IEICE The Institute of Electronics, Information and Communication Engineers

IEICE Proceeding Series

Robust Human Detector based on HLAC and HOG using RGB-D

Miho Morita, Hiroshi Takemura, Hiroshi Mizoguchi

Vol. 1 pp. 110-113
Publication Date: 2014/03/17
Online ISSN: 2188-5079

2012 International Symposium on Nonlinear Theory and its Applications
NOLTA2012, Palma, Majorca, Spain, October 22-26, 2012

Robust Human Detector based on HLAC and HOG using RGB-D

Miho Morita , Hiroshi Takemura, Hiroshi, Mizoguchi

Department of Mechanical Engineering, Tokyo University of Science
2641 Yamazaki, Noda-shi, Chiba, 278-8510 Japan
Email: j7511659@ed.tus.ac.jp

Abstract–This paper proposes a new human detection method using RGB-D data. RGB-D data include RGB images and Depth (stereo) images. Stereo image can represent human's three-dimensional shape. This method is base integration with two features: Higher-order Local Auto-Correlation (HLAC) features and Histogram of Oriented Gradients (HOG) features. We extend HLAC features from RGB images and HOG features from Stereo images. HLAC features can give a broad pattern of gray scale image. HOG features can give an accurate description of contour of human body. To use Stereo images, HOG features can give a contour more accurate than conventional method. We use co-occurrence of multiple features to integrate HOG and HLAC features, called IHH. In our experiments, we obtain 12.0% lower value on false positive per window than other proposed IHH method when miss rates are similar. These results proved the effectiveness of this new method.

1. Introduction

Human detection in image has becoming an important research area. Many researches have been done in this filed. Human detection is expected to application in various fields, such as a video monitoring system and a robot vision.

Human detection has two problems. Figure 1 shows situations of each problem. One of problem is detector cannot decide which object is human, the other is detector confuse human with other things. In order to make up these deficits, we need average features of human. However, shape of human differs widely in pose, change of appearances and directions of taking photographs. Though we take a photograph at same direction, human has individual such as wearing various clothes and occlusion of body part as well as complex background. Using brightness values as are, we can't succeed in human detection because of useless information. We extract features from brightness values such as contours and use these for detection. One of famous method is using Local Edge Oriented Histograms (EOH) features which proposed by Levi and Weiss [1]. It is robust to change of appearance. However, it will be affected by illumination change.

In this paper, we proposed new robust human detection method. This method uses two types of images. One is intensity image that is converted from RGB image. That shows difference of brightness values clearly. The other is stereo image that give distance between camera and each object. We extend Higher-order Local Auto-Correlation (HLAC) features [2] from RGB (intensity) images and Histogram of Oriented Gradients (HOG) features [3] from stereo images. HLAC features can give a broad pattern of image. HLAC features are invariant to shift, but poor at illumination change. Same as EOH, HOG features can give an accurate description of contour of human body, but HOG features are robust to illumination change. Extracting HOG features from stereo image, these are robust of changing distance from camera and objects. Besides, these can give description of contour more clearly. However, they often confuse human with other objects that similar contour. To make up these deficits each other, we integrate these features as Figure 2 shows. In order to integrate these features , we use Real AdaBoost [4]. Real AdaBoost choice features co-occurrence and train one classifier. We use the classifier to detect human.

2. Human Detector Algorithm

In this section we explain human detection method we proposed. We use many images to train classifier *C*. These images are called training images. We extend HLAC features from RGB (intensity) images and HOG features from stereo images. In order to robust human detect, we integrate with these two features and named it IHH (Integrated with HLAC and HOG) features. We calculated

