Real-Time Human Detection in a Restricted Area for Safety in Truck Dumper Control System Using Deep Learning

Apirak Worrakantapon, Wattana Pongsena, Kittisak Kerdprasop, and Nittaya Kerdprasop
Data and Knowledge Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand
Email: apirak.wor@gmail.com; pongsena@hotmail.com; {kerdpras, nittaya}@sut.ac.th

***Abstract*—A process to receive raw materials from suppliers in an animal feed industry utilizes both automatic and semi-automatic machine control systems. The process called "truck dumper system" is the procedure that the suppliers provide raw materials carried by trucks; then, their tailgates open, and the raw materials are discharged by raising front end part of a truck to gather raw materials in a collection area. In general, the truck dumper system has been controlled manually by staff in a control room, not by a truck driver. However, serious accidents may occur during the process because when the dumper lifts up, the staff's vision has been blocked by the raised part of a truck. Therefore, if the staff controls the dumper to lift down by lacking safety awareness, people in the restricted area can be endangered. In this study, we proposed a framework of automatic human detection to prevent any accident that may occur from the truck dumper in the restricted area. The human detection model was developed to detect humans possibly in different blind corners that are difficult for staff in a control room to monitor these unseen areas for safety-awareness. The main technology of the proposed framework was the real-time human detection with fully convolutional neural network architecture called You Only Look Once, or YOLO. The framework has been designed to send a signal to terminate the truck dumper system immediately after the model detects people in the restricted area. In experiments, we discovered that the model could detect a human in all blind corners, including the corners that the staff's sight was completely bloacked by some barriers. The overall efficiency of this framework in an aspect of speed was high. The average time to process per image was 397 milliseconds by using CPUs and only 52 milliseconds by using GPUs. The results also showed that the model was effectively applicable to detect human in real-time due to its high-speed process.**

***Index Terms*—Truck dumper, human detection, safe-dumping system, deep learning, YOLO, convolutional neural network**

I. Introduction

In the present era of industry 4.0, more and more industrial factories have moved toward the smart and automation systems that in the previous decade have usually been operated by humans. However, which industrial procedures being able to be automated can be varied from industry to industry due to the differences in the systems. The focus of this work is automation process in the animal feed industry. Traditionally, a feedmill, which is a factory to produce animal feed, includes many processes: receiving raw material, storing and managing the material, mixing and pelleting, packing, and loading out to silo trucks. Indeed, the mentioned processes involve many heavy machines. Therefore, safety in the work place for employees and other people nearby is very important because accidents from heavy machinery can be easily occurred.

In the process of raw-material receiving, general factories obtain the raw material from suppliers conveyed by big dump trucks. The trucks dump the material to raw-material intake; then, the machine transfers the raw material to storage such as silos or bulk storages. Supplier's dump trucks can be distinguished into two main types: a truck with a self-dumping dump body and a truck fixed the dump body part. For the truck fixed the dump body part, the factories use truck dumper systems to lift the whole truck and dump the material into a designated area. During the process, this are is strictly prohibited by not allowing people to get close because it may cause fatal danger. To prevent any unpleasant accidental situation to occur, a staff in a control room is assigned a particular job to observe the workspace area for making sure that there is no human in the area. However, blind-side corners in the area are difficult to notice by the staff who monitors in a remote room.

We therefore propose an automation system to help securing the controlled access zone in the truck dumping system. As far as we know, this framework is the first proposal to turn a human work toward the automation process. The main benefit of our proposed framework is for human safety in a truck dumping system, especially in the animal feed factory. In the next section, preliminaries regarding human detection with convolutional neural network and You Only Look Once (YOLO) are briefly presented. Our proposed automated safety framework for human detection in the truck dumping system is explained in Section III. Section IV describes experimental details and evaluation metric. Experimental

Manuscript received January 2, 2020 revised March 5, 2020; accepted April 10, 2020.

Corresponding author: Apirak Worrakantapon (email: apirak.wor@gmail.com).

doi: 10.18178/ijeetc.10.1.29-35

results are illustrated in Section V. We finally conclude our work in Section VI.

II. Preliminaries

In the past, computer vision based on various statistical and mathematical methods has been a popular technique for detecting humans in still images and video stream and for recognizing action [1]-[5]. However, rccognition accuracy is sometimes unsatisfiable because light and sight aspects affect the performance of the computer vision techniques. Common solution for such problems are applying the background subtraction technique, but it is inapplicable to various budging objects. Another solution is object capturing with a 3D camera that has been proven satisfiable detected results; however, the 3D camera is costly [6], [7].

