Fall Situation Recognition by Body Centerline Detection using Deep Learning

Dong-hyeon Kim[1], Dong-seok Lee[1], Soon-kak Kwon[1*]

Abstract

In this paper, a method of detecting the emergency situations such as body fall is proposed by using color images. We detect body areas and key parts of a body through a pre-learned Mask R-CNN in the images captured by a camera. Then we find the centerline of the body through the joint points of both shoulders and feet. Also, we calculate an angle to the center line and then calculate the amount of change in the angle per hour. If the angle change is more than a certain value, then it is decided as a suspected fall. Also, if the suspected fall state persists for more than a certain frame, then it is determined as a fall situation. Simulation results show that the proposed method can detect body fall situation accurately.

Key Words: Mask-RCNN, Body Fall, Color Video, Action Recognition.

I. INTRODUCTION

According to the National Statistical Office's estimate of future population, Korea has already become an aging society, with the ratio of people aged 65 or older exceeding 7.2% in 2000 and 14% in 2017. In addition, the elderly population is expected to enter the super-aged society by 2026, accounting for 20 percent of the population. The increase in the number of elderly people living alone due to the aging population become one of the biggest social problems. The elder living alone may face serious threat of survival because it is difficult to get help from other people. One of the threats for the elder living alone is the fall situation. A fall situation threatens for elderly to survival because it is difficult to quickly communicate this situation. Several methods are presented for detecting the fall situation. The fall detection is classified into methods based on equipment device and based on video [1-5].

In this paper, we propose the fall situation method by the deep learning. The object detection method through deep learning using a neural network significantly improves accuracy compared to the existing method [6-10]. Mask R-CNN [11], which is a state-of-the-art method for object detection, are applied to human body detection [12-15].

R-CNN is a type of deep learning neural networks. Also, Fast R-CNN keeps the bounding box information found in the selective search through the CNN and extracts and pools the corresponding region from the final CNN feature map to dramatically shorten the time of the CNN. Faster R-CNN can process even faster speed because it is processed in CNN when creating bounding box. Mask R-CNN adds a network that masks whether each pixel corresponds to an object in Faster R-CNN. In this paper, we use Mask R-CNN, which is the most advanced form for the performance and speed.

The methods of classifying objects by processing the bounding box have a disadvantage of reducing accuracy for a moving model [16]. However, Mask R-CNN increases accuracy because it verifies that each pixel is an object in addition to the processing of the bounding box.

First, the points of both shoulders and knees of the body are detected by Mask R-CNN. After that, a

centerline of the body is extracted from the points of shoulders and knees. The angle of the centerline is calculated and tracked. If the change of the angle exceeds a certain threshold, the current frame is considered suspected of falling. If the suspected fall condition in consecutive frames, the fall situation is detected.

II. CONVENTIONAL FALL DETECTION METHOD

Existing fall detection methods can be classified into equipment mounted detection methods and color image-based detection methods [1]. Fall detection methods based on an equipment-mounted device [2-3] are accurate when the user is equipped this device. However, the user should wear the kit continuously. Video based detection methods [4-5] are to analyze the color image to determine whether the situation is falling, which can detect the fall situation without any inconvenience of having to install special equipment. However, this method has the disadvantage of not being able to detect falls because the screen cannot be determined for situations with little lighting. To solve this problem, the method presented in this paper is a detection method of fall situation in color image through deep learning neural network.

III. METHOD OF FALL SITUATION DETECTION BY NEURAL NETWORK

In this paper, the method of detecting fall situation through deep learning neural network is implemented. The flow chart of the method proposed in this paper is shown in Fig.1.
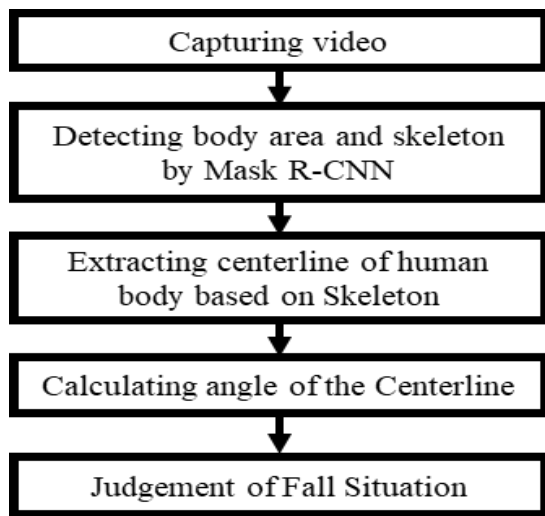
Fig. 1. Flowchart of the proposed method.

