SUGISAKA MASANORI

Journal of Robotics, Networking and Artificial Life
Vol. 9(3); December (2022), pp. 289–294
ISSN (Online):2352-6386; ISSN (Print):2405-9021
https://alife-robotics.org/jrnal.html

Research Article

Detecting Approaching Human Hands in a Human-Robot Coexistent Food Preparation Work Area for Preventing Collision

Joo Kooi Tan[1], Takaaki Yotsumoto[2], Yuta Ono[2]
[1]*Faculty of Engineering, Kyushu Institute of Technology, Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan*
[2]*Graduate y of Engineering, Kyushu Institute of Technology, Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan*

ARTICLE INFO

Article History
Received 26 November 2021
Accepted 24 September 2022

Keywords
Recognition
Hand area extraction
Lab color system
Grab cut

ABSTRACT

In Japan, the population of working age between 15 and 64 years old peaked in 1995 at about 87 million and is expected to continue to decline in the future. Therefore, to solve the labor shortage, the introduction of industrial robots that can perform the same level of work as humans is strongly requested especially in the food preparation industry. In order to prevent danger to workers there, industrial robots must recognize workers and avoid them when there is fear of collision. In this paper, we propose a method for extracting hand regions based on the color distribution of a hand and GrabCut in an experimental environment to recognize human hands and detect their directions of approach. The proposed method was examined experimentally and gave satisfactory results.

1. Introduction

According to an announcement by the Small and Medium Enterprise Agency [1], the working-age population (15-64 years old) in Japan peaked in 1995 at about 87 million and has been declining ever since, reaching about 77 million in 2015. This trend is expected to continue, and it is estimated that the population will decline to about 60% of the 2015 level by 2060. The working population (the total of "employed" and "totally unemployed" among the population aged 15 and over) declined by only about 420,000 between 1995 and 2015, due to an increase in the labor participation rate of women and those people aged 65 and over, but this decline is not as large as that of the working-age population. However, it is easy to imagine that the working population will decline in proportion to the decline in the working-age population. For this reason, the introduction of industrial robots, such as robot arms capable of performing tasks equivalent to those of humans, will be necessary to solve the labor shortage.

According to the Industrial Safety and Health Regulations [2] and the Ministry of Labor Notification No. 51 of 1983 (machines specified by the Minister of Health, Labor and Welfare under the provisions of Article 36, Item 31 of the Industrial Safety and Health Regulations) [3], except for those that meet certain standards, in order to use industrial robots, it is basically necessary to enclose them with fences and take other safety measures to prevent humans from entering. However, according to [4], by implementing measures based on the investigation of danger, etc., joint work between industrial robots and humans can be performed.

Corresponding author's E-mail: etheltan@cntl.kyutech.ac.jp URL: https://www.kyutech.ac.jp/~etheltan/

In order to prevent danger to the worker, it is necessary for the industrial robot to recognize the worker and avoid the worker when necessary.

In this study, as the first step for an industrial robot to avoid collision with workers, we consider a situation where robots and humans collaborate in an assembly line such as a lunch-box assembly line. We propose a method to detect one's neighbors' hands that approach his/her work area by extracting hand areas from ego camera images.

One of the methods for hand area extraction is to use a distance image [5], [6], [7] to capture and recognize its 3-D shape. However, the infrared camera used in this method has a drawback that it is easily affected by sunlight, which limits the locations where it is used. Other methods that use multiple sensors [8], [9] or special sensors [10] have been proposed, but they are considered to be difficult to use because of the high price of the system. On the other hand, extraction or segmentation techniques are typically based on skin color [11], [12], [13], background subtraction with hand tracking [14], or appearance-based model [13]. Technique [11] uses skin color and face color: [12] uses two kinds of color spaces, i.e., HSV and TSL to differentiate between skin color and non-skin color: Moreover, [13] uses a skin color model with fine-tuning to the depth for segmentation of two hands. [14] uses a particle filter for tracking non-skin color (the background) and detecting skin color by the Gaussian Mixture Model (GMM) with a combination of edge information for segmentation of a hand. Both skin color and background subtraction-based techniques use cues of HSV and RGB color space and are limited to capturing the target from a distant fixed camera. [15] adopts a hand model. However, the technique requires three specific views of a hand model for calibration.

