioral analysis [12-13]. Foreground-background segmentation is the first important step for intelligent surveillance system. The goal is to separate the object or moving object (foreground) and the environment (background).

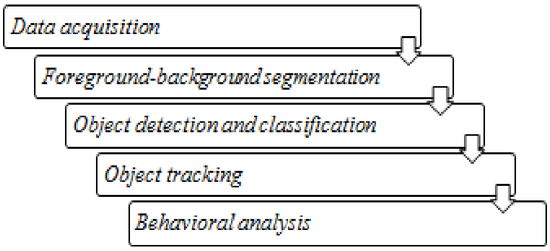
Data acquisition
Foreground-background segmentation
Object detection and classification
Object tracking
Behavioral analysis

Fig. 1. Data processing steps in ISS

A. Data

At the end of 2018, the Faculty of Engineering installed a CCTV system to increase security in the faculty area. The CCTV cameras are spread over 5 (five) points which include: engineering valley, parking lot, front yard of building 3, lobby of building 3, and the dean's hallway in building 3 of the Faculty of Engineering. Figure 2 and 3 shows the CCTV configuration at the Faculty of Engineering UNS with PUSKOM as the communication center at UNS, FO is fiber optic cable, and UTP is UTP cable [14,15].

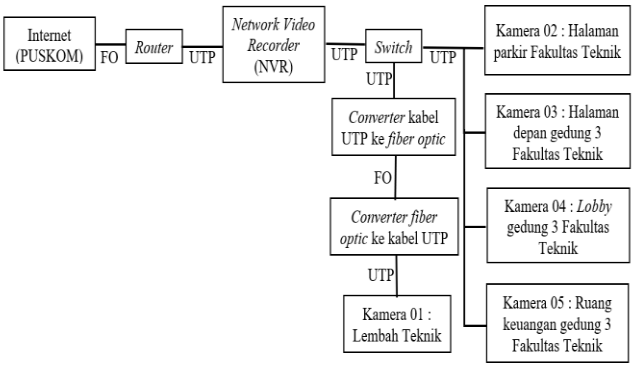
Internet (PUSKOM)
FO
Router
UTP
Network Video Recorder (NVR)
UTP
Switch
UTP
UTP
Converter kabel UTP ke fiber optic
FO
Converter fiber optic ke kabel UTP
UTP
Kamera 01 : Lembah Teknik
Kamera 02 : Halaman parkir Fakultas Teknik
Kamera 03 : Halaman depan gedung 3 Fakultas Teknik
Kamera 04 : Lobby gedung 3 Fakultas Teknik
Kamera 05 : Ruang keuangan gedung 3 Fakultas Teknik

Fig. 2. CCTV configuration at Faculty of Engineering UNS

Fig. 3. CCTV system at Faculty of Engineering UNS

Apart from Faculty of Engineering CCTV data, the data used in this study were obtained from various sources, the details of data collection are as follows.

TABLE I. DATA

Data	References
1	Faculty of Engineering CCTV data
2	Bing Search API (dataset)
3	Adnane Cabani dkk. [16]
4	Prajna Bhandary [17]
5	Bing Search API (database)
6	Personal data

B. Methods

Transfer Learning

Transfer learning is a method of applying a network to an existing classifier as a starting point for the classification and detection of different objects. In other words, transfer learning allows the classifier trained to distinguish between cats and dogs to be used to distinguish between cars and bicycles even though cars and bicycles were not used in the initial training. So that the network can be used in the new object classification, a fine tuning process is carried out. In this study, the network architecture of the imported MobileNetV2 is used by removing the head layer as the base model. Next is the fine tuning process by creating a new head layer. From this fine tuning process, a model will be obtained to detect the use of masks.

DNS Face Detector

DNN Face Detector which is a face detection model built with an SSD framework based on the ResNet-10 network architecture. This model is trained in the Caffe framework for the WIDER FACE dataset with 140,000 iterations. As the name implies, this model is used to detect faces in a frame or image.