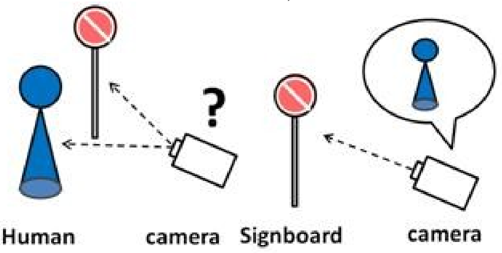


Fig.1 Problem of human detection

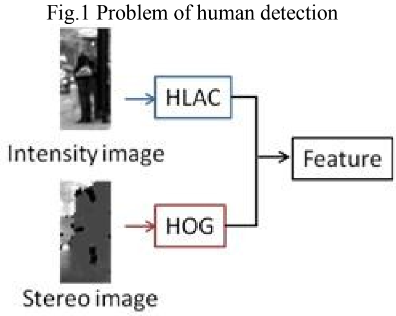


Fig.2 Proposed method

IHH features from other images called test images. Compared test image's IHH feature with classifier's. Calculated evaluation value *H*, and if H is larger than threshold value, classifier decide that IHH value is human's. We get coordinate of that IHH's and compare classifier's coordinate. Potion of human is decided. Figure 3 shows overview of human detection.

2.1. Features

We use HLAC (Higher-order Local Auto-Correlation) feature and HOG (Histograms of Oriented Gradients) feature. HLAC features often use for human face image detection [2], [4]. The feature value of HLAC is the auto-correlation function calculated from the object image. HLAC can give broad pattern of images. HLAC is invariant to shift, but poor at illumination change. When a picture illumination lowers, HLAC features values became smaller than the same situation without effect of illumination changes. For this reason, HLAC features are poor at illumination change. HOG features can give an accurate description of contour of human body. HOG features are robust to illumination change. However, HOG features are confused by pattern of clothes. Extracting HOG features from stereo image, this problem is solved. These are robust of changing distance from camera and objects.

2.2. Method of Integration

The method is based on integration with HLAC and HOG features. To use these two features simultaneously, we represent the co-occurrence of multiple features [6].

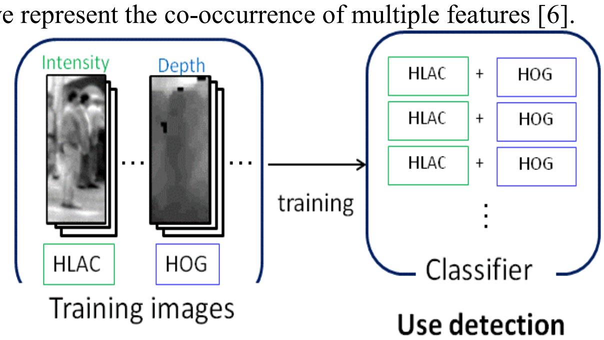


Fig.3 Overview of human detection

2.2.1. Select features

First we label N sample images two classes, positive class (human) and negative class (no human) before train classifier. HLAC and HOG features are calculated images divided the window into small spatial regions ("cells"). The feature co-occurrence is represent several features are observed at the same time as shown in Figure 4. In this research, we can get four types of values, both are positive classes, HLAC features are positive class and HOG features are negative class, the opposite combination, and both are negative classes. For each class statistical dependencies between the features are obtained for each training sample image. We use such dependencies for classification. The input image is classified by evaluating from which class the feature co-occurrence is likely to be observed. The features are selected to capture discriminative similarities of the samples that belong to the class. The classifier using these features is generated in the Real AdaBoost using joint probability. The join probabilities W_+, W_-(W_+ is positive class's, W_-(is negative class's probability) defined as follows:

$$\mathrm{W}_+ = \sum \mathrm{D}_{+\mathrm{t}}$$

$$\mathrm{W}_- = \sum \mathrm{D}_{-\mathrm{t}} \tag{1}$$

$$\mathrm{D}_1 = \frac{1}{\mathrm{N}}$$

$$\mathrm{D}_{\mathrm{t}-1} = \mathrm{D}_{\mathrm{t}} \exp\left(-\frac{1}{2}\ln\frac{\mathrm{W}_+ + \varepsilon}{\mathrm{W}_- + \varepsilon}\right) \tag{2}$$

where D is weights of samples, N is the number of sample images, t (0<t<T) is stage of learning and ε is a coefficient for preventing division by zero problems.

