

Directional Entropy Feature for Human Detection

Long Meng, Liang Li, Shuqi Mei, Weiguo Wu
Sony China Research Lab

{Long.meng, Liang.Li, Shuqi.Mei, Weiguo.wu}@sony.com.cn

Abstract

In this paper we propose a novel feature, called directional entropy feature (DEF), to improve the performance of human detection under complicated background in images. DEF describe the regularity of region by computing the entropy value of edge points' spatial distribution in specific direction, so DEF has the discriminating power for regular and random pattern. We combine Histogram of Oriented Gradient (HOG) feature with DEF to construct a human detection classifier to test DEF's performance. Experimental results show that DEF can help HOG to decrease false alarms caused by random complicated and rigid shaped background.

1. Introduction

The framework of object detection generally includes three parts: feature selection, design and training of classifier and design of detection procedure. This paper is focused on the feature design. Nowadays, a variety of features have been used in publications for object detection, such as corner, edge (or edgelet [1]), patch [2, 3], point descriptor [4], part and component [5, 6] or Haar-like [7, 8], HOG [9, 10], covariance descriptor [11] etc.

In fact, different features have different discriminating power. Haar-like feature describes the difference between object parts. HOG focuses on the gradient's orientation. Edgelet captures local shape patches character. A boosting cascade with multi-type feature pool could combine the different discriminating power of different type of features and achieve better performance than single type of features [12, 13]. In [12], Haar-like, HOG and LBP features are combined by online boosting method. Similarly, Haar-like, Gabor and EOH are utilized by a dynamic boosting method in [13]. In this paper, we try to find new kind of features with different characteristic from others especially for human detection.

Human detection is very difficult owing to the variance of human pose and clothes, complicated background. For example, trees, columns, body parts often bring false alarms for human detection. We implemented Dalal & Triggs algorithm as [9] and analysis the false alarms. HOG feature divides block into cells, which focuses on the local distribution of gradient intensity of several directions and cells inside one block. It is calculated by summarizing of oriented gradient of all pixels inside a cell. Because it does not care the edge points' distribution inside a cell, when gradient summary of cells for a complicated background image is similar to that of human image, the classifier by HOG will give a false alarm.

We propose a new feature, called Directional Entropy Feature (DEF), to record the regularity of object parts, which has never been considered before. It is used to remove false alarms of complicated background which are difficult for other kinds of features. Combined with other features, DEF can help remove more false alarms. In this paper, we use HOG + DEF to construct the classifier by SVM training to test the performance of DEF.

This paper is organized as follows, Section 2 describes the principle and design for DEF, the implementation of a human detection classifier based on HOG+DEF, experimental results and discussion are given in Section 3, finally conclusion is draw in Section 4.

2. Directional Entropy Feature

2.1. Principle of DEF

Fig. 1 (a, b) shows an example of human image and its gradient image. Fig. 1 (c) is a false detection by HOG classifier. From its gradient image as shown in Fig. 1 (d), it can be clearly seen that its texture looks like a human, thus HOG classifier gives false decision. We want to find a feature to discriminate human from such complicated background. Obviously, the distribution pattern of edge points in (b) and (d) is visually different, although they result into similar

HOG feature. The difference motivates us to find novel features to show the difference between human and human-like textures.

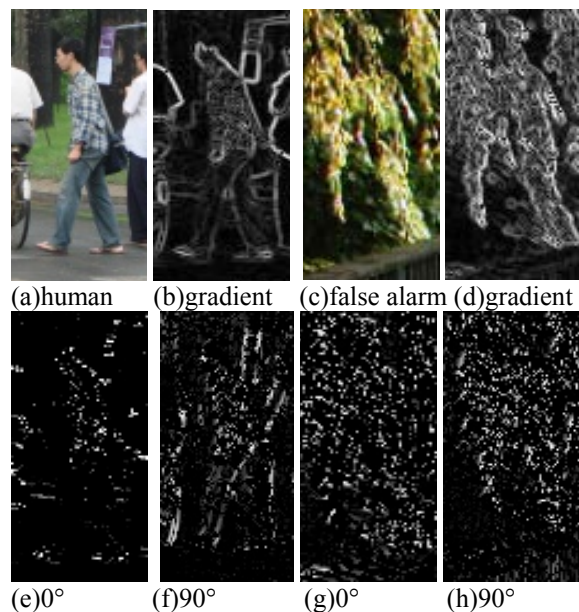


Figure 1 Object and background example

Fig. 1(e, f) is directional sub-images of Fig. 1(b) and Fig. 1(g, h) is that of Fig. 1(d). Fig.1(e, g) remains all edge points whose gradient orientation is near 0°; Fig.1(f, h) remains edge points whose gradient orientation is near 90°. For regular objects, the edge points tend to form lines orthogonal to the direction of their gradient's as shown in Fig. 1 (e, f). For background image, the edge points' distribution may be in a random pattern as shown in Fig. 1(g, h). The regularity will not change too much as human pose changes, so regularity of edge point's distribution is a good indicator which can reflect the difference between regular objects and background, meanwhile, tolerates the variance inside regular objects at the same time.