We Apply Mask R-CNN to the neural network for falls. Mask R-CNN is a neural network designed to cover areas of objects that are actually found. Mask R-CNN consists of ResNet for extracting feature maps, Feature Pyramid Network (FPN) for extracting classes and boxes from feature maps for efficient channel numbers, and Region Proposal Network (RPN) for which Mask predictions are added in RoI. Figure 2 shows the structure of Mask R-CNN for body detection and bouncing box skeleton extraction in this paper. In the FPN, the scale of the input image is reduced through the bottom-up layer, expanded through the top-down layer, and various sizes of objects can be detected through the bottom-up layer and top-down layer within the FPN. ResNet has applied a skip connection that adds the input value of the layer to the output value, and the learning efficiency increases as the size of the output value that must be learned for each layer decreases through the skip connection. The RPN detects the RoI of the object and the pixels to which the object belongs in the feature map. The RoI align is applied to the RPN of Mask R-CNN to improve the predicted accuracy of the bounding box and mask. The RoI Pool, applied to Fast R-CNN and Faster R-CNN, round up the decimal coordinates of the predicted bound box, while the RoI Align corrects the decimal coordinates through double linear interpolation, thus improving the accuracy of the RoIs prediction.
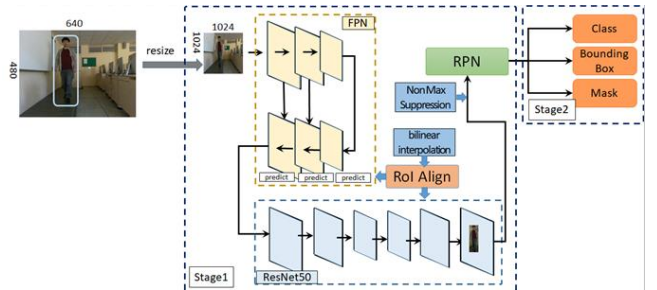
Fig. 2. Structure of Mask R-CNN.

We perform the learning of the body area and the main body part in Mask R-CNN. The main body parts learned are shoulder and knee points.

First, we install the camera at a point where you can take a picture of a person's body. At this point, the camera is installed parallel to the ground. In the captured image, the body area and the body's shoulder points $p_{sl} \equiv (x_{sl}, y_{sl})$, p_{sr} and knee points p_{nl} and p_{nr} are detected through Mask R-CNN. Then we locate the centerline of the body following the $p_{sc} \equiv ((x_{sl}+x_{sr})/2(y_{sl}+y_{sr})/2)$, and the p_{nc}. An angle of the centerline is then calculated as the following:

