

Human Detection with Uprisen Angle of a Camera for the Service Robot

Watcharin Tangsuksant¹, Amornphun Phunopas², Pornthep Sarakon³, Aran Blattler⁴

Department of Production and Robotics Engineering,

Center of Innovative Robotics and Advanced Precision Systems: iRAPs,

Faculty of Engineering, King Mongkut's University of Technology North Bangkok,

1581 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok, 10800, THAILAND

watcharin.t@eng.kmutnb.ac.th¹, amornphun.p@eng.kmutnb.ac.th², pornthep.s@eng.kmutnb.ac.th³, aran.b@eng.kmutnb.ac.th⁴

In order to improve the intelligent service robot, the visual perception is a crucial. This paper presents the human detection for a service robot. The camera is installed at the robot by uprisen angle around 30 degree. Two feasible algorithms for the real-time detection between Haar cascade and Single Shot Detector (SSD) algorithm are compared. This research collects the training data as 1,000 images by the uprisen angle and the different views of the human, within 60 cm of user range. between a camera and human. The result shows that our proposed model of SSD method is higher performance than another by 0.933 of the average of IoU. Therefore, the proposed method is suitable to apply for the service robot.

Keywords: Human detection, Uprisen angle, Service robot, Single shot detection, Haar cascade.

1. Introduction

The service robot can see like a human using the camera. The camera position is initially fixed with an uprisen angle to see people at a 60 cm range distance. The field of view is designed for people with a height minimum of 120 cm to see the upper body in the frame grabber. However, it is just the first procedure of data acquisition. Next, the robot needs to understand the semantics of acquired images. This paper focuses on distinguishing humans from other objects. The robot must prioritize safety for every movement to beware of whether a human is in front of it. Moreover, the robot can interact with humans with different behaviors, such as greeting and playing with the people around. Many vision-based human detection techniques [1] are reviewed in various machine learning techniques and vision-based on revolutionary methods such as SIFT, SURF, PCA, and HOG. The human detection data sets are accessible publicly, such as MS COCO, ImageNet, and WiderPerson. However, this paper prepares its own data sets for the experiment.

The human face is a biometric feature or identity for feature extraction. Anirudha B Shetty et al. in 2021 [2] have compared two facial recognition techniques, Local

Binary Pattern and Haar Cascade. The results showed that Haar Cascade has higher accuracy than the Local Binary Pattern, and Haar Cascade uses more execution time than the Local Binary Pattern. Many researchers use some primary object detection techniques with good performance, such as Feature extraction using stepwise convolutional self-encoder, Faster R-CNN, YOLO, and SSD. Jintong Cai et al., in 2022 [3], selected the SSD technique to analyze bee behavior to detect honeybees. The result exhibited increased accuracy on multiple datasets, but the model's performance in complex scenarios is still limited. Rohit Raj et al., in 2020 [4], mentioned that Haar Cascade is advantageous over deep learning speed. In deep learning, such CNNs perform much better results at object detection, but they need a high-performance chipset like GPU or TPU for efficient inference. There are improvements in DSSD, FSSD, and ASSD [5] from SSD-based detectors. The result showed that ASSD has better the accuracy of SSD by a large margin at a small extra computation cost.

2. Proposed method

Although, the first version of service robot can detect the obstacle around itself with sensors co-working to other systems, but it cannot distinguish the obstacle types such as human and others.

According to the existing service robot function, the robot cannot appropriately response to the human or other obstacles. For example, the robot tries to avoid the human who want to interact with the robot instead of stopping itself or interacting to the user. Therefore, the visual perception is a crucial part for solving these situations. For the existing robot design, the webcam is placed on the body of the robot which is uprisen angle around 30 degree and high 120 cm. from the floor as same as the surface screen of the robot as shown in Fig 1.



Fig.1 Service robot and webcam installed position.

In order to classify the human and other objects appearing in front of the robot, this research focuses to two main algorithm that is suitable for a real-time system. First is Haar Cascade algorithm which is the machine learning base for object detection [6]. Haar features is used for extracting the value of each subarea in the images. Each feature is calculated as a single value by subtracting the number of pixels under white and black rectangles. Fig.2 is shown the Haar features such as edge features, line features and rectangle features.