The regularity or randomness in opposite viewpoint, of the edge points' distribution could be represented by the entropy of the directional edge points' locations. Therefore we propose the Directional Entropy Feature to record the regularity of edge points' distribution in certain direction.

We collect 2000 human images and 2000 false detections by HOG classifier. What we want to do is to find features to discriminate these images. We calculate one DEF value between the 2 kinds of images. As Fig. 2, more than 40% human image's DEF value is about 0.8, while false alarmed images' are concentrated on 0.85. The DEF value's distributions

for the two classes of images are different. That is to say, DEF can help HOG for discriminate human and background. For images in Fig. 1, the feature value of the human image is 0.78, while that of false alarm image is 0.91.

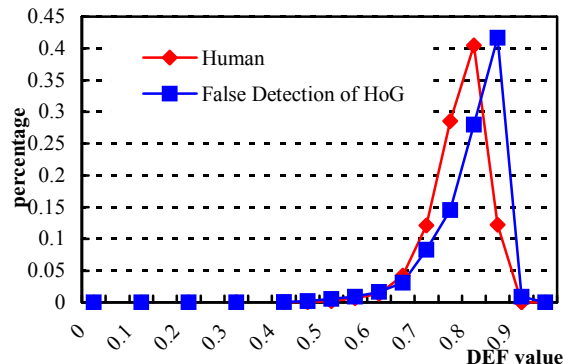


Figure 2. The DEF value distribution

2.2. DEF design

For each gradient sub-image with gradient direction a , the left top pixel's location is the origin, x-coordinate is from left to right, and y-coordinate is from top to down. The new axis $x'-y'$ is rotated by the axis $x-y$ with angle a . The new location of edge point with location (x, y) is (x', y') . As shown in Fig.3, (x', y') can be calculated as follows:

$$s = \sqrt{x^2 + y^2}$$

$$b = \text{arctg}(y/x)$$

$$x' = s \times \cos(b - a)$$

$$y' = s \times \sin(b - a)$$

s is the distance of point $P(x, y)$ and origin $O(0, 0)$. a is the direction of the gradient sub-image, that is the angle between old axis and new axis. b is the angle between line OP and x-axis.

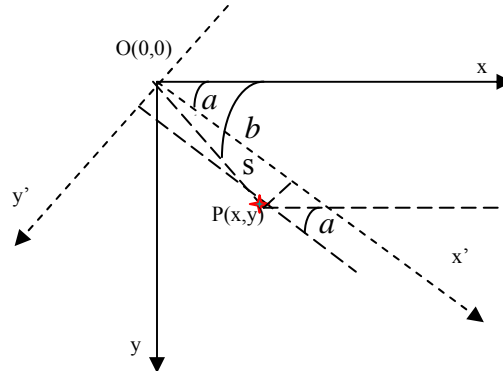


Figure 3. The relationship between (x, y) and (x', y')

DEF is designed to be block based. From anyone of the gradient sub-image, we can get blocks with any size and location. For block $B(x_0, y_0, w, h)$ as the rectangle in Fig. 4 ($a=90^\circ$), left top position (x_0, y_0) , width = w , height = h , the pixels, whose gradient directions are near a , form line orthogonal to a , if the image is a regular object. As shown in Fig. 5, the projections to y' -coordinate of pixels from line orthogonal to a are concentrated in smaller range than that of random pattern.

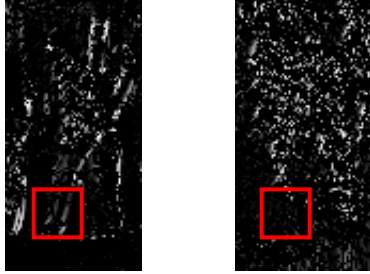


Figure 4. 90° gradient sub-image

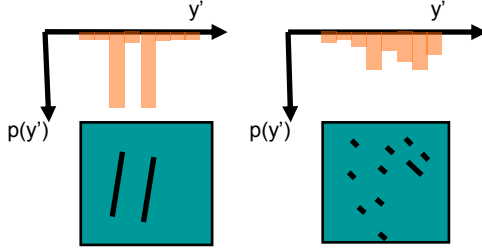


Figure 5. Edge points projection

The gradient weighted density of edge points' distribution on y' -coordinate of the block B can be calculated as follows:

$$p(y') = \frac{\sum_{\substack{x \in (x_0, x_0+w) \\ y \in (y_0, y_0+h) \\ |\theta(x,y)-a| < \Delta}} |Grad(x,y)| \cdot s \cdot \sin(b-a) = y'}{\sum_{\substack{x \in (x_0, x_0+w) \\ y \in (y_0, y_0+h) \\ |\theta(x,y)-a| < \Delta}} |Grad(x,y)|}$$