$$\theta = \tan^{-1}\left(\frac{y_{nc}-y_{sc}}{x_{nc}-x_{sc}}\right). \quad (1)$$
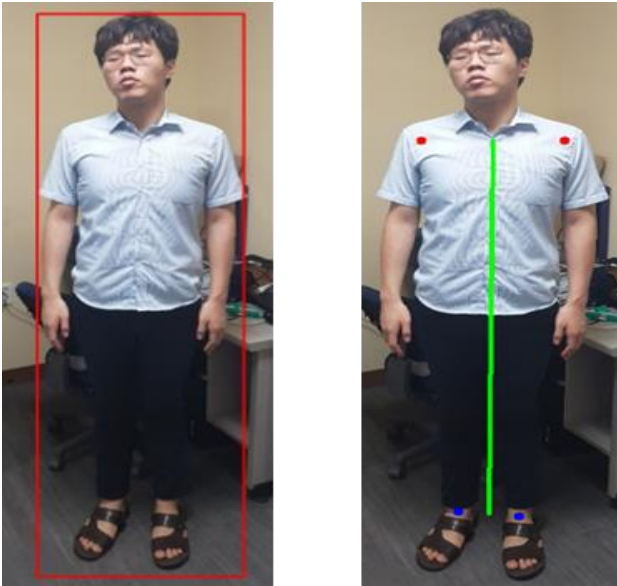
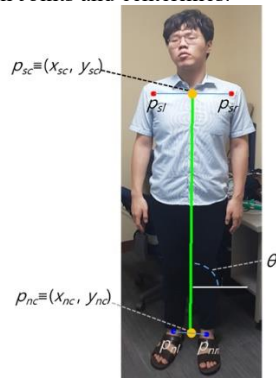Fig. 3. Tracking bounding box through Mask R-CNN and Detecting Skeleton Joints and centerlines.

Fig. 4. Calculate the center angle for determining fall situation.

Also, we track θ of the centerline in the image. If the body is collapsing, then θ will get smaller. If θ falls below a certain value, then it can be judged that a person has collapsed. However, even in normal situations, such as falling sharply and lying on one's stomach to sleep, the situation becomes smaller. Instead of using θ to distinguish these cases, we use θ's amount of change. In the case of a sudden fall, the change in θ per frame will be dramatically reduced. The variation of θ in the *n*-th frame is calculated as the following:

$$w(n) = \theta(n) - \theta(n-1). \qquad (2)$$

If w(*n*) is less than T_θ, then it is determined that the frame is suspected as body falling. A sudden change in body posture or a sudden misdetection of a major part of the body can cause a sudden change in the amount of θ even if it is not in a fall situation. For this purpose, if the suspected fall condition persists for the duration of *N* frames, it shall be detected as a fall situation.

IV. SIMULATION RESULTS

The R50-FPN model is applied as a backbone of Mask R-CNN. COCO Dataset [17] is used for training set. 200,000 images are trained. 250,000 persons are included in the training images. The average precisions are 55.4 for detecting the boundary box and 65.5 for detecting the key points, respectively.

To measure the accuracy of the fall situation detection implemented in this paper, the experiment is conducted through images containing five scenarios shown in