Histogram of Oriented Gradient (HOG)

Histogram of oriented gradients (HOG) is a method in machine learning that is used to extract features on objects, in this case human faces. Based on the steps, the initial process in the HOG method is to convert the RGB (Red, Green, Blue) image to grayscale, then proceed with calculating the gradient value of each pixel. As for obtaining the gradient value of each pixel, the steps taken are to compare each pixel in the image with the surrounding pixels. This is done to find out how dark a pixel is when compared to the pixels that surround it. After that, the algorithm will place an arrow image indicating which direction the image is getting darker in. This process is repeated for each pixel in the image until all pixels are replaced with arrows. This arrow is referred to as a gradient and shows the flow of color from light to dark throughout the image. To find faces in an image, HOG searches for the part of the image that most closely resembles a known HOG pattern extracted from a set of training faces.

Deep Metric Learning

Deep metric learning (distance matrix learning) is used to create face embeddings in the face recognition process. In principle, metric learning aims to measure the similarity between samples by paying attention to the optimal distance metric between samples. The output of metric learning is a 128-d number (a list of 128 numbers) vector that represents face attributes. The training process runs by looking at 3 face images at once (single triplet training step), namely:

1. Image of a recognized person's face training (the anchor)

2. Another picture of the same person (positive image)

3. Pictures of different people (negative image)

Each image will produce a 128-d number which is the result of measuring the attributes of each image. Then the weights of the neural network will be changed slightly so as to ensure the resulting measurements the same person or different people.

III. RESULTS AND DISCUSSION

The proposed method is tested in two scenarios: real system (CCTV at Faculty of Engineering) and on dataset available in the internet.

A. *Faculty of Engineering CCTV data*

Face detection is applied to 3 (three) cameras: camera #02 which is located in the parking lot of building 3 of the Faculty of Engineering, camera #03 which is located in the front yard of building 3 of the Faculty of Engineering, and camera #04 which is located in the lobby of building 3 of the Faculty of Engineering. The "No Mask" label indicates the person in the frame is not wearing a mask. In contrast, people wearing masks are identified with the label "With Mask".

