An Efficient Scheme to Obtain Background Image in Video for YOLO-based Static Object Recognition

Hyeong-Jin Kim, Min-Cheol Shin, Man-Wook Han,
Chung-pyo Hong and Ho-Woong Lee*

Division of Computer Engineering, Hoseo University, Republic of Korea
E-mail: psvm9514@gmail.com; shinmc9@gmail.com; aksdnr507@gmail.com; cphong@hoseo.edu; always14@hoseo.edu
**Corresponding Author*

Received 14 February 2022; Accepted 07 July 2022;
Publication 27 August 2022

Abstract

Detecting backgrounds in videos is an important technology that can be used for many applications such as management of major facilities and military surveillance depending on the purpose. It is difficult to accurately find and identify important objects in the background if there are obstacles such as pedestrian or car in the video. In order to overcome this problem, the following method is used to detect the background. First, a pixel area histogram is generated to determine the amount of change in pixel units of an image over time. Based on the histogram, we propose an algorithm that estimates the background by selecting the case with the smallest rate of change. In addition, in order to strongly respond to changes in the surrounding environment, even when a change in brightness occurs, this is solved through frame overlap. Finally, the desired object is identified by applying YOLO v3 as a model for object detection in the obtained background. Through the above process, this study proposes a method for effectively identifying static objects in the background by precisely estimated background of the video. Experimental

Journal of Web Engineering, Vol. 21_5, 1691–1706.
doi: 10.13052/jwe1540-9589.21513

results show that the non-detection and false detection rate for the background object is enhanced by 60.2% and 11.2%, respectively, in comparison with when the proposed method was not applied.

Keywords: Background obtainment, histogram, object detection, YOLO.

1 Introduction

Technology for detecting backgrounds except moving objects in videos has long been studied in the field of computer vision. This is an important technology that can be used for many applications such as management of major facilities and military surveillance depending on the purpose. In the video, the movement of various objects or human movements can cover the background. For this reason, if the object in the foreground is not removed, it is difficult to accurately find and identify important objects in the background. In order to overcome this problem, the process of obtaining only accurate background information must precede. Various existing background extraction algorithms have limitations due to many obstacles such as dynamic background changes, lighting changes, and occlusion.

Recently, research applying deep learning techniques to detect background is being actively conducted. However, it has the disadvantage of requiring relatively many operations. In order to utilize deep learning techniques, numerous training data must be established. In some cases, training may be difficult if limited data is required for collection. Medical data is an example. In addition, repetitive tasks such as data labelling and appropriate threshold selection tasks through hyperparameter adjustment are required. Moreover, it may require a high level of hardware performance. In other words, deep learning techniques have the disadvantage of having multiple overheads. Therefore, rather than solving problems arising from only deep learning techniques, hybrid methodologies that estimate necessary backgrounds through computer vision technology and recognize necessary objects through deep learning techniques are being considered. This may be possible to recognize high-performance video background objects at low cost through a combination of computer vision technology and deep learning techniques.

In this study, the following method is used to detect the background. First, a pixel area histogram is generated to determine the amount of change in pixel units of an image over time. Based on the histogram, we propose an algorithm that estimates the background by selecting the case with the smallest rate of change. In addition, in order to strongly respond to changes

in the surrounding environment, even when a change in brightness occurs, this is solved through frame overlap. Finally, the desired object is identified by applying YOLO v3 as a model for object detection in the estimated background. Through the above process, this study proposes a method for effectively identifying static objects in the background by precisely securing the background of the video.

The rest of this paper consists of 4 parts. Section 2 introduces the related work. And the Proposed Scheme is presented in Section 3. Section 4 shows the experimental results, and finally, the conclusion is described in Section 5.