(a)   (b)   (c)
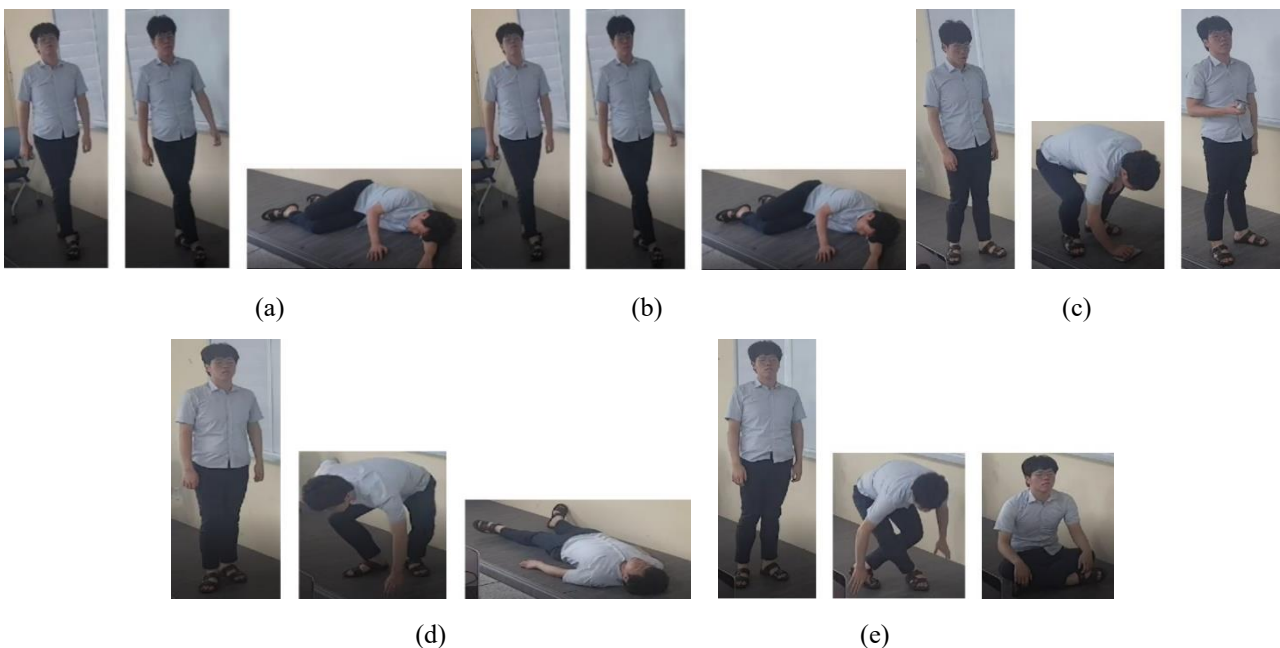

(d)   (e)

Fig. 5. Fall detection for experimental image: (a) falling while walking, (b) standing fall, (c) head down, (d) lying down, and (e) sit down.
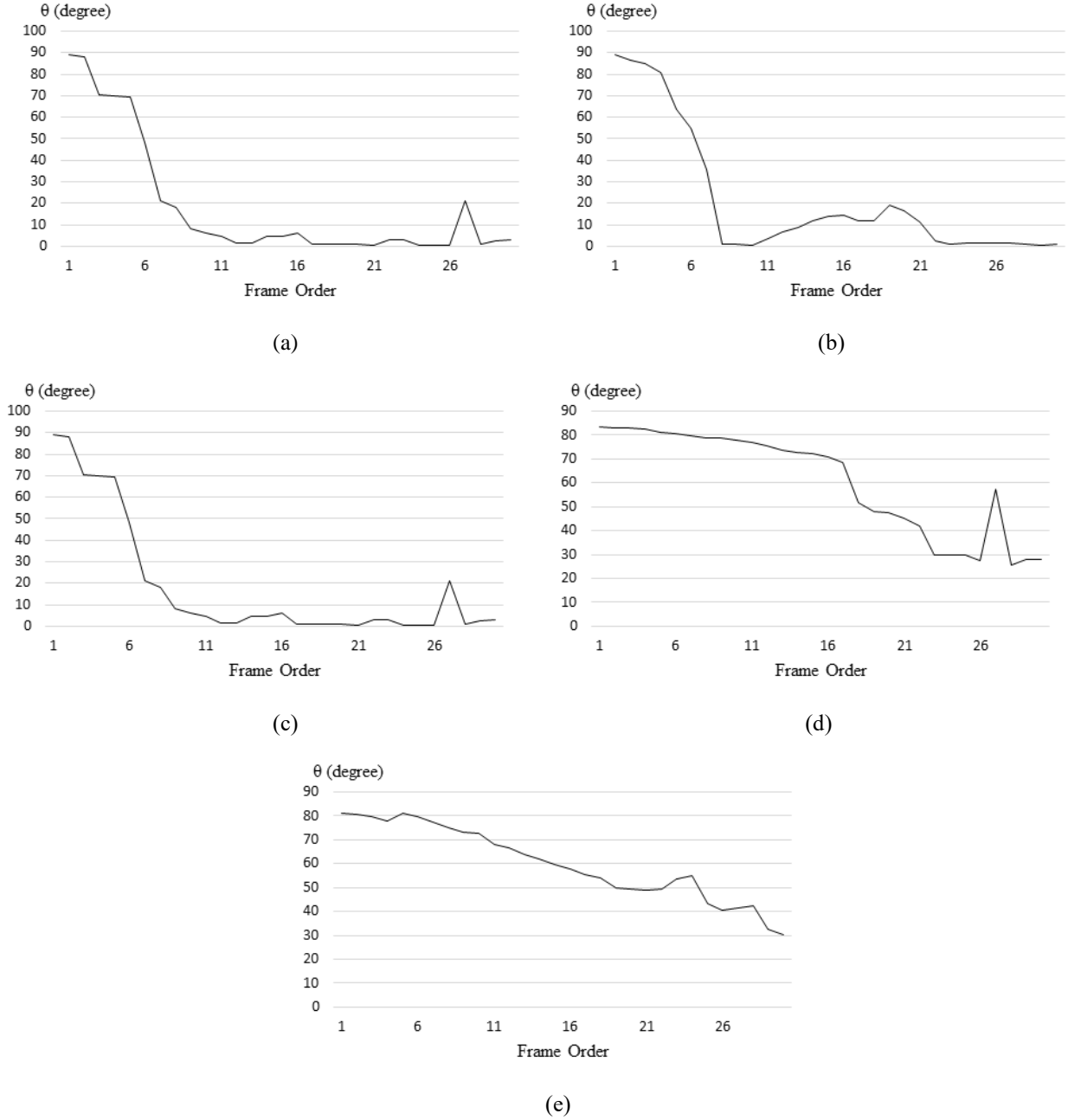
Fig. 6. Changes of θ: (a) falling while walking, (b) standing fall, (c) head down, (d) lying down, and (e) sit down.