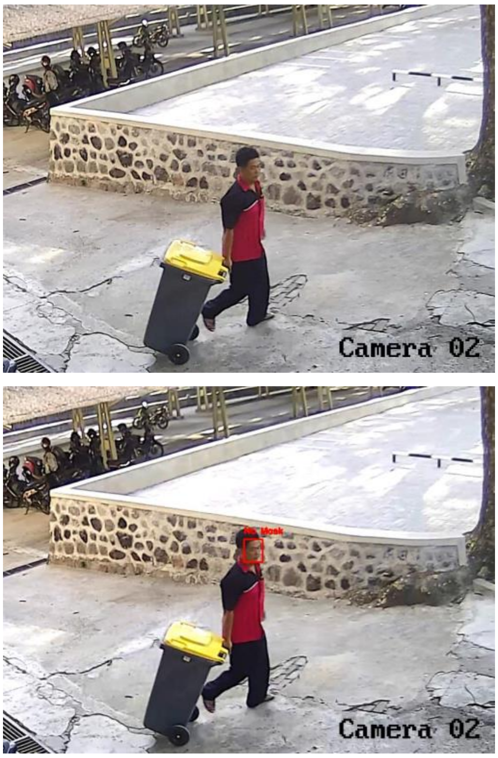

Fig. 4. Results in camera #02

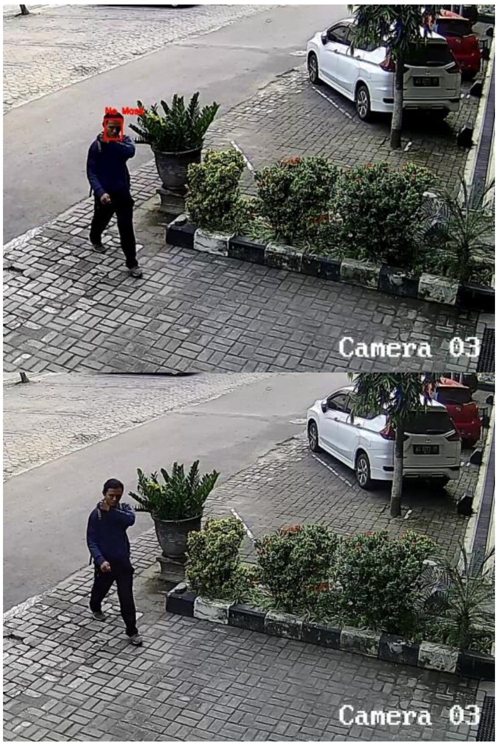

Fig. 5. Results in camera #03

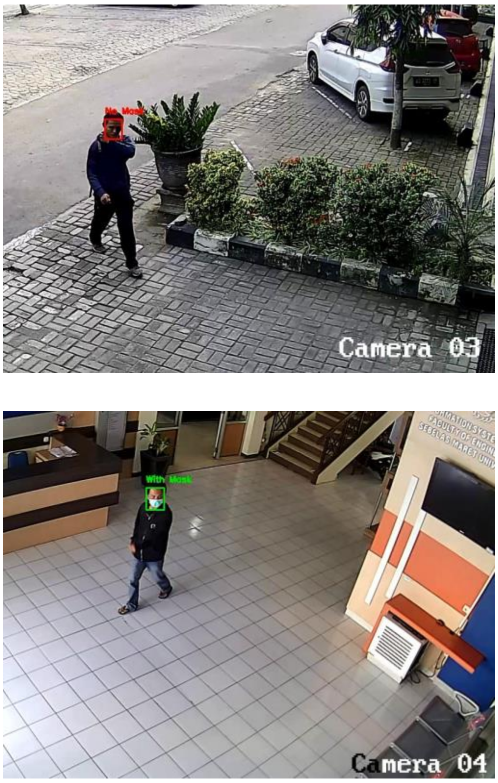

Fig. 6. Results in camera #04

From the results of the implementation of face detection on CCTV cameras, the results detected on camera #04 show better results. In contrast to camera #02 and camera #03 which are outdoors, camera 04 is installed indoors so that the effect of sunlight on the video frame is not too large. With less sunlight, the recognition of “With Mask” and “No Mask” becomes more accurate. In addition, the position of a person in the frame also affects the level of accuracy. This is because the system will be better at detecting faces that are more or less parallel to or facing the camera. While in installation, the CCTV of the Faculty of Engineering UNS is installed in the corners of the building so that the captured image looks higher – considering that its initial purpose was for security.

TABLE II. ACCURACY ON CCTV DATA

Data	Accuracy
camera #02	87,5%
camera #03	75,0%
camera #04	86,4%

B. Implementation on dataset

Detection in images is divided into two, namely detection for images of people's faces with masks and detection for images of people without masks. Detection of face images of people with masks uses more or less the same method as detection on CCTV. From the detection results, it can be seen that face recognition with masks can be implemented properly. Even the face of the person facing the side in the third image is recognizable.

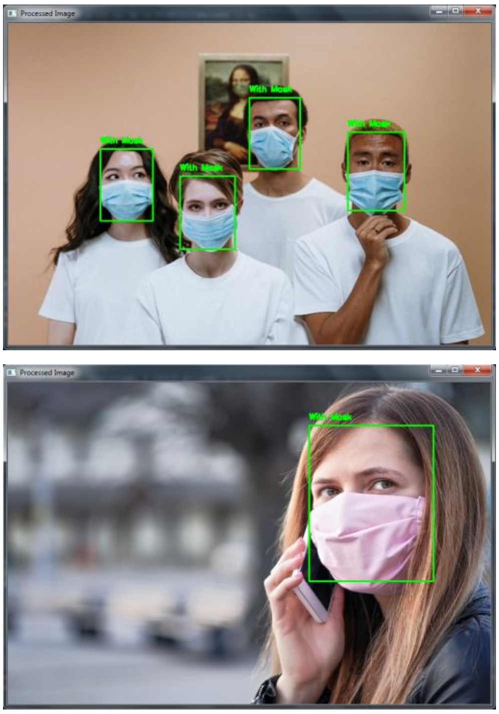

Fig. 7. Detected as “with mask”