With the high performances of the central processing unit (CPU) and graphics processing unit (GPU), the complex deep learning algorithms such as convolutional neural networks (CNN) have been extensively applied to solve the human detection problem [8]-[11], [12]. The success of CNN application for object detection is because of the multi-layer network-based architecture that makes it excellent in extracting representative features from images [13], [14]. Its architecture contains two parts: a convolution part and a classification part. A convolution part is for feature extraction, whilst the classification part is for object recognition. However, CNN is still hard to use in real-time applications because of its huge time-consuming during the learning phase.

To improve time complexity during learning phase of a CNN deep learning method, a faster image recognition scheme called You Only Look Once, or YOLO, has been proposed by Redmon and colleagues in 2015 [15]. As the name suggested, YOLO speeds up typical CNN learning style by performing a one-stage scan over image and transforming an image classification problem to be a logistic regressional learning. Presently, YOLO has been adopted to solve a wide rage of machine vision and engineering problems that need object detection in real-time [16]-[22].

YOLO works faster than ordinary CNNs by locating objects in the image and classifying the objects at the same time. It firstly divides an image into SxS grids. Then, it predicts positions and size of interest objects and estimates their confidence scores for possibly object types on each grid simultaneously [15], [20]. This concept is shown in Fig. 1. The high confidence score represents the high probability of the target. This concept helps reducing time consumption potentially.

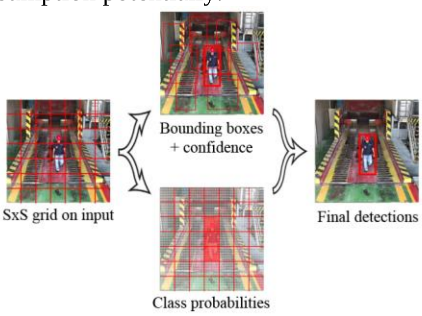


Fig. 1. The overall object detection concept of YOLO

In this study, we proposed a framework to adopt YOLO for real-time human detection in the restricted area for the safety of truck dumper control systems. The main focus of our framework is for preventing accidents that may harm people involved in truck dumper systems.

III. A Safe-Dumping System Framework

In this study, we proposed a framework (as shown in Fig 2) for preventing accidents during the raw-material receiving process managed by truck dumper control systems. This framework operates in a real-time scenario. Firstly, the system captures images from the IP Camera (step 1) and processes them using the human detection system implemented with YOLO (step 2). It counts the number of people appearing in the risky areas of the image. Then, the counting results are forwarded to the Object Linking and Embedding for Process Control Server, or OPC Server, system (step 3).

After that, the number of detected people were recorded in the Programmable Logic Controller (PLC) memory (step 4). We add additional conditions to a ladder logic in the PLC program to operate the truck dumper control system. The ladder logic is a program written in a graphical diagram that specifies particular conditions for signaling to the machine. Details of steps 1 throuh 4 are explained in the sub-sections: A, B, and C.

In the proposed framework, if a human detection system in step 2 can detect people in the restricted areas, that is, counting result is greater than zero. A signal should be submitted to the machine, and all machine operations must stop. If the specified memory in the PLC is equal to 0, the machine operation should resume to normal situations.