2. Summary of the Research

This section gives an overview of the system proposed in this paper. The proposed system detects a human hand or a robot arm approaching the work area in front of a user (a worker) from the images of a head-mounted camera

Fig.1. Head-mounted camera

attached to the user (See Fig. 1) to avoid collision between the user's hand and the neighbors' hands.

Initially, in order to detect the hand approaching from the side of the user, the system acquires hand color information of a hand area on the spot where the system is used in the given lighting conditions. The skin color information is acquired from image frames prepared in advance and integrated as the knowledge for reference. Then, the system detects the hand area using the knowledge and identifies the hand from the area.

2.1. *Acquisition of color information of a hand area*

The proposed system acquires the color information of a hand area in a practical environment before judging the approach of a human hand, and creating a color distribution map of the hand at the place and the time.

2.2. *Judging approach of a human hand*

In the first place, the hand region is extracted from the input image based on the hand region color distribution map obtained in advance, and noise is removed. Then the system identifies whether or not a hand area is approaching a user's work area, and if it is true, it displays a bounding box containing the hand area in the image as the result of the detection.

3. Acquisition of Hand Area Information

In this section, we describe a method to acquire color information of a hand area from the captured video.

The first frame of the video to acquire the hand area information is the background frame without a hand. Then, while keeping the orientation of the camera unchanged, frames containing several patterns of hand areas are acquired for a few seconds as shown in Fig.2. These frames are used as the frames for extracting the hand color information.

3.1. *Extraction of candidate hand areas*

As shown in Fig. 3(a), most of the image frames acquired from the head-mounted camera used in the proposed system have a large background area except for hand areas. Since these areas interfere with the acquisition of the color information of a hand area, we delete these areas. We detect candidate hand areas by extracting the difference region between each color information extraction frame and the background frame described above (See Fig.3(b)). For each pixel, a foreground pixel is extracted to prevent false extraction due to changes in brightness. The $L^*a^*b^*$ color system is used for this process. After that, GrabCut and fine line segmentation are applied to the hand area extraction image for possible improvement in detection accuracy.

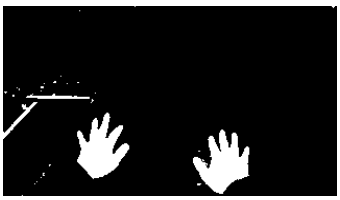
Since the areas which do not have a skin color are not considered to be hand areas, the skin color areas are extracted using the HSV color system.

3.2. *Extraction of hand areas*

The color information is extracted from the color information extraction frame using the candidate hand area extracted in Section 3.1 as a mask. To select the color distribution of the hand area, we use the difference between the color distribution of the color information extraction frame and that of the background frame. The color distribution histograms of the two frames are compared to select the hand area color distribution in the processing frame. Finally, the color distributions in the hand area obtained from the color information extraction frames are combined by logical OR, and the distribution having the largest area is selected to remove noise in the case of selecting a color distribution other than the hand area by mistake. In the proposed method, the above process is used to obtain the color distribution of the hand area dependent on the usage environment.

Fig.2. Color information extraction frame

(a)



(b)

Fig. 3 Extraction of area candidate: (a) Input frame, (b) candidate hand areas

4. Judgment on the Approach of a Human Hand

In this section, we describe a method of judging the approach of a human hand from an input image.

To judge the approach of a human hand, the proposed method identifies if the approach is occurring, and if so, it displays a bounding box containing a hand area in the image where the hand approach is occurring.

4.1. *Extraction of hand areas*

The hand area is extracted from the input video. The hand area color distributions of the camera wearer and his/her neighbors, which were created in advance by acquiring hand area color information, are used to extract the pixels having a^* and b^* values that are included in the color distribution. Then, by using a median filter as a denoiser, the actual hand area is finally extracted from the input video. An example of hand area extraction after denoising is shown in Fig.4.

4.2. *Judging the approach of a human hand*

The proposed method judges the approach of neighbors' hands to the work area in front of a camera wearer/robot.

