The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

# On Detecting a Human and Its Body Direction from a Video

Yuuki Nakashima, Joo Kooi Tan, Seiji Ishikawa, Takashi Morie
Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan
E-mail: {nakashima, etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp, morie@brain.kyutech.ac.jp

***Abstract***: This paper describes a novel technique for detecting a human and its body direction using HOG feature. The HOG feature is a well-known feature for the judgment of a human. But normally it contains the background feature, which gives negative influence on the judgment. This paper proposes the employment of the HOG feature based on a human model. The feature is also employed for detecting human body direction. Experimental results show effectiveness of the proposed technique compared to the conventional one.

***Keywords***: HOG, human models, body direction, SVM

## 1. Introduction

Recently, car vision technologies have been paid much attention in the field of ITS (Intelligent Transport System), aiming at reducing traffic accidents to a large extent. Various techniques for detecting cars, car lanes, traffic signs, etc., have been developed by the employment of a PC with high specification and cheap digital cameras. In particular, techniques for automatic detection of pedestrians or humans have been studied enthusiastically.

Although human detection from outdoor scenes is one of the difficult problems in computer vision because of a variety of clothes, occlusion, illumination change, image resolution, etc., a robust human detection algorithm has been reported recently. The HOG feature proposed by Dalal and Triggs [1] is a well known feature for representing and recognizing a human image. But, in its original algorithm, it includes an unnecessary part of the image such as the background. It is not conceivable that the background feature contributes to high recognition rate of a human image. Zhu et al. [2] employ the HOG feature based on variable block size, but they also include the background in the feature.

In this paper, we propose a method of detecting a human and its body direction employing the HOG feature based on a human model. The reason for detecting body orientation is to predict his/her sudden step into a road. The present method differs from other existent methods in that (i) the HOG feature is calculated referring to a human model and therefore it does not include the background in a given image, and that (ii) human body direction is also recognized by the HOG feature based on the human models which represent respective body directions. To the best of our knowledge, the HOG feature has been employed to date only for detecting a human and not for recognizing its body direction. The proposed method is described in the following and the performance of the method is shown by the experiments employing real outdoor scenes.

## 2. Detecting a human

### 2.1 The HOG feature

The HOG (Histograms of Oriented Gradients) feature [1] is based on image gradients vectors. An image is separated into non-overlapping cells containing $5\times5$ pixels. Gradients are calculated at each pixel on the cell and the histogram on the orientation of the calculated gradients is made with the cell. These cells are collected into a block containing $3\times3$ cells and all the nine histograms are normalized on the block. This block is defined at every cell: Namely, the blocks overlap spatially. Finally, all the blocks containing the normalized histograms are collected into a single vector. This is the HOG feature of an image. As is understood from the above procedure, the background image is inevitably included in the HOG feature.

### 2. 2 A human model

The main idea of the proposed method is to calculate the HOG feature on a human model instead of calculating it on the whole image. A human model is created in the following way.

**1:** The intensity of gradient is calculated of all the images in an image database.
**2:** The average values of the intensity of the gradient are calculated at every pixel on the images and a normalized average gradient intensity image is produced.
**3:** An edge image is produced from the average gradient intensity image.
**4:** A silhouette image is created from the average gradient intensity image and a skeleton image is made from the silhouette image.
**5:** Finally a human model image is yielded by performing logical OR operation between the edge image obtained at step 3 and the skeleton image obtained at step 4.

The flowchart of the above steps is given in **Fig. 1**. Examples of the images at each step of the above procedure are depicted in **Fig. 2**.

### 2. 3 A feature vector

Using the human model, the method calculates the HOG feature as shown in **Fig. 3**. The histograms are made only at the points specified on the human model. We set a cell ($5 \times 5$ pixels) at each white pixel on a human model. A gradient orientation histogram $\boldsymbol{a}_i = (a_{ij})$ is made at the $i$th cell ($i$=1,2,…,$N$; $j$=1,2,…,9). These histograms are collected into a single vector $\hat{\boldsymbol{v}}$ of the form

$$\hat{\boldsymbol{v}} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N) \equiv (a_{ij}) \\ (i = 1,2,\ldots,N;\ j = 1,2,\ldots,9) . \tag{1}$$

Vector $\hat{\boldsymbol{v}}$ is normalized into vector $\boldsymbol{v}$ of the form

$$\boldsymbol{v} = (v_1, v_2, \ldots, v_{9N}) \tag{2a}$$

by

$$v_k = \frac{a_{ij}}{\sqrt{\|\hat{\boldsymbol{v}}\|^2 + \varepsilon^2}},\ k = 9(i-1) + j \tag{2b}$$