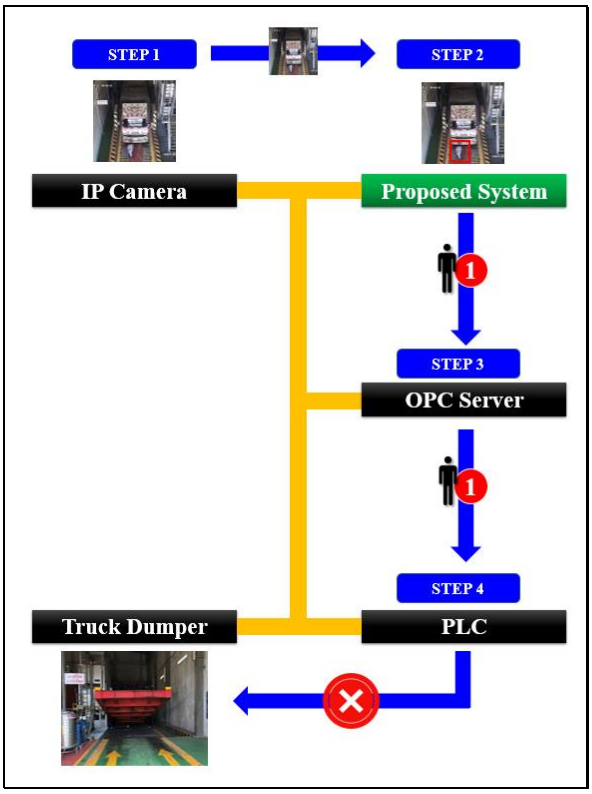
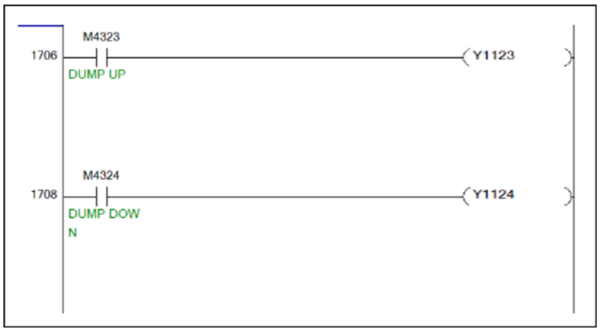
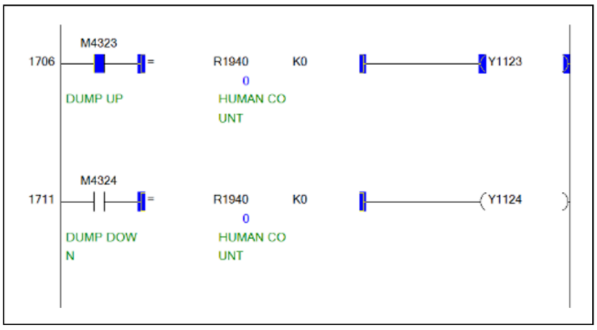


Fig. 2. A proposed framework for human safety in the truck dumping system in a feedmill factory



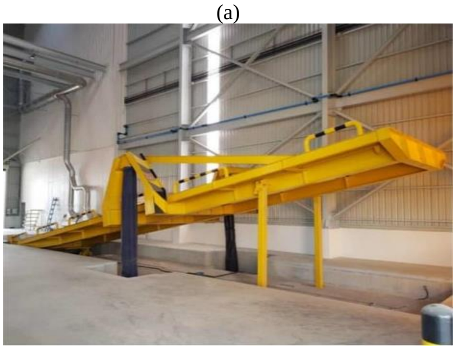
(a)



(b)

Fig. 3. PLC program: (a) Normal PLC program: interface of a program before adding conditions to a ladder logic and (b) a PLC program with safety conditions: a modified version according to the proposed framework that safety conditions have been added to a logic

(a)



(b)

Fig. 4. Illustration of the truck dumper: (a) machine in the idle stage, and (b) machine in the working stage (Photos courtesy: Kasetphand Industry Co., Ltd., Thailand)

Fig. 3 (a) shows the original PLC program that was implemented in the truck dumper system to control raw-material dumping in the animal feedmill factory. The program control two operations: dump up and dump down. The modification of this original PLC program to provide a safer environment for employees is the one shown in Fig. 3 (b). The number of human count as a result of human detection system has been added to the program. The new program controls truck dumping like the original version if the number of human count in the restricted area is zero. For some unusal situations that a human appears in the prohibited zone, the truck dumping operations will be blocked. Fig. 4 illustrates truck dumper system at work. In Fig. 4 (a), the dumper system is in the idle stage, whereas it is in the working stage (dump-up) in Fig. 4 (b).

A. Internet Protocol Camera (IP Camera) (Fig. 2 Step 1)

IP Camera is a type of digital video camera that can directly connect to a network system. It is usually used for surveillance camera around houses, organizations, or factories for security monitoring. This camera can be set as a web server; thus, all devices located in the same network can access the camera in real-time.

B. Proposed Safety System: Human Detection with Deep Learning (Fig. 2 Step 2)

The human detection system had been implemented by Python programming. Our system has been designed to access the camera streaming in real-time. Images are then imported to our YOLO model in order to detect human in the image. Comparing to other deep learning models such as CNN and Faster Region-based Convolutional Neural Networks (Faster R-CNN) [23], the YOLO model works faster with acceptable recognition accuracy.

C. OPC Server (Fig. 2 Step 3) and PLC (Fig. 2 Step 4)

OPC Server is a software interface standard used for communicating among devices in the network of the control system. It communicates with other devices through the Human Machine Interface (HMI), which is a graphical screen allowing persons to control the system comfortably.