First, labeling of extracted regions on a given image is performed, and among the obtained labeled regions, judgment is made on the regions with a size larger than a specified value. Two attention areas are defined at the right-hand side and the left-hand side of the work area. If a labeled hand area is found in the attention area, it is judged dangerous, because the found hand area may have a fear of coming into the frontal work area.

There are four patterns of judgments from the camera wearer's/robot side: no approach, approach from the right-hand side, approach from the left-hand side, and approach

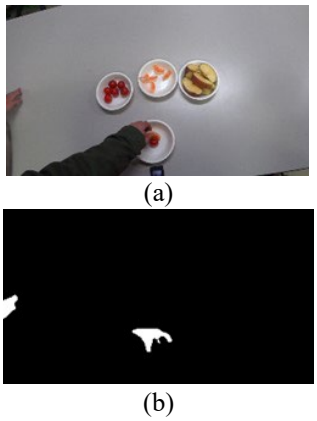


Fig. 4 Hand area extraction: (a) Input images, (b) extracted hand areas.

from both sides. Note that the proposed system assumes there are three people; the camera wearer him/herself, the right neighbor, and the left neighbor.

5. Experimental Results

An experiment was conducted to examine the performance of the proposed method. Since the focus is on the detection of a hand approaching from the side toward the center work area in front of a user/robot, a rather simple scene containing some dishes with food was set (See Fig. 5). The scene was captured by a head-mounted camera of the user whose play as a robot and the video provided by the camera was processed for detecting an approaching hand.

It is noted that, although robot hands should be included in the experiment, all the hands for detection are human hands in this particular experiment for simplicity. Namely the user and the neighbors in the experiment are all humans.

The system performed the analysis on 150 images extracted every 10 frames from the input video. The accuracy of correct detection was evaluated by creating a ground truth image and by calculating three major indices; *recall*, *precision,* and *F-measure*. As the result of the experiment, the *recall* was 81.0%, *precision* was 98.6%, and *F-measure* was 88.9% as shown in Table 1. Some results on the detection of approaching hands are given in Fig. 5.

6. Discussion and Conclusion

This paper proposed a system for detecting a hand approaching from the side of a center work area in front of a worker who is a user of the proposed system. For the detection, a head-mounted camera was used which the user wears. Advantages of the employment of the head-mounted camera include that the system can escape from careful camera setting at the work spot and that it can gain mobility in, e.g., a factory. The effectiveness of the proposed system was shown experimentally.

The system has originality in that it analyzes the work at hand based on the first-person vision. The idea of detecting hands approaching from the side toward the center work area is also the originality of the paper. The

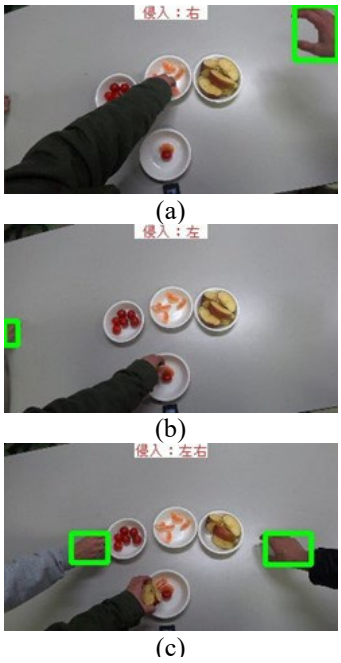


Fig. 5 Experimental results: (a) A hand approaching from the right-side, (b) approaching from the left-side, (c) approaching from the both sides.

Table 1 Accuracy evaluation

Accuracy evaluation frames[frame]	150
TP [frame]	68
FP [frame]	1
FN [frame]	16
$recall$ [%]	81.0
$precision$ [%]	98.6
F- measure [%]	88.9

hand detection described in this paper will become necessary in the future for the line robots working with human line workers to avoid mutual hand collision.

Some issues need further investigation to achieve higher performance of the proposed system. Advanced hand-part color calibration should be done more accurately since people have a variety of skin colors. The case of wearing colored gloves should also be taken into account. Alternatively, the shape information of a hand may be useful for hand detection, since it is less influenced by the change in illumination.