2 Related Work

In this section, we discuss the need for object detection in the background and related studies for it. Figure 1 shows an example in which proper detection was not achieved due to an object passing by an object existing in the background. If an object is covered when a car passes by, humans can recognize that there is an object behind it, but there is no way to detect it because the simple object detection process obscures the object in the background. In addition to dynamic foreground objects, there are limitations due to many obstacles such as dynamic background change, lighting change, and occlusion [1]. Recently, studies using deep learning techniques for background detection have been actively conducted [2]. Background estimation machine learning algorithm methodology using Gaussian techniques such as Mixture Of Gaussian (MOG) and Gaussian Mixture Model (GMM) is the most representatively used, and there are studies to improve it [3]. However, this method has a disadvantage in that there are various overheads. Due to these limitations, it can be a problem due to various obstacles such as moving objects or bad weather in various environments that require autonomous driving or precise object detection [4].

Figure 1 Example of failure to detect objects in the background.

In other studies, for background inference, a method of applying a different algorithm to each layer by dividing the layers for inference by step has been proposed [5]. In order to improve the time-consuming problem of object detection in large-volume video and the difficult problem of predicting multiple objects that change according to the direction and time of the object, a method to solve it by dividing it into each step was proposed [6]. Fast Multi-Level proposed an algorithm to improve performance with fast computation time and less memory usage for foreground detection [7]. A new histogram pixel modelling with fuzzy segmentation applied for background removal has been studied [8]. We proposed a method for detecting and tracking in complex road traffic using background detection and improved the problem of tracking multiple moving objects using a Kalman filter and counting one object repeatedly [9]. A study was conducted to track the detected object using background subtraction and Kalman filter techniques [10]. Research has been conducted to significantly reduce the calculation of the MOG-based background removal algorithm through image resizing and to increase the performance by using local information [11]. A study was conducted to monitor the suspected landslide area in real time using GMM [12]. Various studies have been conducted on learning convolutional networks for background detection and background subtraction [13, 14].

In addition, various studies talked about the publication of many studies on background detection and object detection in the deep learning field, organized CNN papers on background detection, and talked about additional directions [15]. We proposed an encoder-decoder fully convolutional neural network architecture to fuse the results generated by combining various background detection algorithms and output more accurate results [16].

As previously described, a dynamically moving foreground becomes an obstacle to background detection. As a study to complement this, an effective background subtraction method was proposed by combining colours and texture features [13, 17]. A specific study has proposed a method of using a fuzzy model for data uncertainty for moving foreground detection [18].

You Only Look Once (YOLO) v3 proposes a method to find objects effectively in still images [18]. However, YOLO does not accurately recognize the background object when the background is obscured by a moving object in the video.

However, in this paper, we will combine the computer vision method and deep learning method to recognize any given object of interest which exists in the background. We will obtain the background with the computer vision method, and also we will recognize the static object with the deep learning

technique. In particular, the computer vision method is described in detail through the rest of this paper.

3 Proposed Scheme

3.1 Frame Overlay

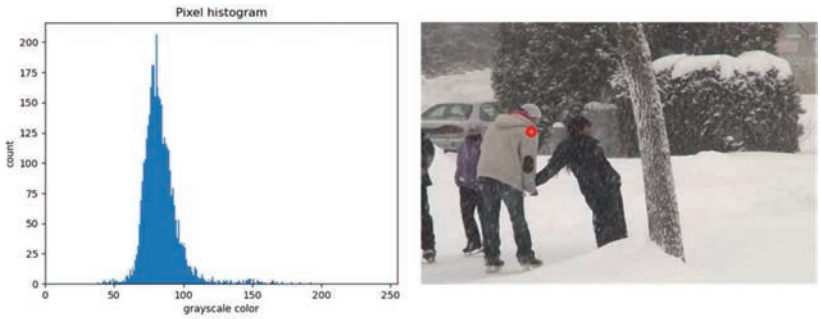
A given input image has frames according to the passage of time. Each frame consists of the colours of a pixel area. By overlapping the frames of the input image, it is possible to create an overlapping array of frames with time-dependent information. As the amount of information accumulated in the histogram by overlapping frames in the image increases, the accuracy increases. This gives better results under normal circumstances when many frames are stacked in a small-time span.