Fig. 5. At this point, the resolution of the images is 852x480. Fig. 5 (a)-(b) must be detected in the fall situation, and the remaining images must not be detected in the fall situation. In the experiment, T_θ and N are set at 76 and 6, respectively.

Fig. 6 shows the change in θ for each experimental image. For the images of Figure 5 (a)-(b) which are falling, the variation of θ in the fall situation is continuously significant. Conversely, in Figure 5 (c)-(e), which is not a fall situation, the change in θ is not significant, and is temporarily limited, if large.

Table 1 shows the number of frames identified by the suspected fall situation in accordance with T_θ. At this point, the larger the T_θ, the smaller the number of frames

mis-detected to the suspected fall situation in the scenario of Fig. 5 (c)-(e).

Table 2 shows whether a fall situation is detected in accordance with N, the basis for the fall situation. If N is less than 3, it may be detected as a fall even if it is not a fall situation. On the other hand, when N is greater than 7, no normal fall situation was detected. This situation means that if a fall situation occurs in a moment and N is too large, it is rather impossible to determine the fall situation accurately.

Through the results, the proposed method accurately calculates the angle of the centerline for detecting falls, and the accuracy of the fall detection is extremely high. However, in the case of a person lying down (d), an

260

inaccurate result occurs. In the future research direction, additional conditions should be studied, not just the amount of angular change in the centerline, in situations in which a person lies down.

Table 1. Number of suspected fall state frame detections for T_θ.

Images	T_θ					
	5	10	15	20	25	30
Falling while walking	15	6	6	5	2	2
Standing fall	16	7	5	5	3	2
Head down	0	0	0	0	0	0
Lying down	13	7	3	3	2	1
Sit down	10	4	4	3	2	2

Table 2. Fall detection for *N*.

Images	N					
	3	4	5	6	7	8
Falling while walking	O	O	O	O	X	X
Standing fall	O	O	O	O	O	X
Head down	X	X	X	X	X	X
Lying down	O	X	X	X	X	X
Sit down	X	X	X	X	X	X

V. CONCLUSION

In this paper, the method of detecting the area and major parts of the body accurately using Mask R-CNN and detecting the fall situation by detecting the center line was implemented. Through the proposed method, the fall situation could be detected more accurately through CCTVs installed indoors, and it would be possible to cope with an emergency. This will enable the detection of emergency situations of the elderly accurately, which will prevent human and property damage by informing them of the danger situation more quickly and acting against them.

REFERENCES

[1] N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. O. Laighin, V. Rialle, and J. E. Lundy, "Fall detection - Principles and Methods," in *Proceeding of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1663-1666, 2007.

[2] A. Diaz, M. Prado, L. M. Roa, J. Reina-Tosina, and G. Sanchez, "Preliminary evaluation of a full-time falling monitor for the elderly," in *Proceeding of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2180-2183, 2004.