In the present system, the detection of a hand area is done every 10 frames independently. In order to assure correct detection of a hand area, the detected hand area needs tracking in the video, so as to compensate for false or missed detection of the hand area at certain image frames.

References

1. The Small and Medium Enterprise Agency, "2019 White Paper on Small and Medium Enterprises in Japan", The Small and Medium Enterprise Agency, p.121, 2018. (In Japanese).
2. Ministry of Health, Labour and Welfare, "Ordinance on Industrial Safety and Health", Ministry of Health, Labour and Welfare, 1972. (In Japanese).
3. Ministry of Health, Labour and Welfare, "Ministry of Labor Notification No. 51 of 1983 (Machinery specified by the Minister of Health, Labour and Welfare pursuant to the provisions of Article 36, item 31 of the Industrial Safety and Health Regulations)", Ministry of Health, Labour and Welfare, 1983. (In japanese).
4. Ministry of Health, Labour and Welfare, "Partial revision of the enforcement notice of article 150-4 of the industrial safety and health regulations concerning industrial robots", Ministry of Health, Labour and Welfare, 2012. (In Japanese).
5. S. Tanaka, A. Okazaki, N. Kato, H. Hino, K. Fukui, "Spotting fingerspelled words from sign language video by temporally regularized canonical component analysis", 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), 7pages, 2016.
6. K.O. Rodriguez, G.C. Chavez, "Finger spelling recognition from RGB-D information using kernel descriptor", IEEE Conference on SIBGRAPI, 7 pages, 2013.
7. N. Pugeault, R Bowden, "Spelling it out: Real-time ASL fingerspelling recognition", IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp.1114-1119, 2011.
8. H. Brashear, V. Henderson, K. H. Park, H. Hamilton, S. Lee, T. Starner, "American sign language recognition in game development for deaf children", Proc. of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, pp.79- 86, 2006.
9. H. Brashear, T. Starner, P. Lukowicz, H. Junker, "Using multiple sensors for mobile sign language recognition", Proc. of the 7th IEEE International Symposium on Wearable Computers, pp.45-52, 2003.
10. Yu Wang, Seiji Itai, Satoshi Ono, Nakayama Shigeru, "Human Recognition with Ear Image by Principal Component Analysis", Journal of Japan Society of Information and Knowledge, Vol. 16, No. 1, pp.15-27, 2005. (In Japanese)
11. A. Bapat, V. Kanhangad, "Segmentation of hand from cluttered backgrounds for hand geometry biometrics", IEEE Region 10 Symposium, TENSYMP, 4 pages, 2017.
12. L. Gu, X-y. Yuan, T. Ikenaga,"Hand gesture interface based on improved adaptive hand area detection and contour signature", IEEE International Symposium on Intelligent Signal Processing and Communication System, PP.463-468, 2012.
13. H-S. Shih, C-H. Ma,"Hand segmentation with skin color Fine-turning using Kinect Depth Sensor", IEEE Global Conference on Consumer Electronic, 2Pages, 2017.
14. V.Spruyt, A. Ledda, S. Geerts, "Real time multi-colourspace hand segmentation", IEEE International Conference on Image Processing, pp.3117-3120, 2010.
15. J.J. Kuch, T. S. Huang, "Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration", IEEE International Conference on Computer Vision, PP. 666-671, 2002.

Authors Introduction

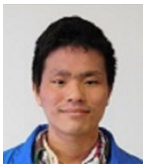
Prof. Dr. Joo Kooi Tan

Professor Tan is currently with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor. Current research interests include ego-motion analysis, three-dimensional shape and motion recovery, human detection and its motion analysis from videos. Professor Tan was awarded SICE Kyushu Branch Young Author's in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. Professor Tan is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.

Mr. Takaaki Yotsumoto

He obtained his B.E. degree in the Department of Control Engineering from Kyushu Institute of Technology in 2019, Japan. He is presently a master's course student at the same university. His research interests include human detection and recognition.

Mr. Yuta Ono



He received his M.E. from Kyushu Institute of Technology. He is now in the Ph.D. course at the Graduate School of Engineering, Kyushu Institute of Technology. His research includes computer vision, machine learning, and rush-out pedestrian detection from videos.