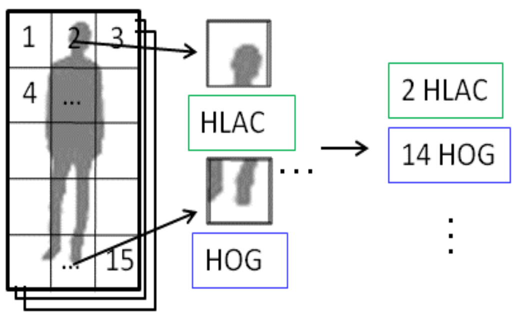
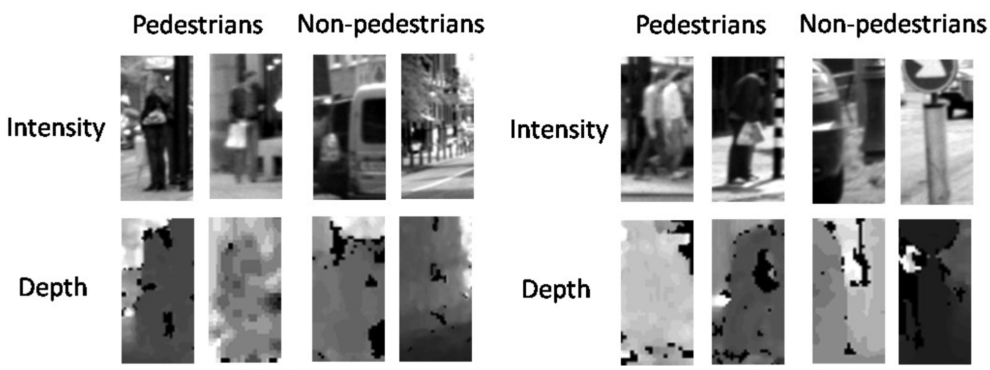
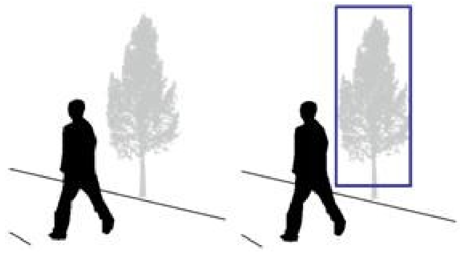


Fig.4 Divide the sample image



(a) Training images (b) Test images

Fig.5 Examples of sample images



(a) Miss (b) False positive

Fig.6 Situation of detection

2.2.2. Integration with each feature

After learning, the classifier has data of common combination of HLAC and HOG features and cell numbers and joins probabilities of these features. When detection of test image, we calculate combination of HLAC and HOG features. We compare these and classifier's data and calculate *h(x)* defined as follows:

$$h(x) = \frac{1}{2}\ln\frac{W_{+}+\varepsilon}{W_{-}+\varepsilon} \quad (3)$$

where x is input image data. If the $h(x)$is larger than any threshold which is suitable, we regard the sample is human.

The features make it possible to classify difficult samples that are misclassified by weak classifiers using a single feature. Its can support various appearances and different points of view, and this method is robust to occlusions.

3. Experiment condition

Demonstration of the effectiveness of the proposed method, we conducted experiments to evaluate the human detection. We trained classifiers of 4 distinct types in the experiments. The first one is proposed IHH method that integrated with RGB (intensity) image's HLAC and stereo image's HOG features. The second is jointed multiple HOG features. We extend HOG from intensity images and stereo images. The third is integrated with intensity image's HLAC and intensity image's HOG. The fourth is integrated with stereo image's HLAC and intensity image's HOG. Method of integrated with intensity image's HLAC and HOG is more effective than full body HOG method and jointed multiple intensity image's HOG features [7]. Therefore we compared these four methods.