However, as many frames are overlapped, the calculation through the histogram also increases. Consequently, balancing the time gap and the threshold of the number of overlapping frames affects speed and accuracy. Therefore, it is important to set an appropriate threshold. As time goes by, older information is removed in the oldest order. Then, the new frame is overlapped. When new frames overlap, the histogram of the pixel area is updated to estimate a new background. This process also helps them adapt to changes in weather and light. It uses a frame overlapping array that is maintained for a certain period of time and updated periodically by containing the amount of change in the pixel area over time.

3.2 Pixel Histogram

A frame has the form of a pixel area x, y, channel. Through the process of Section 3.1, the overlapped frame containing the change amount of the pixel area over time has a time,x,y dimensional vector. Also, each pixel area contains a gray-scaled pixel value from 0 to 255. The histogram has an amount of change according to the time flow for each channel. Each pixel has a histogram of its variation. The accumulated amount of change can be identified through the histogram overlapped in each pixel area. Using the overlapped histogram information, the pixel value of the histogram with the highest probability is selected and used for background estimation.

In the overlapped array, each pixel area of x, y has overlapped change amount information over time. The background pixel of the pixel area is inferred using the most frequent value among this information. Figure 2(a) shows an example histogram of overlay pixels obtained by overlapping. The



(a)

(b)

Figure 2 An example of pixel histogram representation.

area indicated by the red circle in the figure on the right means a single pixel of 150,150 coordinates. The graph on the left means a grayscale histogram of the entire video frame of 150 and 150 coordinate pixels. The dominant value of the histogram is defined as the color value of the corresponding position of the background image. A background image is created based on the pixel color value obtained after performing the same operation on all coordinates in the video. Figure 2(b) shows sample frames in which the dominant color values of 150, 150 coordinate pixels appear.

3.3 Background Obtainment

Through Section 3.2, the background is obtained by merging the pixels inferred using the histogram of the unit pixel through the background estimation process in the pixel area. Figure 3 is a pseudocode of the algorithm

Algorithm : Proposed Background Obtainment Algorithm

```
overlay_frames = queue                    // overlayed frame queueby elapsed time
skip_overlay_frame = threshold      // frames to skip
overlay_threshold = threshold    // number of frames to be overlayed
estimation_background = matrix // estimated background matrix
change_ threshold = threshold // histogram variation threshold
// Frame overlap (when the number of overlapping frames reaches the limit, old frames are
removed and new frames are added)
for idx in video:
      if idx % skip_overlay_frame == 0:
            if len(overlay_frames) == overlay_threshold:
                  overlay_frames.pop()
            overlay_frames.append(video[idx])

// Transforms the overlapped frame based on the time axis
      transpose_overlay_frames = transpose() [t(pixel), width, height] → [width, height, t(pixel)]

// Updates the background estimation matrix by finding the color values that change the most
based on the pixel histogram
      for x in width:
         for y in height:
            color = max_histogram(transpose_overlay_frames[x][y])
         if estimation_background[x][y] > color:
               rate_of_change= (estimation_background[x][y] – color) / 255
                 else :
               rate_of_change = (color - estimation_background[x][y]) / 255
            if rate_of_change > change_ threshold:
                  estimation_background[x][y] = color
```

Figure 3 Background estimation algorithm.

process for the proposed method. For the estimated pixel, the amount of change between the two pixels can be quantified through the divide operation and the value of the estimation background matrix matching the pixel area. When the amount of change has a value greater than or equal to a specific threshold, the value of a pixel matching the pixel area of the estimated background area is updated.

In the proposed algorithm, three variables exist: skip_overlay_frame, overlay_threshold, and change_threshold. Skip_overlay_frame is used to avoid excessive overlap. After the frame overlap process, the number of frames corresponding to skip_overlay_frame is skipped. Overlay_threshold is a number for how many frames will be used for background inference. If the values