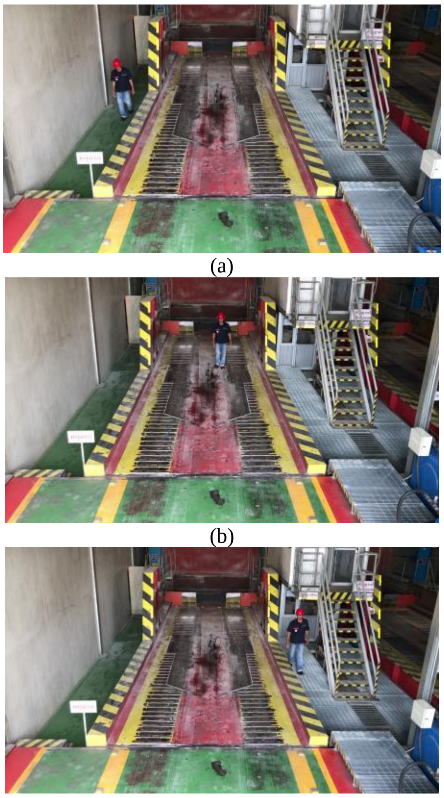
PLC is a small modular device with multiple inputs and outputs (I/O) used for controling the machine. A traditional controler controls the machine via electrical circuit wiring which is difficult to modify and costly to maintain the circuit afterward. A better solution is to use PLC for machine control. Programmers or technicians can write a control program into a PLC memory and can easily modify the program subsequently. The PLC has its own CPU and input signal receivers including sensors for program processing. Also, it can generate output signals for activating the machine. In this study, we add a program into PLC to identify the conditions for filtering output signals to the machine.

HMI or OPC Client can send and receive both data and commands to OPC Server. Then, the server passess data and command to PLC memory. After that, the data and command in the memory are processed into the designed ladder logic. Finally, the machine address mapped by ladder logic should be activated. In this study, we implement a program as an OPC Client. It sends the number of detected people from the human detection system to OPC Server and records the number into the PLC memory for the program processing.

IV. Experiment and Evaluation

A. Data Preparation

The input data of this framework are a collection of images in the real situation. These images are extracted from video clips captured by the IP Camera installed at the animal feedmill factory in Thailand. The data set of this work contains 300 images with only one perspective camera but the image data contain three different human positions in the images, i.e., a human at left-side of the image (Fig. 5 (a)), a human at the center of the image (Fig. 5 (b)), and a human at right-side of the image (Fig. 5 (c)).



(a) (b) (c)

Fig. 5. Example of raw input data in a format of color images with three different human positions at the dumper station: (a) at the left-side, (b) at the center, and (c) at the right-side.

The image with a human at the left side of the dumper station represents a situation that there is a human in the prohibited area which is the opposite side of the control room. The human at the center of the image represents scenario that there is a human standing at the top of the truck dumper. The human at the right side of the image represents the event that a human appears at the side of the control room. We collect 100 images of each of the three scenarios. The image resolution was 1928×1080 pixels.

B. Tools used in the Experiment

The hardware and software tools used for training the deep learning model and for testing the recognition accuracy of the human detection system are listed as follws.

- Hardware
  - OS: Windows 10 Pro
  - CPU: Inter® Core™ i7-9750H 2.60GHz
  - RAM: 16.0 GB
  - GPU: NVIDIA Geforce RTX 2070
- Software
  - NVIDIA CUDA 10.1 and cuDNN 7.6

C. Evaluation

In this research, the performance of the deep model for detecting human in the restricted area was evaluated using precision as a measurement metric for assessing performance. The precision can be calculated as the proportion between the number of results that the human detection system predicts accurately and the total number of the test data. The calculation is shown in equation (1).

$$\text{Precision} = \frac{\text{True positive}}{\text{Actual results}} \tag{1}$$

where True positive is the number of positive results that the model predicts accurately and Actual results are the total number of the test data.

V. Experimental Results

Our input data was collected from the fixed-perspective IP Camera with three different human positions (as shown some examples in Fig. 5). After inquiring about possible unsafe situations from the factory staff, we found that only two human positions had been blocked from the staff's vision: a human at the left-side of the image and a human at the center of the image. These may lead to serious accidents if any people exist in the restricted area while the truck dumper system is activated. Fortunately, the model based on deep learning can detect people who are in the restricted area from all positions.