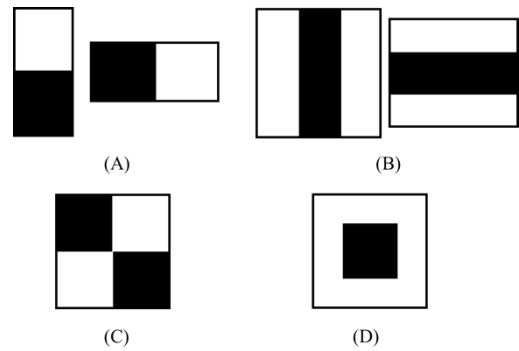


Fig.2 (A) Edge Feature, (B) Line Feature, (C) Rectangle Feature, (D) Center-Surround Feature.

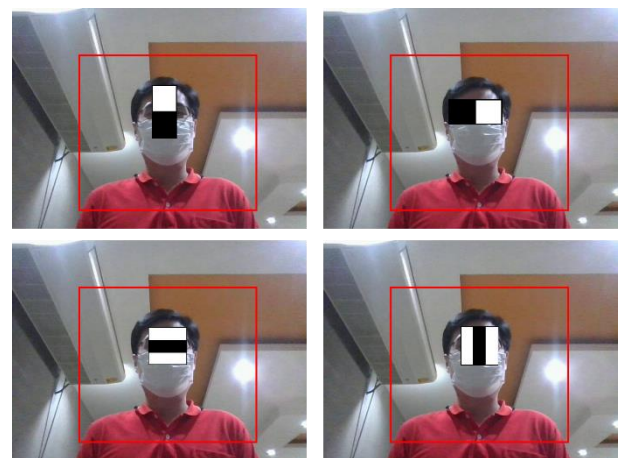


Fig.3 Example of features scanning on an image

Then, Haar features are processed as similar as the convolutional kernel. These will be scanning from top-left to the bottom right of the images for the several times with the various features of Haar as shown in Fig.3. The concept of integral images is applied for calculating rectangular features quickly. Instead of calculating at every pixel, this concept creates sub-rectangles and array references for each those sub-rectangles. Next, Adaboost training come into play for selecting the best features from the more 100,000 features. Final step is to use the cascade classifier for detecting the sub-region that contain the human face.

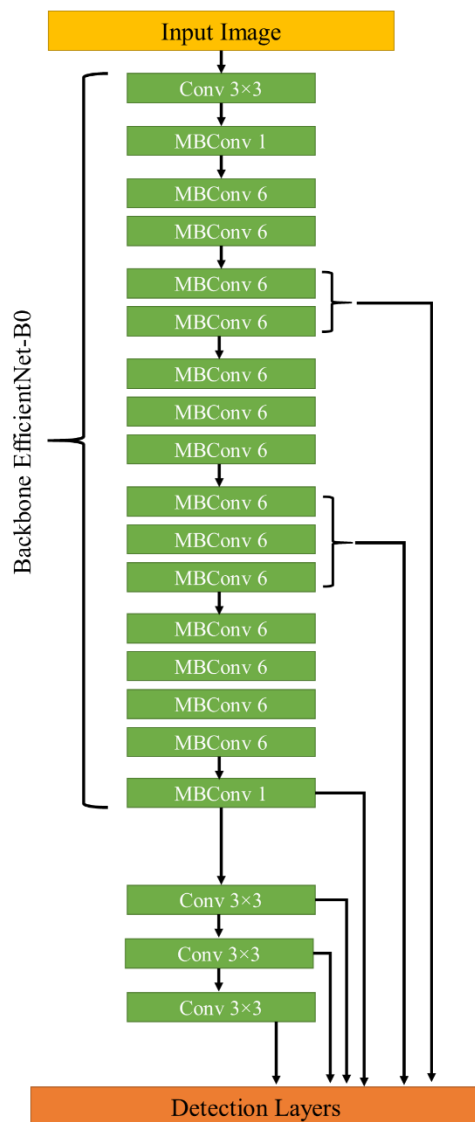


Fig.4 Single Shot Detector (SSD) of Effcientnet-B0 architecture.