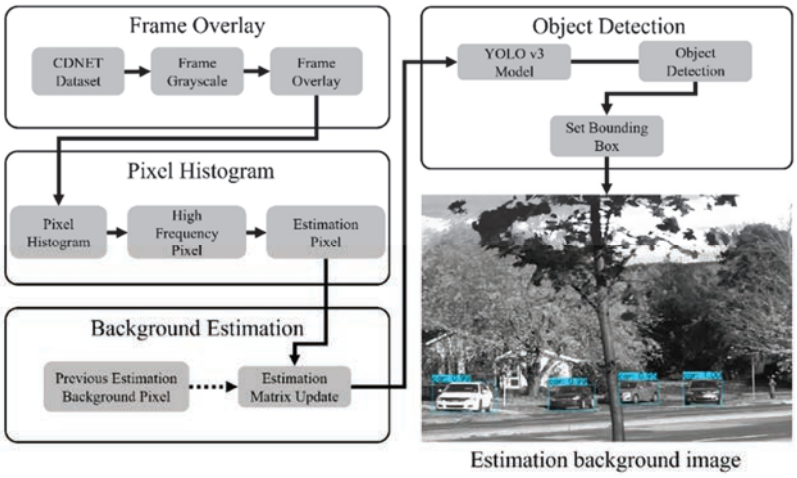


Figure 4 Overall operational flow.

of Skip_overlay_frame and Overlay_threshold are set too low, background information may be insufficient, which may adversely affect inference. The Change_threshold value causes small background changes to be ignored. Tolerance to noise can be achieved by ignoring small range changes in the histogram in the pixel area.

3.4 Background Object Detection

The YOLO v3 model is used for the object detection process Figure 4. shows the overall operational flow of the proposed method. This paper proceeds with the object detection process by using the basically trained YOLO v3 model and the obtained estimated background matrix.

4 Evaluation

In this section, we present experimental results to compare the proposed scheme with the simple YOLO method. The simple YOLO method performs the object recognition without background obtainment.

Table 1 shows the experimental environment. We adopt CDNET2014 as the dataset. For evaluation, an experiment is conducted on 80 classes trained in YOLO v3. For accurate comparison, image frames with interested objects passed by obstacles were selected. Images that are difficult to detect properly,

Table 1 Experiment environment

OS	RAM	CPU	GPU
Windows 10	48GB	Intel® Core™ i9-10900 2.8 GHz	NVIDIA GeForce GTX 1660 SUPER 6GB

Table 2 Experimental result of simple YOLO method

	Skating	Fall	Fountain01	Backdoor	Bungalows
Number of Frames	544	881	87	130	549
Totally undetectable	6	39	0	11	113
Some undetectable	0	637	52	105	387
Class misdetection	149	14	0	0	0
Duplicate detection	81	3	0	0	0

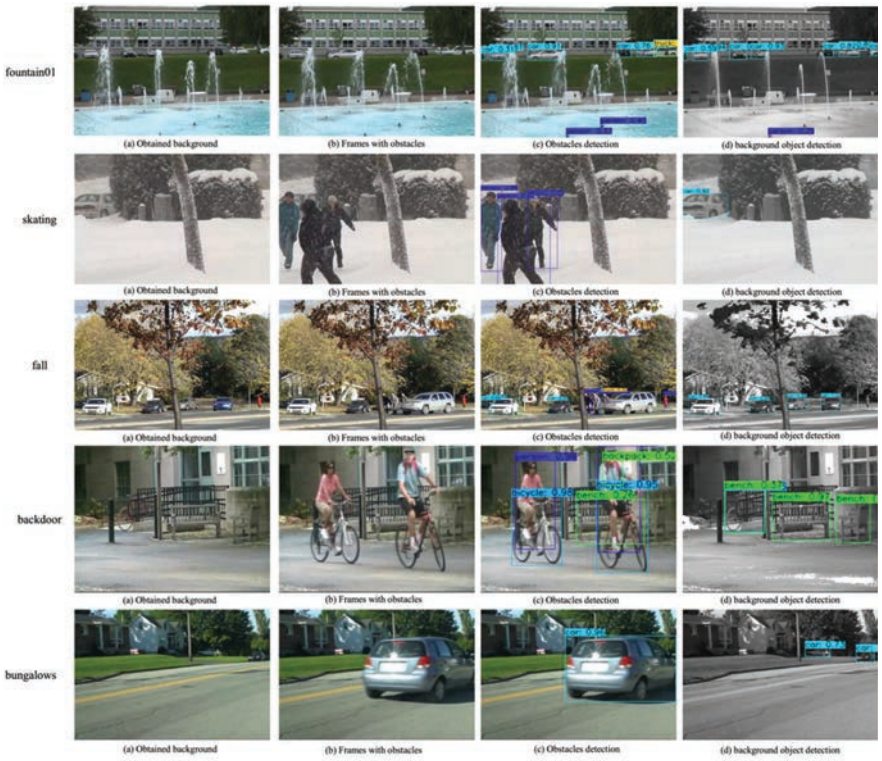
such as images with large camera shake and images with severe light blur, were also selected. The YOLO parameter, Input shape, is a grayscale image with a size of 416 by 416. The value of IoU threshold was 0.45 and the score threshold was 0.25. Some images in which the interested objects are classes trained in the YOLO v3 model exists in the background was selected. As a result, 5 videos of skating, fountaion01, fall, backdoor, and bungalows of CDNET2014 were used. Among them, 2191 comparison data frames were created as a dataset for YOLO v3 application. The result of detecting an object by selecting an arbitrary frame among the entire image is compared with the search result by applying the proposed algorithm to the same frame. The proposed algorithm repeats the process of inferring and updating the background while overlapping frame windows of a given size. YOLO is used as the object detection technique for each case. Table 2 shows the experimental results when an object is detected without background information to which the proposed method is not applied. And Table 3 shows the experimental results using the proposed method.