Results of human detection at all three positions are summarized and presented in Table I. We implement both YOLO versions 2 and 3, and then compare human detection performance against a human staff working in a control room. Results from YOLOv3 human detection for both detectable and undetectable cases are demonstrated in Fig. 6.

Table I. Results Showing Effectiveness of the Model for Human Detection in the Restricted Areas

Experimentation		Human Staff	Model-based YOLOv2	Model-based YOLOv3
left-side	Precision	0.00	0.85	1.00
	Processing Time (milliseconds)	-	203	399
center	Precision	0.00	0.72	1.00
	Processing Time (milliseconds)	-	209	396
right-side	Precision	1.00	0.60	0.91
	Processing Time (milliseconds)	-	195	398

(a)

(b)

Fig. 6. Performance of a human detection model: (a) detectable a human image at the left-side, center and right-side, and (b) undetectable a human image at the left-side, center and right-side.

In terms of processing time, the experimental results in Table I show that the model based on YOLOv2 can recognize human faster than the model based on YOLOv3, approximately two times faster. In the aspect of human safety, the speed of the model is highly important because it obviously helps reducing any accident that may occur when people enter the restricted area. Once the human is detected, the control system can automatically pause the dumper machine immediately.

In terms of accuracy, the results demonstrate that the model based on YOLOv3 provides significantly higher accuracy than the model based on YOLOv2. The model based on YOLOv2 provides the average precision for automatic detection in three different human positions as 72%, whereas the average accuracy of YOLOv3 is as high as 97%. The precision of the YOLOv2 model to detect the human at the left-side, center, and right side of the image was 85%, 72%, and 60%, respectively. YOLOv3 can detect human at the same positions with accuracy rate 100%, 100%, and 91%, respectively. It can be observed from the result that the model based on YOLOv3 provides the average precision significantly higher than the model based on YOLOv2 up to 25%.

The main objective of this study was to develop the framework to prevent accidents in the industrial factory that might occur in the restricted area. If the accidents in the restricted area really happen in the truck dumper area, it will risk a human's life. Thus, we select the model based on YOLOv3 due to its high precision and acceptable speed. To compare the results obtained by staff monitoring, our model can detect the human at the blind-side of the image in which the staff cannot notice human in that area. Therefore, we are confident to apply this framework in a real workplace of truck dumper.

VI. Conclusion

This work presents the design of a framework for detecting human unexpected to appear in the prohibited zone of truck dumping in the feedmill factory. The main purpose of this design is for improving safety in the factory. The automatic process of the proposed framework is anticipated to provide a better safety system than the current human staff controling system. The experimental results of this study show that the proposed framework has a high potential to prevent serious accidents for people working in the restricted area.

In the proposed framework, we adopt a deep learning model to detect the number of people in all three possible positions. The model can detect human in restricted area with precision as high as 97%. The precision of the proposed model can possibly be further improved by increasing the number of images in the training dataset. However, the current model is effectively applicable even when the human appears in blind spots that invisible by the staff in a control room. The processing time of the model is quite low with an average time per image of 397 milliseconds. Moreover, it can process images in real-time with the GPU that a rate of processing speed was up to 19.2 images per second. This indicated that the proposed framework with the deep learning model can be used in real-world situations.

Based on the success of this preliminary design and implementation, the proposed idea can be deployed not only in the truck dumper control system of the feedmill factory, but the idea can also be applied to many production and manufacturing sectors. Since the main concern of the proposed framework is for improving safety for workers in the factory, any kind of application that seeks for safety improvement is thus able to adopt and adjust the proposed fraework to suit a specific application area. For instance, the safety control system in loading/unloading goods in large warehouses is one area of application.

In future work, we plan to improve the performance of the human detection model that can work under different environments such as different illumination conditions. Such conditions can normally occur in a workplace of our truck dumper control system, for example, the blur scene during the cloudy days. Therefore, some preprocessing steps to handle sub-optimal illumination should be helpful for the increase in accuracy of our system. We also plan to include a classifier to distinguish between outsiders and employees. Even though this aspect has no direct effect toward the performance of our system, such extension is more or less improving the intelligent level of our system. Then, we plan to combine this framework of human-detection and a notification system to notify security guards when people invade the prohibited areas, for instance, a confined space and under-construction area, to reduce the probability of accidents. This kind of extension is necassary for a practical application aspect of the proposed framework to aid other kinds of workplaces such as hardware manufacturing, construction sites, and many dangerous industries.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Apirak Worrakantapon conducted the research and experiments, analyzed the data as well as proposed the method of the study; Apirak Worrakantapon and Wattana Pongsena wrote the paper; Kittisak Kerdprasop and Nittaya Kerdprasop proved the correctness of the results and had approved the final version of this manuscript

References

[1] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, vol. 2, pp. 1491–1498.