First, we explain how to collect positive or negative class's samples. Samples were corrected from Daimler Multi-Cue Occluded Pedestrian Classification Benchmark [8]. It contains collection of images labeled pedestrian and non-pedestrian bounding boxes in images captured from a vehicle-mounted calibrated stereo camera rig in an urban environment. Dense stereo is computed using the semi-global matching algorithm [9]. We was scaled all samples to the size 30×60 pixels. The positive samples include various pose variation.

We tested detectors on images contain two type of images collected by DaimlerChrysler. The "Non-

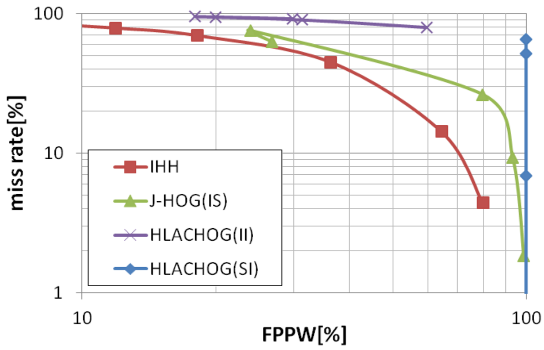


Fig.7 Detector performance of "Non-Occluded"

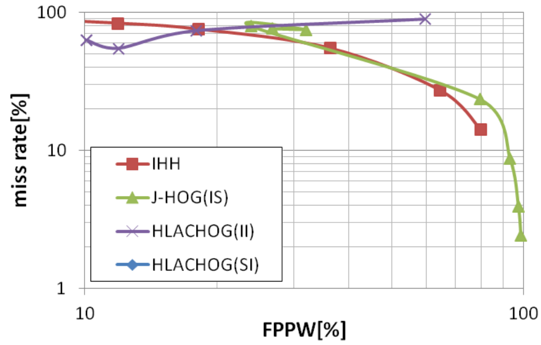


Fig.8 Detector performance of "Partially Occluded"

Occluded" contains pedestrian images without inter-human occlusion. It has 25608 samples. The "Partially Occluded" contains pedestrian include inter-human occlusion. It has 11160 samples. Non-pedestrians images have 16235 samples. Figure 5 shows a part of samples of train and test images by DaimlerChrysler.

For comparisons of detection performance, we used DE T (Detection Error Tradeoff) curve [10]. Detection tasks c an be viewed as involving a tradeoff between two error ty pes: miss rate and FPPW (false positive per window). Mis s rate is performance rate that classifiers cannot detect hu man when we experiment use pedestrian images. FPPW is rate that classifier confuse other objects with human in an experiment using non-pedestrians images. On DET curve s, the x-axis corresponds to FPPW, and the y-axis shows t he miss rate on log-log scales.

We explain about a situation of each rate means in Figure 6. These show the tradeoff. The lower both of rates are the higher efficient method is. In DET Curve we plot these error rates on both axes. Better distinguishes different well performing systems are close to linear.

4. Experiment

We compared four methods, proposed IHH, jointed HOG (J-HOG(IS)), integrated with intensity's HLAC and HOG (HLACHOG (II)), and integrated stereo's HLAC and intensity's HOG (HLACHOG(SI)). Figure 9 and Figure 10 shows a part of result of human detection. We use 9000 pedestrian (positive) images and 13000 non-pedestrian (negative) images to train classifiers. We mark miss rate and FPPW of the results on Figure 7 and Figure8.

Table 1 Mean value of detector performance rate of the "Non-Occluded"

	IHH	J-HOG(IH)	HLACHOG(II)	HLACHOG(SI)
Distances[-]	59.8	73.3	96.1	101.6
Miss rate [%]	42.5	35.1	90.7	18.0
FPPW[%]	42.1	64.4	31.7	99.9

Table 2 Mean value of detector performance rate of the "Partially Occluded"

	IHH	J-HOG(IH)	HLACHOG(II)	HLACHOG(SI)
Distances[-]	68.0	75.1	72.0	100.0
Miss rate [%]	57.4	38.5	68.6	0.027
FPPW[%]	36.3	64.5	21.9	99.9

We calculated mean value of distance from origin of coordinates to each point on linear graphs. When the distance is a small value, both of the error rates are low. The mean values of the distance and error rates show in Table 1and Table 2.