Figure 5 shows the examples of object detection after obtaining the background from the selected 5 videos. Figure 5(b) is an example showing a random frame selected from an image. (c) shows the result of the object detection process through YOLO in the selected frame. (d) is a background image inferred through the algorithm proposed in the selected frame and shows the object detection result in the inferred background.

In the graph of Figure 6, you can see the comparison results before and after applying the proposed method to each video. It indicates the number of

Table 3 Experimental result of the proposed scheme

	Skating	Fall	Fountain01	Backdoor	Bungalows
Number of Frames	544	881	87	130	549
Totally undetectable	0	0	0	0	0
Some undetectable	0	16	0	4	11
Class misdetection	0	0	0	0	0
Duplicate detection	0	0	0	0	0



(a) obtained background (b) frames with obstacles (c) obstacles detection (d) background object detection

Figure 5 Background Obtainment and Object Detection experiment.

frames for a case in which all objects to be detected in the background are not detected, some are not detected, an incorrect class is detected, and an object is detected separately. The Y-axis means the number of frames in which false detection and non-detection events occurred in the image. As a result of the

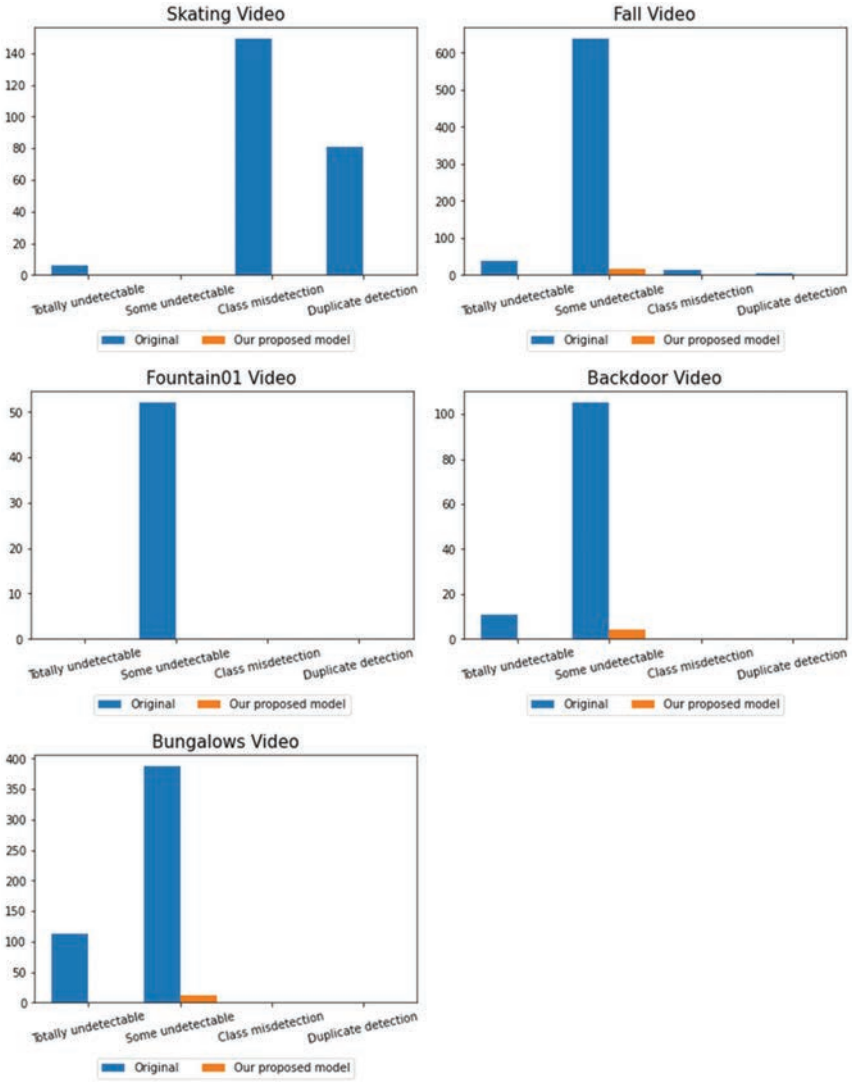


Figure 6 Comparison between simple YOLO method and the proposed scheme.

experiment, when there is an obstacle in the background object in the image, the precision of class classification is lowered with a high probability and the frequency of object detection failure is also increased. It can be seen that the accuracy increases when an image background is inferred, and a background object is detected using the background information of the inferred image.

Skating video had a problem in that it was classified into an incorrect class, or it was divided into several areas and counted into several. Other images often failed to be detected because they were covered by obstacles, and if you use the background information using the estimated background, you can see that it has been remarkably reduced. It was verified that the performance of object detection in the background can be improved if the background information estimated using the histogram-based background estimation algorithm is used. As a result of the experiment for 5 videos, the average object undetected rate improved by 60.2% and the false detection rate improved by 11.2%, respectively. Through this, when there is an object to be detected in the background, the proposed algorithm removes the object in the foreground, thereby showing a significant result in improving the detection performance.