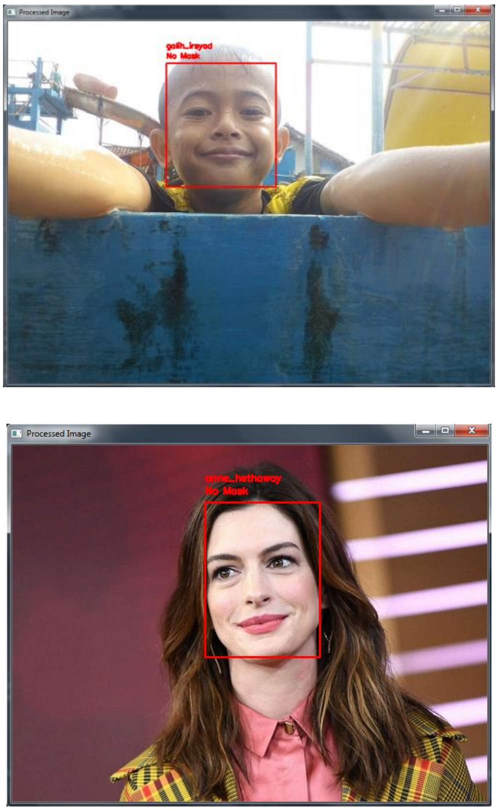

Fig. 8. Detected as “no mask”

The detection of images without masks begins with the same method as detection on CCTV so that it will be known whether the person in the video frame is wearing a mask or not. Because the person in the image frame is not wearing a mask, facial recognition of the person who is not wearing the mask will be carried out according to a database that has been processed using the deep metric learning method. If someone in the image frame is not available in the database, then the system will not be able to recognize that person so it will be labeled “unknown”.

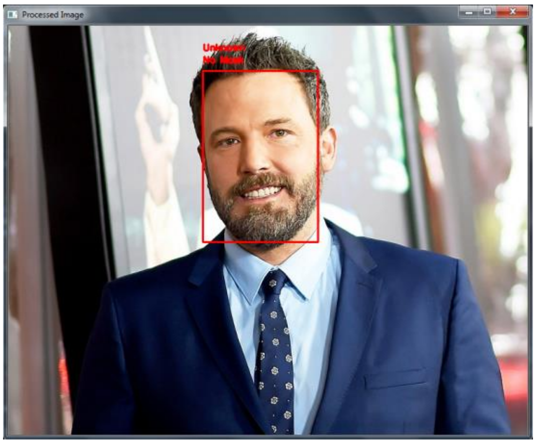

Fig. 9. Ben Affleck unrecognized or “unknown”

From the results of the implementation test in the image, the results obtained that Ratu Riyaning's accuracy value is very small when compared to the accuracy values of other people. This can be caused by the lack of a database used in training. As previously discussed, database training is carried out using deep metric learning techniques that will generate measurement values for each trained image. So that the more training data, the more accurate the measurement value will be. In addition, image quality also affects the results obtained. The clearer the image used in training and in testing, the better the level of accuracy obtained.

TABLE III. ACCURACY ON CCTV DATA

Data	Accuracy
Anne Hethaway	100%
Christian Bale	100%
Dwayne Johnson	100%
Febri Abdul	94%
Galih Irsyad	95%
Iko Uwais	100%
Jason Statham	100%
Michael Caine	86%
Nisa Nurul	94%
Ratu Riyaning	69%

IV. CONCLUSIONS

From the research that has been done, it can be concluded that face detection has been implemented for the purposes of the surveillance system at the Faculty of Engineering, UNS by considering the use of masks. The detection accuracy rates on camera #02, camera #03, and camera #04 are 87%, 75%, and 86%, respectively. In addition, facial recognition is also applied to images with the highest accuracy being 100% of the images of the faces of artists including Anne Hethaway, Christian Bale, Dwayne Johnson, Iko Uwais, and Jason Statham. While the lowest level of accuracy is 69% of Ratu Riyaning's face. The direction of future research is to increase accuracy by increasing the amount of data used in training. In addition, additional features can also be implemented, such as warnings when someone is not wearing a mask.

