Collision Avoidance in a Human-Robot Coexistent Food Preparation Environment Using Hands Area Extraction

Takaaki Yotsumoto
Graduate of Engineering, Kyushu Institute of Technology, Japan, Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan

Yuta Ono, Joo Kooi Tan*
**Faculty of Engineering, Kyushu Institute of Technology, Japan, Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan*
E-mail: {Takaaki-yotsumoto , etheltan}@ss10.cntl.kyutech.ac.jp
https://www.kyutech.ac.jp/~etheltan/

Abstract

In Japan, the population of the working-age between 15 and 64 years old is peaked in 1995 at about 87 million and is expected to continue to decline in the future. Therefore, in order 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 a food preparation industry. In order to prevent danger to workers there, it is necessary for industrial robots to recognize workers and avoid them when there is fear of collision. In this paper, we propose a method of extracting hand regions based on the color distributions of a hand and GrabCut in an experimental environment to recognize human hands and detect their directions of approach. The performance of the proposed method is shown experimentally.

Keywords: Recognition, Hand Area Extraction, Lab Color System, Grab cut

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 into the future, 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 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 the partial revision[4] of the enforcement notice of Article 150-4 of the Industrial Safety and Health Regulations concerning industrial robots, when it can be assessed that the risk of danger to workers due to contact with industrial robots has been eliminated by implementing measures based on the investigation of danger, etc., joint work between industrial robots and humans can be performed. 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 a worker, we propose a method to recognize a human hand and to detect the direction of its approach by extracting the hand region.

One of the methods for hand region extraction is to use a distance image[5, 6, 7] to capture and recognize the shape in 3D. 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 expensive sensors[10] have been proposed, but they are considered to be difficult to use because of the high price of the system.

2. Summary of Research

This section gives an overview of the system proposed in this paper. The proposed system for detecting the proximity of a human hand or a robot arm recognizes the proximity of a human hand in front of a head-mounted camera attached to the user as shown in **Fig. 1**. Before recognizing the approaching hand, the system obtains the color information of the hand area in the lighting environment where the system is used. Then, the system extracts the hand region from the fingerprint input frame, and identifies the hand from the region.

2.1. *Acquisition of hand area color information*

The proposed system obtains the color information of the hand area in the system environment before judging the proximity of a human hand, and creates the color distribution of the hand at that place and time.

2.2. *Judging the proximity of a human hand*

First, the hand region is extracted from the input image based on the hand region color distribution obtained in

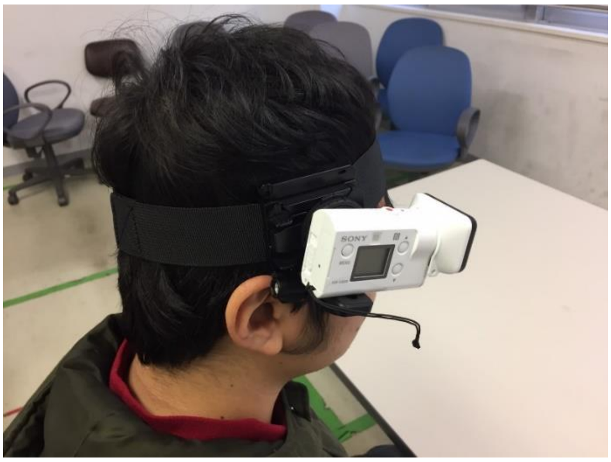


Fig.1. Head mounted camera

the system's operating environment as described in Section 2.1, and noise is removed. After that, the system identifies whether proximity is occurring, and if so, it displays a rectangle on the image where the proximity is occurring as the result of the recognition.

3. Acquisition of Hand Area Information

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

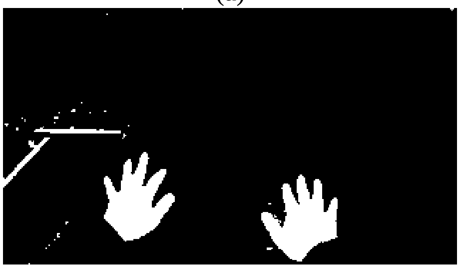
The first frame of the video to acquire the hand area information is the background frame without a hand. After that, while keeping the orientation of the camera unchanged, frames containing several patterns of hand regions are input for a few seconds as shown in **Fig.2**, which are used as the frames for extracting color information.