5 Conclusion

Through this paper, we conducted a study for object detection in the image background with many obstacles, and how to estimate the background using the histogram-based background inference algorithm and detect the background object using the estimated background information. Suggested In the object detection in the image, it was confirmed that the detection failed with a high probability when the background object was covered without background information, and a high improvement could be seen by using the background information. However, when the camera's shooting position is not fixed or in an image in which light changes frequently with a short period, the background estimation performance deteriorates, and it may be difficult to obtain good results. Therefore, in order to utilize background information, precision of background estimation is essential, and additional research is needed.

Acknowledgement

"This research was supported by the MIST(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP(Institute of Information & communications Technology Planing & Evaluation) in 2022" (2019-0-01834).

References

[1] A. Darwich, P. A. Hébert, A. Bigand, Y. Mohanna, 'Background subtraction based on a new fuzzy mixture of Gaussians for moving object detection', Journal of Imaging, pp. 92, July, 2018.

[2] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, J. Walsh, 'Deep learning vs. traditional computer vision', In Science and information conference, pp. 128–144, April, 2019.

[3] A. Sofwan, M. S. Hariyanto, A. Hidayatno, E. Handoyo, M. Arfan, M. Somantri, 'Design of Smart Open Parking Using Background Subtraction in the IoT Architecture', In 2018 2nd International Conference on Electrical Engineering and Informatics (ICon EEI), pp. 7–11, October, 2018.

[4] C. Naimeng, Y. Wanjun, W. Xiaoyu, 'Smoke detection for early forest fire in aerial photography based on GMM background and wavelet energy', In 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), pp. 763–765, January, 2021.