Second is Single Shot Detector (SSD) which is the deep learning base [7]. SSD is a backbone network of Efficientnet-B0 as shown in Fig.4., that is adjustable the fully connected layer (FC). There are two different sizes of images between 300×300 px. and 512×512 px. applying with SSD, namely SSD300 and SSD512. Generally, the final output of SSD will show the boundary box of detected object and its category name with predicting score.

According to the original architecture, 20 classes of output are provided. This research changes the

The 2023 International Conference on Artificial Life and Robotics (ICAROB2023), Feb. 9 to 12, on line, Oita, Japan

transferring trained model to 2 classes output between human and others. Moreover, the SSD300 is selected for detection.

3. Experiments and Results

Based on these two possible algorithms, which are Haar cascade and SSD, for a real-time application. This section shows the experiment for comparing the performance of human detection.

Although, both of them already have the pre-trained data of human, but the uprisen angle of a camera still has some mistakes. Therefore, this research will compare and show the improvement between these algorithms.

As the position of the camera on the robot's body, it is installed around 120 cm. from the floor and the uprisen angle about 30 degrees. In addition, the user range is set by 60 cm. from the camera approximately. Therefore, the experiment compares the accuracy of the detector algorithm between Haar cascade and SSD which are shown before and after training the images of above condition.



Fig.5 Training images in the various views of people.

A thousand images are collected from 9 people with the various angles such as front side, right side, left side, back side of each person as shown in Fig.5. Then, all images are trained by Haar cascade and SSD with two classes of human and other objects. The experiment tests the 99 unknown images with various side of people, the images resolution is 640×480 px.

For the Haar cascade technique, this research uses pre-trained of upper-body for detection with scale factor of 1.05 and min-neighbor of 2. Our trained model of Haar cascade uses scale factor of 1.2 and min-neighbor 2.

Moreover, the SSD of pre-trained model is used for two classes of output, compared to our trained model. The configuration of our trained model defines the IoU (intersection of union) threshold, center variance and size variance as 0.5, 0.1 and 0.2, respectively.

Table 1. The Comparison of Average IoU between pre-trained and proposed models.

	Pre-trained model	Proposed
Haar cascade	0.120	0.607
SSD	0.868	0.933

Table 1. shows the comparison results of the average IoU for human detection as the user range with uprisen angle (30 degree) of a camera between Haar cascade and SSD. Eq. 1 shows the average of IoU which N is the number of testing images which the intersection and union areas are shown in Fig.6. As the showing result, Haar cascade presents the average IoU value as 0.120 and 0.607 for upper-body (traditional model) and our proposed. In addition, SSD technique shows the average of IoU of 0.868 and 0.933 for pre-trained and proposed models.

$$Average\ IoU = \frac{1}{N} \sum_{i=1}^N \frac{Area\ of\ Intersection}{Area\ of\ Union} \quad (1)$$

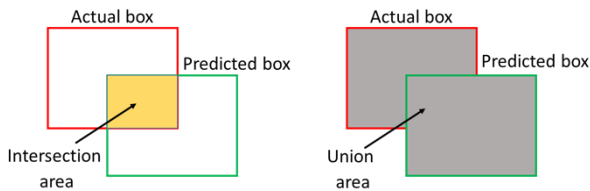


Fig.6 Intersection and Union area for actual and predicted boxes.

4. Conclusion and Discussion

This research proposes the human detection method that is feasible for the service robot with the uprisen angle of a digital camera around 30 degrees. Two possible techniques are selected for real-time detection such as Haar cascade and Single Shot Detector.



Fig.7 Example human detection between pre-trained and our proposed models of SSD.

As the results, both of Haar cascade and SSD shows the better performance in average of IoU for our proposed method. In addition, the highest performance is shown by SSD with our proposed model as 0.933 of average for IoU value. Fig.7 shows the comparison between pre-trained and our proposed model for SSD. Several times of pre-trained detection shows some mistakes, for example, tv_monitor or other classes are detected at the right side of the images while those areas have no any TV. These errors are calculated around 54% for pre-trained model, but the proposed model can extremely reduce the error. Furthermore, our proposed can detect the human with the various viewpoints.