3.1. *Extraction of candidate hand region candidates*

As shown in **Fig. 3(a)**, most of the images acquired from the head-mounted camera used in this study are the background except for the hand region. Since these areas interfere with the acquisition of the color information of the hand area and are clearly background areas, we delete these areas. We detect the hand region candidates by extracting the difference region between each color information extraction frame and the background frame described above. For each pixel, a foreground region f 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

Fig.2. Color information extraction frame

(a)



(b)

Fig. 3 Extraction of a hand region candidate: (a) Input frame, (b) hand region candidates.

are applied to the hand region extraction image for possible improvement in detection accuracy.

Since areas other than skin color are not considered to be hand areas, the skin color areas are extracted by using the HSV color system.

3.2. *Extraction of candidate hand regions*

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

4. Judging the Proximity of a Human Hand

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

To judge the proximity of a human hand, the method identifies whether proximity is occurring, and if so, it displays a rectangle in the image where the proximity is occurring.

4.1. *Extraction of hand regions*

First, the hand region is extracted from the input video. The hand region color distributions of the camera wearer and his neighbors, which were created by acquiring hand region color information, are used to extract the fruits of a^* and b^* values that are included in the color distribution. Then, by using a median filter as a denoiser, the hand region is first extracted from the input video. Using the color distributions of the hand regions of the camera wearer and his neighbors, which are created by acquiring the hand region color information, the real part is extracted, although the values of a^* and b^* are included in the color distribution. After that, a median filter is used for denoising to obtain the hand region extracted image.

4.2. *Judging the proximity of a human hand*

The proposed method judges the proximity of a human hand to the obtained hand region extracted image.

First, labeling is performed, and among the labels obtained by labeling, judgment is made for the labels with an area of a certain area or more.

If the area of the hand is more than a certain area in the left or right danger area, it is judged to be dangerous, because there is approach on the left or the right. The judgment results of the hand region extracted images are ‘no approach’, ‘approach to the right side’, ‘approach to the left side’, ‘approach to both the right and the left side’,

Table 1 Accuracy evaluation

Accuracy evaluation frames[frame]	149
TP [frame]	114
FP [frame]	27
FN [frame]	9
$recall$ [%]	92.68
$precision$ [%]	80.85
F- measure [%]	86.36

'approach to the right side', 'approach to the left side'. There are four patterns of judgments from the camera wearer's side: 'no approach,' 'approach to the right side,' 'approach to the left side,' and 'approach to both the right and the left sides'. method judges the proximity of a human hand to the obtained hand region extracted image. First, labeling is performed, and among the labels obtained by labeling, judgment is made for the labels with an area of a certain area or more.

If the area of the hand is more than a certain area in the left or right danger area, it is judged to be dangerous because there is approach on the left or right. The judgment results of the hand region extracted images are "no approach", "approach to the right side", "approach to the left side", "approach to both the right and left side", "approach to the right side", "approach to the left side. There are four patterns of judgments from the camera wearer's side: "no approach," "approach to the right side," "approach to the left side," and "approach to both the right and left sides.

5. Experimental Results

The system recognizes the presence or absence of approach and the location of approach in images acquired by a head-mounted camera. The input images do not have a single background, and the recognition is performed on 149 images extracted every 10 frames from the input images. The accuracy is calculated by creating a ground truth image and calculating the recall, precision and *F*-measure. As the result of the experiment, recall was 92.68%, precision was 80.85% and *F*-measure was 86.36% as shown in **Table 1**.

6. Conclusion

The causes of false positives are discussed here. First, there are cases where the correct image is not created when viewed in a single frame, but it is created in the previous frame. In this case, the approach detection of a person cannot be performed correctly in a single frame. Second, it is considered difficult to acquire the hand region information every time. Therefore, it is necessary to devise a system that can detect a person, even if the hand region information is acquired only by the person wearing the camera.

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, 7pages, 2013.
7. N. Pugeault, R Bowden, "Spelling it out: Real-time ASL fingerspelling recognition", IEEE International Conference on Computer Vision Workshops (ICCV Workshops), p.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

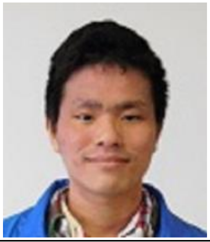
Information and Knowledge, Vol. 16, No. 1, pp.15-27, 2005. (In Japanese)

Authors Introduction

Mr. Takaaki Yotsumoto

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

Mr. Yuta Ono



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

Prof. Dr. Joo Kooi Tan

She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor Her current research interests include ego-motion, three-dimensional shape and motion recovery, human detection and its motion analysis from videos. She 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. She 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.