[5] J. Han, C. Liu, Y. Liu, Z. Luo, X. Zhang, Q. Niu, 'Infrared small target detection utilizing the enhanced closest-mean background estimation', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 645–662, November, 2020.

[6] T. Mahalingam, M. Subramoniam, 'A robust single and multiple moving object detection, tracking and classification', Applied Computing and Informatics. July, 2020.

[7] T. Germer, T. Uelwer, S. Conrad, S. Harmeling, 'Fast multi-level foreground estimation', In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 1104–1111, January, 2021.

[8] Z. Zeng, J. Jia, D. Yu, Y. Chen, Z. Zhu, 'Pixel modeling using histograms based on fuzzy partitions for dynamic background subtraction', IEEE Transactions on Fuzzy Systems, pp. 584–593, June, 2017.

[9] H. Yang, S. Qu, 'Real-time vehicle detection and counting in complex traffic scenes using background subtraction model with low-rank decomposition', IET Intelligent Transport Systems, pp. 75–85, January, 2018.

[10] S. H. Jeevith, S. Lakshmikanth, 'Detection and tracking of moving object using modified background subtraction and Kalman filter', International Journal of Electrical and Computer Engineering, pp. 217–223, February, 2021.

[11] S. B. Song, J. H. Kim, 'SFMOG: Super Fast MOG Based Background Subtraction Algorithm', Journal of IKEEE, pp. 1415–1422, December, 2019.
[12] Y. Liu, G. Tang, W. Zou, 'Video monitoring of Landslide based on background subtraction with Gaussian mixture model algorithm', In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 8432–8435. July, 2021.
[13] M. Babaee, D. T. Dinh, G. Rigoll, 'A deep convolutional neural network for background subtraction', arXiv preprint arXiv:1702.01731. February, 2017.
[14] M. Vijayan, P. Raguraman, R. Mohan, 'A Fully Residual Convolutional Neural Network for Background Subtraction', Pattern Recognition Letters, pp. 63–69, June, 2021.
[15] T. Bouwmans, S. Javed, M. Sultana, S. K. Jung, 'Deep neural network concepts for background subtraction: A systematic review and comparative evaluation', Neural Networks, pp. 8–66. September, 2019.
[16] D. Zeng, M. Zhu, A. Kuijper, 'Combining background subtraction algorithms with convolutional neural network', Journal of Electronic Imaging, January, 2019.
[17] W. He, K. Yong, W. Kim, H. L. Ko, J. Wu, W. Li, B. Tu, 'Local compact binary count based nonparametric background modeling for foreground detection in dynamic scenes', IEEE Access, pp. 92329–92340, 2019.
[18] J. Redmon, A. Farhadi, 'Yolov3: An incremental improvement', arXiv preprint arXiv:1804.02767. April, 2018.

Biographies

Hyeong-Jin Kim received BS degree in computer science from Hoseo University in 2021. He is currently an MS of Computer Engineering at Hoseo University, Asan, Korea. His research interests include machine learning, Healthcare AI, and data science.

Min-Cheol Shin received BS degree in computer science from Kongju University in 2022. He is currently an MS of Computer Engineering at Hoseo University, Asan, Korea. His research interests include AI, Computer Vision, Deep Learning.

Man-Wook Han entered the Department of Computer Science at Hoseo University in 2018. He is currently a BA in Computer Science, Hoseo University, Asan. His research interests are AI, backend servers, and deep learning.

Chung-Pyo Hong received BS and MS degrees in computer science from Yonsei University, Seoul, Korea, in 2004 and 2006, respectively. In 2012, he received a PhD in computer science from Yonsei University, Seoul, Korea. He is currently an associate professor of Computer Engineering at Hoseo University, Asan, Korea. His research interests include machine learning, explainable AI, and data science.

Ho-Woong Lee completed a PhD in computer science from Dankook University, Gyenggi-do, Korea. From 2000 to 2020, he served as CTO of AhnLab, Inc. and adjunct professor at Seoul Women's University and Dankook University. He is currently an associate professor in the Department of Computer Science and Engineering, Hoseo University, Asan. His research interests are machine learning, computer security, digital healthcare and blockchain.