REFERENCES

[1] G. N. Sulman, T. Sanocki, D. Goldgof and R. Kasturi, "How effective is human video surveillance performance?," 19th International Conference on Pattern Recognition, 2008. ICPR 2008, Tampa, FL, 2008, pp. 1-3.R.

[2] S. Ibrahim, “A comprehensive review on intelligent surveillance systems,” Communications in Science and Technology, vol. 1, pp. 7-14, 2016.

[3] R.A. Shahad, M.H.M. Saad, A. Hussain, "Activity Recognition for Smart Building Application Using Complex Event Processing Approach, International Journal on Advanced Science, Engineering and Information Technology," vol. 8, pp. 315-322, 2018.

[4] Kementrian Kesehatan. " Situasi Terkini Perkembangan Coronavirus Disease Covid-19." [Online]. Available: https://covid19.kemkes.go.id/situasi-infeksi-emerging/info-corona-virus/situasi-terkini-perkembangan-coronavirus-disease-covid-19-09-oktober-2020/#.X4AXfRIxXIU [Accesed: 9 October 2020]

[5] Badan Pusat Statistik. "Perilaku Masyarakat di Masa Pandemi Covid-19." [Online]. Available: https://www.bps.go.id/publication/2020/09/28/f376dc 33cfcdeec4a514f09c/perilaku-masyarakat-di-masa-pandemi-covid-19.html [Accesed 9 October 2020]

[6] M. Cristani, M. Farenzena, D. Bloisi, V. Murino, “Background Subtraction for Automated Multisensor Surveillance: A Comprehensive Review,” *EURASIP Journal on Advances in Signal Processing*, 24 pages, Volume 2010, 2010.

[7] T. Bouwmans, F. El Baf, and B. Vachon, “Statistical background modeling for foreground detection: A survey,” *in Handbook of Pattern Recognition and Computer Vision (Volume 4)*. Singapore: World Scientific, Jan. 2010, ch. 3, pp. 181–199.

[8] M. Paul, S. M. Haque and S. Chakraborty, "Human detection in surveillance videos and its applications-a review", EURASIP Journal onar Advances in Signal Processing, vol. 2013, no. 1, pp. 1-16, 2013

[9] M. Enzweiler and D. M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” IEEE Trans. Pattern Analysis and Machine Intelligence, 2009. 1, 2, 6, 9, 14, 15, 18.

[10] A. Yilmaz, M. Shah, Object tracking: A survey. Journal ACM Computing Surveys 38(4) (2006).

[11] Arnold W. M. Smeulders, Dung M. Chu, Rita Cucchiara, Simone Calderara, Afshin Dehghan and, and Mubarak Shah, Visual Tracking: an Experimental Survey, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 36, NO. 7, July 2014.

[12] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," 37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR '08), pp.1,8, 15-17 Oct. 2008. doi: 10.1109/AIPR.2008.4906450.

[13] M. Cristani, R. Raghavendra, A. Del Bue, V. Murino, Human behavior analysis in video surveillance: A Social Signal Processing perspective, Neurocomputing, Volume 100, 16 January 2013, Pages 86-97, ISSN 0925-2312.

[14] Wahyu Kurniawan, “Implementasi Metode People Tracking untuk Sistem Pengawasan Cerdas di Fakultas Teknik Universitas Sebelas Maret (UNS) Surakarta,” Final Project (2019).

[15] Wahyu Kurniawan, Sutrisno Ibrahim, and Meiyanto Sulistyo, “People detection and tracking methods for intelligent surveillance system,” AIP Conference Proceedings 2217, 030110 (2020); https://doi.org/10.1063/5.0001022

[16] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. "MaskedFace-Net--A Dataset of Correctly/Incorrectly Masked Face Images in the Context of COVID-19." arXiv preprint arXiv:2008.08016 (2020).

[17] Prajna Bhandary. "Mask Classifier." [Online]. Available: https://github.com/prajnasb/observations [Accesed 15 September 2020]