Therefore, this research ensures that the proposed model of SSD is feasible for applying for our service robot with the uprisen angle (around 30 degrees) and 120 cm. from the based. Future work will implement to the other systems in the service robot.

Acknowledgements

This research was funded by the Ministry of Higher Education, Science, Research and Innovation, Thailand (Grant No.175613)

References

1. Sumit, Shahriar Shakir, Dayang Rohaya Awang Rambli, and Seyedali Mirjalili. "Vision-based human detection techniques: a descriptive review." *IEEE Access* 9 (2021): 42724-42761.
2. Shetty, Anirudha B., and Jeevan Rebeiro. "Facial recognition using Haar cascade and LBP classifiers." *Global Transitions Proceedings 2.2* (2021): 330-335.
3. Cai, Jintong, et al. "Single shot multibox detector for honeybee detection." *Computers and Electrical Engineering* 104 (2022): 108465.
4. Raj, Rohit, et al. "Feature based video stabilization based on boosted HAAR Cascade and representative point matching algorithm." *Image and Vision Computing* 101 (2020): 103957.
5. Yi, Jingru, Pengxiang Wu, and Dimitris N. Metaxas. "ASSD: Attentive single shot multibox detector." *Computer Vision and Image Understanding* 189 (2019): 102827.
6. Paul Viola and Michael J. Jones, "Robust real-time face detection", *International Journal of Computer Vision*, 57(2):137–154, 2004.
7. Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, pp.21-37, 2016.

Authors Introduction

Dr. Watcharin Tangsuksant



He received B.Eng. degree in Biomedical Engineering from Srinakharinwirot University, Bangkok, Thailand in 2013. His M.Eng. degree in Biomedical Engineering from King Mongkut's institute of technology Ladkrabang, Bangkok, Thailand in 2015. His D.Eng. degree in Department of life science and system engineering from Kyushu institute of technology, Wakamatsu campus, Japan, in 2019. He is currently a lecturer in Department of production and robotics engineering at King Mongkut's University of Technology North Bangkok. His research interest is imageprocessing and machine learning.

Asst.Prof.Dr. Amornphun Phonopas



He received his B.S. degree in Electronics Physics in 2005 from the Faculty of Science, Thammasat University in Thailand. His M.Sc. degree in Robotics and Automation from Institute of Field robotics (FIBO), Thailand. His D.Eng degree in Department of Computer Science and Systems Engineering from Kyushu institute of technology, Iizuka campus, Japan, in 2012. He is currently an instructor in Department of production and robotics engineering at King Mongkut's University of Technology North Bangkok. His research interest is robotics.

Dr. Pornthep Sarakon



He is a lecturer in the Department of Production and Robotics Engineering, King Mongkut's University of Technology North Bangkok, Thailand. He received a B.Eng. (Hons.) degree in Biomedical Engineering from Srinakharinwirot University, Nakhon Nayok, Thailand in 2015 and an M.Eng. degree in Information and Communication Technology for Embedded Systems from the Sirindhorn Inter-national Institute of Technology, Thammasat University, Thailand in 2017. He received a Ph.D. degree in Electrical and Electronic Engineering from Kyushu Institute of Technology (Kyutech), in 2021. His research interests include artificial intelligence, model compression, computer vision, and 3D body data.

Dr. Aran Blattler



He received B.Eng. degree in Production Engineering from King Mongkut's University of Technology North Bangkok, Bangkok, Thailand in 2013. His M.Eng. degree in Materials and Production Engineering (TGGS) from King Mongkut's University of Technology North Bangkok, Bangkok, Thailand in 2016. His D.Eng. degree in Department of Computer Science and Systems Engineering from Kyushu institute of technology, Iizuka campus, Japan, in 2021. He is currently a lecturer in Department of production and robotics engineering at King Mongkut's University of Technology North Bangkok. His research interest is high-precision measurement and high-precision robotics.