[3] G. Wu, "Distinguishing fall activities from normal activities by velocity characteristics," *Journal of Biomechanics*, vol. 33, no. 11, pp. 1497-1500, 2000.

[4] K. de Miguel, A. Brunete, M. Hernando, and E. Gambao, "Home Camera-Based Fall Detection System for the Elderly," *Sensors*, vol. 17, no. 12, pp. 1-21, 2017.

[5] E. E. Geertsema, G. H. Visser, and M. A. Viergever, "Automated remote fall detection using impact features from video and audio," *Journal of Biomechanics*, vol. 88, no. 9, pp. 25-32, 2019.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *Proceeding of the Conference on Computer Vision and Pattern Recognition*, pp. 580-587, 2014.

[7] R. Girshick, "Fast R-CNN," in *Proceeding of the International Conference on Computer Vision*, Santiago, Chile, pp. 1440-1448, Dec. 2015.

[8] R. Shaoqing, H. Kaiming, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *IEEE Transection on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2016.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-time Object Detection," in *Proceeding of the Conference on Computer Vision and Pattern Recognition*, pp. 779-788, Jun. 2016.

[10] S. K. Kwon and D. S. Lee, "Zoom motion estimation for color and depth videos using depth information," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 11, pp. 1-13, 2020.

[11] K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN," *IEEE Transection on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, 2020.

[12] R. Saini, P. Kumar, B. Kaur, P. P. Roy, D. P. Dogra and K. C. Santosh, "Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2529-2540, 2018.

[13] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P. P. Roy, and D. P. Dogra, "Multimodal Gait Recognition

with Inertial Sensor Data and Video Using Evolutionary Algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 1-10, 2018.

[14] O. Mazumder, S. Tripathy, S. Roy, K. Chakravarty, D. Chatterjee, and A. Sinha, "Postural sway based geriatric fall risk assessment using Kinect," in *Proceeding of 2017 IEEE Sensors*, pp. 1-3, Nov. 2017.

[15] S. Roy and T. Chattopadhyay, "View-Invariant Human Detection from RGB-D Data of Kinect Using Continuous Hidden Markov Model," in *Proceeding of International Conference on Human-Computer Interaction,* pp. 325-336, Jun. 2014.

[16] K. M. Sudeep, V. Amarnath, A. R. Pamaar, K. De, R. Saini and P. P. Roy, "Tracking Players in Broadcast Sports," *Journal of Multimedia Information System*, vol. 5, no. 4, pp. 257-264, 2018.

[17] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Proceeding of the European Conference on Computer Vision*, pp. 740-755, Sep. 2014.

Authors

Dong-hyeon Kim is currently an undergraduate student in the Department of Computer Software Engineering at Dong-eui University. His research interest is in the areas of image recognition.

Dong-seok Lee received the B.S. and M.S. degrees in Computer Software Engineering from Dong-eui University in 2015 and 2017, respectively, and is currently a Ph.D. course in the Department of Computer Software Engineering at Dong-eui University. His research interest is in the areas of image processing and video processing.

Soon-kak Kwon received the B.S. degree in Electronic Engineering from Kyungpook National University, in 1990, the M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1992 and 1998, respectively. From 1998 to 2000, he was a team manager at Technology Appraisal Center of Korea Technology Guarantee Fund. Since 2001, he has been a faculty member of Dong-eui University, where he is now a professor in the Department of Computer Software Engineering. From 2003 to 2004, he was a visiting professor of the Department of Electrical Engineering in the University of Texas at Arlington. From 2010 to 2011, he was an international visiting research associate in the School of Engineering and Advanced Technology in Massey University. Prof. Kwon received the awards, Leading Engineers of the World 2008 and Foremost Engineers of the World 2008, from IBC, and best papers from Korea Multimedia Society, respectively. His biographical profile has been included in the 2008~2014, 2017~2020 Editions of Marquis Who's Who in the World and the 2009/2010 Edition of IBC Outstanding 2000 Intellectuals of the 21st Century. His research interests are in the areas of image processing, video processing, and video transmission.