[2] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proc. European Conference on Computer Vision*, 2006, pp. 428–441.

[3] C. Raghavachari, V. Aparna, S. Chithira, and V. Balasubramanian, "A comparative study of vision based human detection techniques in people counting applications," *Procedia Computer Science*, vol. 58, pp. 461-469, 2015.

[4] H. Zhang, Y. Zhang, B. Zhong, Q. Lei, L. Yang, J. Du, and D. Chen, "A comprehensive survey of vision-based human action recognition methods," *Sensors*, vol. 19, 2019.

[5] J. K. Agrawal and M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Survey*, vol. 43, nn. 3, 2011.

[6] X. Yang and Y. Tian, "Effective 3D action recognition using EigenJoints," *Journal of Visual Coomunication and Image Reoresentation*, vol. 25, no. 1, pp. 2-11, 2014.

[7] D. I. B. Hagebeuker and P. Marketing, "A 3D time of flight camera for object detection," *PMD Technol. GmbH, Siegen*, 2007.

[8] J. Zhao, G. Zhang, L. Tian, and Y. Q. Chen, "Real-time human detection with depth camera via a physical radius-depth detector and a CNN descriptor," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 1536–1541.

[9] J. H. Kim, H. G. Hong, and K. R. Park, "Convolutional neural network-based human detection in nighttime images using visible light camera sensors," *Sensors*, vol. 17, no. 5, 2017.

[10] T. Liu and T. Stathaki, "Faster R-CNN for robust pedestrian detection using semantic segmentation network," *Fromtiers in Neurorobotics*, vol. 12, 2018.

[11] N. AlDahoul, A. Q. M. Sabri, and A. M. Mansoor, "Real-time human detection for aerial captured video sequence via deep models," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[12] X. Wang and S. Hosseinyalamdary, "Human detection based on a sequence of thermal images using deep learning," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W13, pp. 127-132, 2019.

[13] D. Strigl, K. Kofler, and S. Podlipnig, "Performance and scalability of GPU-based convolutional neural networks," in *Proc. 18th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 2010, pp. 317–324.

[14] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 650–657.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[16] R. Laroca, *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–10.

[17] Q. Peng, *et al.*, "Pedestrian detection for transformer substation based on gaussian mixture model and YOLO," in *Proc. 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2016, vol. 2, pp. 562–565.

[18] D. Shen, X. Chen, M. Nguyen, and W. Q. Yan, "Flame detection using deep learning," in *Proc. 4th International Conference on Control, Automation and Robotics (ICCAR)*, 2018, pp. 416–420.

[19] C. Tang, Y. Ling, X. Yang, W. Jin, and C. Zheng, "Multi-view object detection based on deep learning," *Applied Sciences*, vol. 8, no. 9, 2018.

[20] J. Li, J. Gu, Z. Huang, and J. Wen, "Application research of improved YOLO v3 algorithm in PCB electronic component detection," *Applied Sciences*, vol. 9, 2019.

[21] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417-426, 2019.

[22] X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, and L. Kurach, "A deep learning-based framework for an automated defect detection system for sewer pipes," *Automation in Construction*, vol. 109, 2020.

[23] C. Han, G. Gao, and Y. Zhang, "Real-time small traffic sign detection with revised faster-RCNN," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13263-13278, 2019.

Apirak Worrakantapon is a master student, School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. He received his B.E. in computer engineering from SUT, in 2008. He has long experiences with automation system in the feedmill factory. His current research of interest includes software engineering, automation system, artificial intelligence, machine learning and deep learning.

Watthana Pongsena is a Ph.D.student at School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. He received his B.E. and M.E. in computer engineering from SUT in 2008 and 2012, respectively. His research of interest includes software engineering, data mining, artificial intelligence, and human-computer interaction.

Kittisak Kerdprasop is an Associate Professor and Chair, School of Computer Engineering, SUT. He received his bachelor's degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes machine learning and artificial intelligence.

Nittaya Kerdprasop is an associate professor and the head of Data and Knowledge Engineering Research Unit, School of Computer Engineering, SUT. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her research of interest includes data mining, artificial intelligence, logic, and constraint programming.