From Figure 7, the most effective method is proposed method IHH. Table 1 shows the minimum value of distance is IHH. It support that the most effective method is the proposed method, next on is the jointed HOG. The distance of proposed method is 13.5 lower than jointed HOG.

From Figure 8 and Table 2, IHH is the most effective method of all. The second effective method is HLACHOG (II). The distance of IHH is 7.1 lower than integrated intensity's HLAC and HOG.

Compared to the jointed HOG, our method is 12.0% lower on FPPW when miss rate is similar value the experiment using "Non-Occlude" images.

Thus, our method is robust to distinguish human from other thing as well as detect human which have occlusion problem. Table 2 gives the same result. The distance of proposed method is 25.8 lower than jointed HOG. When miss rate of proposed method is 28.8%, FPPW is 68.2%. Compared to the jointed HOG, our method is 11.0% lower on FPPW when miss rate is similar value.

We see that the proposed method detection performance the most effective in other methods. Integrating different features, we can discriminate between human and other things better.

5. Conclusion

This paper proposed a new human detection system using RGB-D data. RGB-D data include RGB images and Depth (stereo) images. This method is base integration with two features: Higher-order Local Auto-Correlation (HLAC) features and Histogram of Oriented Gradients (HOG) features. We extend HLAC features from RGB images and HOG features from Stereo images. We use co-occurrence of multiple features to integrate HOG and HLAC features. We called proposed method IHH. In our experiments, we obtain 12.0% lower value on false positive per window than other proposed IHH method when miss rates are similar. These results proved the effectiveness of this new method. We compared four methods, proposed IHH, jointed HOG, integrated with intensity's HLAC and HOG, and integrated stereo's HLAC and intensity's HOG. The experimental results proved that the effectiveness of the proposed method.

Our method is 12.0% lower on false positive per window than proposed method when miss rates are similar value. These experimental results proved the effectiveness of our new method. In the feature, we need to detect human on real time using video camera by this proposed method.

Acknowledgments

This work is supported by The Hara Research Foundation (tentative).

References

[1] K. Levi et al., "Learning Object Detection from a Small Number of Example: the Importance of Good Features," n *Proc. IEEE Conf. on CVPR*, vol. 2, pp. 53–60, 2004.

[2] K. Hotta et al., "Scale Invariant Face Detection Method using Higher-Order Local Auttocorrelation Features extracted from Log-Polar Image", *AFGR*, pp. 70–75, 1998.

[3] N. Dalal et al., "Histograms of Oriented Gradients for Human Detection", *CVPR*, pp.886-893, 2005.

[4] W. Lin, M. Oakes and J. Tait, "Real AdaBoost for Large Vocabulary Image Classification", *CBMI* pp.192-199, 2008.

[5] A. Hidaka et al., "Object Detection by Integration of HLAC Mask Features", *Proc. Of ECSIC Symposium on BLISS* 2008.

[6] T. Mita et al., "Discriminative Feature Co-Occurrence Selection for Object Detection", *PAMI*, vol.30, No.7, pp.1257-1269, 2008.

[7] M. Morita et al.,"Human Detection Method Based on Feature Co-occurrence of HLAC and HOG" *Proceedings of the 2011 IEEE ROBIO*, pp.1689-1694, 2011.

[8] M. Enzweiler et al., "Multi-Cue Pedestrian Classification with Partial Occlusion Handling", *IEEE Conference on CVPR*, pp.990-997, 2010.

[9] H. Hirschmueller et al., "Stereo Processing by Semi-global Matching and mutual Information," *IEEE Trans Pattern Anal Mach Intel.* , 30(2) pp.328–341, 2008.

[10] A. Martin et al., "The DET Curve in Assessment of Detection Task Performance", *Proc. Eurospeech '97*, vol. 4, pp. 1899-1903, 1997.