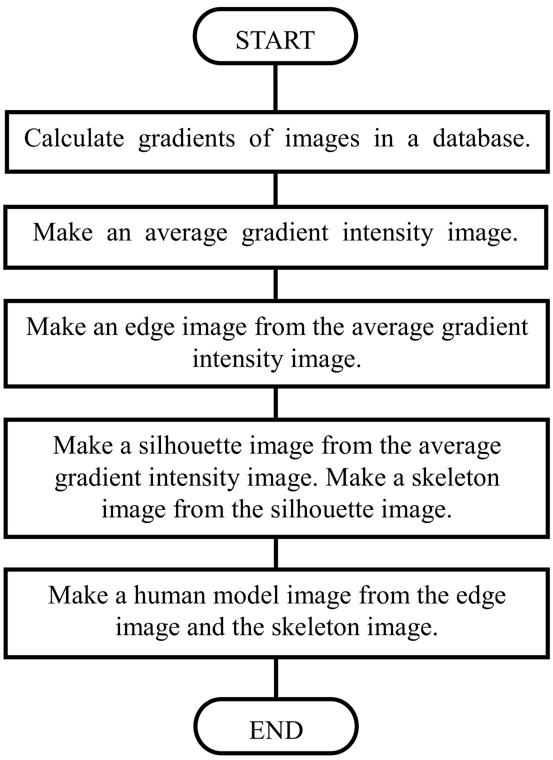


Fig.1. Flow chart of making a human model.

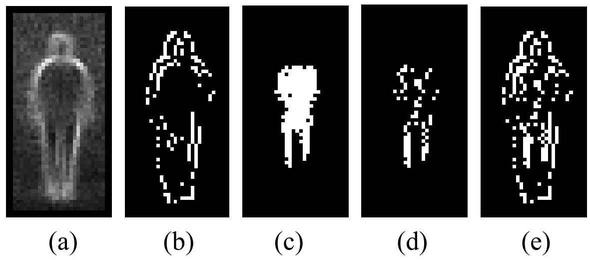


Fig. 2. Images obtained at each step of the procedure: (a) An average gradient intensity image, (b) an edge image, (c) a silhouette image, (d) a skeleton image, and (e) a human model image.

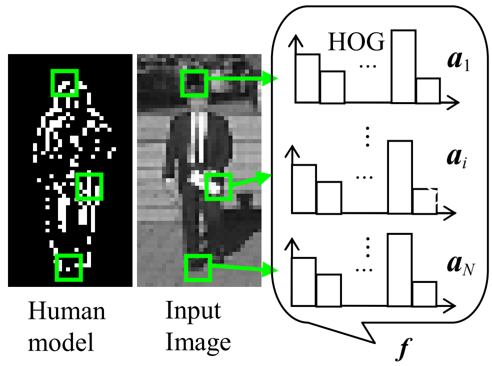


Fig. 3. Histogram creation employing a human model.

The vector $\boldsymbol{v}$ of Eq. (2) finally represents the given image based on the human model. The proposed method employs this vector for the judgment of human or non-human images. It is noted that the present method does not employ a block introduced in [1]. Instead, the cells scattered and overlapped on an image according to the human model are collected into a single feature vector $\boldsymbol{v}$.

A database is prepared for the development of a recognition system. The database contains a number of human images as positive data and a number of images without a human as negative data. The feature vectors of Eq. (2) are calculated from these images to define a non-linear SVM in the feature space. Human image recognition is then performed employing the SVM.

## 3. Detecting human body direction

The proposed method is applied to detecting human body direction. The direction is roughly separated into front(F), right(R) and left(L), and three human models are defined, respectively. In the first step of the detection, three SVM classifiers are generated and used:

They are based on one vs. rest method, i.e., left or others, front or others, and right or others. If not decided, it proceeds to the second step where two SVM classifiers are generated and used based on one vs. one method, i.e., left or front, and front or right. The judgment scheme is given in **Fig. 4**.

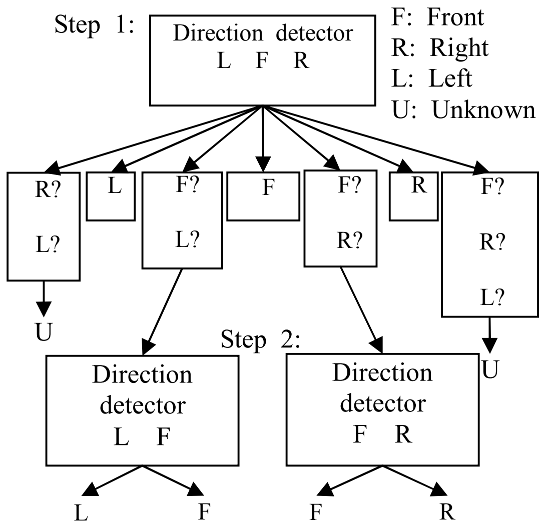


Fig. 4. Judgment of body direction.

## 4. Experimental results

### 4. 1 Detecting a human

We use 1602 human images as positive training dataset and 3235 images without a human as negative training dataset. For validation, we employ a dataset containing 1000 human images and the images without a human. The image size is 30 × 60 pixels. The employed human model is illustrated in **Fig. 5a**. Some learning images and test images are shown in Fig. 5b and Fig. 5c.

The proposed method is compared with the conventional method [1] in order to evaluate its performance. The result is provided by DET (Detection Error Trade-off) curves as shown in **Fig. 6**. When the true negative rate is 0.5[%], the false positive rate is 7.5[%] by the proposed method, whereas it is 19.1[%] by the conventional method. Thus the false positive rate has decreased by 11.6[%] in the proposed method compared to the conventional one. The computation time was 29.1 [msec/window] in the proposed method, whereas it was 54.4[msec/window] in the conventional one.

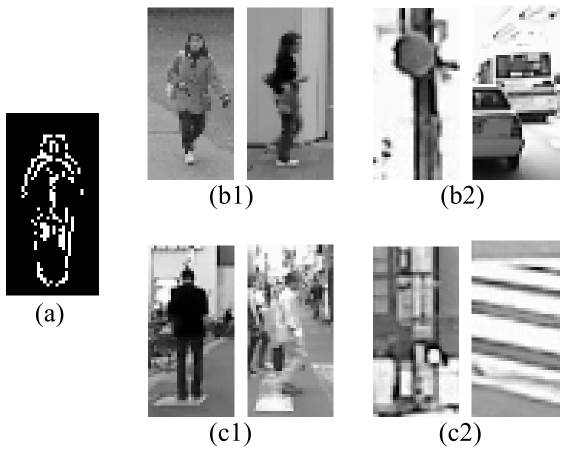


Fig. 5. Employed human model and some of the learning images and test images: (a) Human model; (b) Learning images, (c) test images; (1) human images, (2) images without a human.

### 4. 2 Detecting human body direction

In this experiment, 217 images with each direction are used as learning data. As for validation, we use 2 datasets. Dataset 1 contains 200 images with each direction captured in a specified environment (no occlusion and human natural standing posture). On the other hand, dataset 2 is a set of images taken in a real environment. It also contains 200 images with each direction. The proposed method is compared its performance with the conventional method [1]. The result of recognition (the rate of correct detection of body direction) is given in **Table 1**. The computation time was 26.7 [msec/image] in the proposed method, whereas it was 60.6[msec/image] in the conventional one.

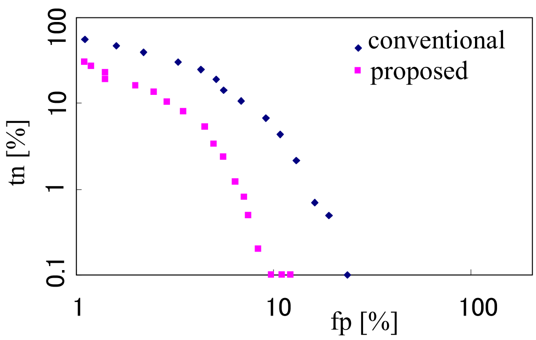


Fig. 6. The DET curve of the proposed method and the conventional method.

Table 1. Recognition rate (%)

| method | Dataset | success | error | unknown |
|---|---|---|---|---|
| conventional | 1 | 67.8 | 8.2 | 24.0 |
| | 2 | 56.0 | 22.3 | 21.7 |
| proposed | 1 | 94.2 | 1.0 | 4.8 |
| | 2 | 79.0 | 15.3 | 5.7 |

## 5. Discussion

We have proposed a method of detecting a human and its body direction employing the HOG feature [1]. The novelty of the present paper is that the HOG feature is calculated on a human model in the proposed method, unlike other reported techniques in which the HOG feature is calculated in a cell that spreads all over a given image. Since HOG features calculated in the background of an image may not contribute much in human recognition, the proposed method can discard the undesired background effect to a large extent. In the original literature [1], three by three adjacent cells are collected into a single block and the block is chosen one after another on the image so that they may have enough mutual overlap. This overlap among blocks is realized by the overlap among cells in the present method. After all, the feature vector defined by Eq. (2) keeps the HOG features only within and on the border of a human figure.

Another novelty of the present paper is that the HOG feature is employed in detecting human body direction. The proposed method detects not only a human facing front but also a human facing sideways by the employment of respective human models. It is well known that the HOG feature is a strong and robust feature for recognizing a human in an arbitrary environment. Extending the recognition technique employing the HOG feature to detecting human body direction may have application to car vision, where a person should be predicted who might step or run into a road out of a side walk.

The experimental results show effectiveness of the proposed method. We understand that undesired background effect has decreased to a certain extent, as the HOG feature was calculated based on a human model. However, not all HOG features may be necessary for the detection. We have a plan to select effective feature locations, which may result in better performance of the proposed method.

## 6. Conclusions

A method was proposed for detecting a human and its body direction employing the HOG features based on a human model. Performance of the method was satisfactory. Refinement of the human model by selecting effective feature locations remains for further study.

## Acknowledgement

This work was partly supported by a grant of Knowledge Cluster Initiative implemented by Ministry of Education, Culture, Sports, Science and Technology (MEXT).

## References

[1] N. Dalal, B. Triggs: "Histograms of oriented gradients for human detection", Proc. IEEE Conf. on Computer Vision & Pattern Recognition, 886-893, 2005.
[2] Q. Zhu, et al.: "Fast human detection using a cascade of histograms of oriented gradients", Proc. IEEE Conf. on Computer Vision & Pattern Recognition, 2006.