$Grad(x, y)$ is the gradient intensity, $\theta(x, y)$ is the gradient orientation of pixel (x, y) . Δ is the tolerance range of orientation, which is set to be half of the angle difference between two continued gradient sub-images in our test. The DEF of the block B in gradient sub-image whose direction is a can be calculated as follows:

$$E(x_0, y_0, w, h, a) = - \sum_{\substack{x \in (x_0, x_0+w) \\ y \in (y_0, y_0+h)}} p(y') \log_h p(y')$$

3. Implementation and Result

In [9], Dalal and Triggs use a detection window with size $64*128$ pixel. Each detection window is divided into 105 blocks, 4 cells and 9 bins for a block, resulting into 3780-D vector. These features are trained by linear SVM to form a classifier.

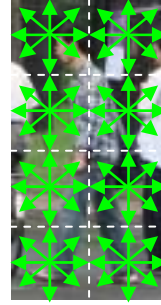


Figure 6. Block and direction for DEF

We use the 3780-D HOG and the new DEF at the same time. As shown in Fig. 6, each detection window is divided into 8 blocks with size $32*32$, 4 directions for a block. DEF vector adopted in our test has 32 dimensions. The classifier is trained by linear SVM. We use the same training and testing databases as [9].

For HOG classifier, false detections appear at complicated background such as Fig. 7, HOG + DEF classifier has better results (Fig. 8). On the same testing database, HOG + DEF can decrease around 10% false detections by HOG classifier under the same detection rate, especially on the complicated but random background or man-made objects such as arrow and car which are much more regular than human.

As shown in Fig. 7, HOG classifier has many false detections in the background, while the HOG + DEF classifier can effectively reduce these false detections, resulting in a cleaner detection result.

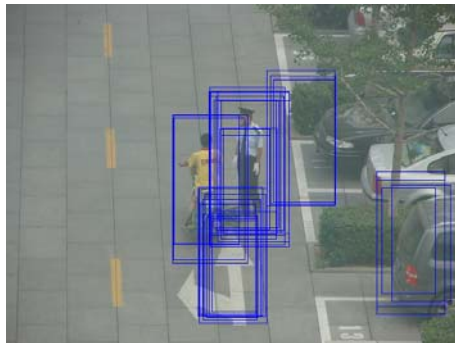
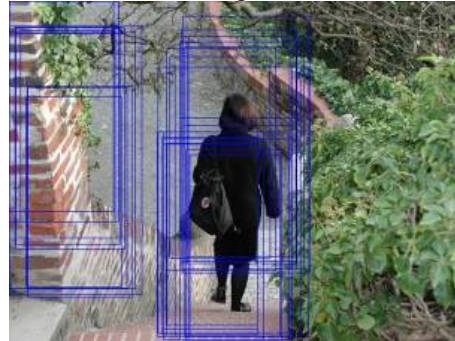


Figure 7. Detection result by HOG classifier

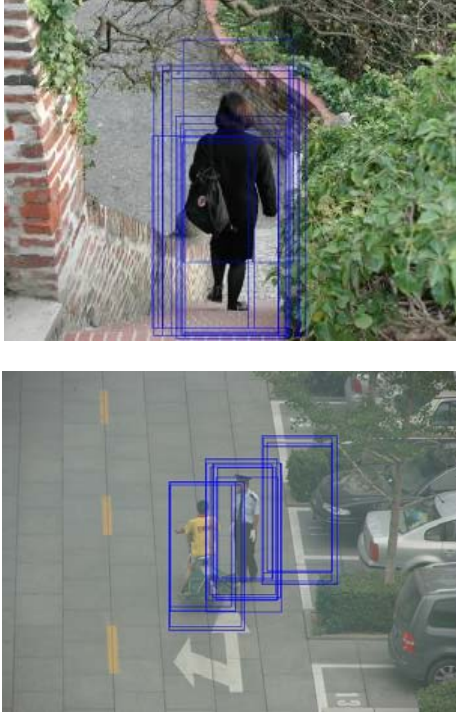


Figure 8. Detection result by HOG+DEF classifier

In our test, HOG is used to be main part of the classifier. The total number of features is near 4000, while added DEF vector is only 32-D. Consider the dimensions of DEF, the improvement ratio is acceptable. Although 32-D DEF vector is used in this paper, the number 32 is not optimized. If we change the dimension, or block size, or directions, the improvement may be more obvious.

4. Conclusion

For human-like object, the difference inside objects is in a wide range. DEF is a block based feature. When changes of human body's pose and location are inside blocks, the edge points' locations are changed while the regularities keep same, so the entropy will not change. That is to say, DEF tolerates object changes inside some range.

The discriminating power of DEF is regularity and random, which is a new character never be discussed before. Since it is calculated from the directional location distribution of key points inside a block, the

information used is different from other features, thus DEF could complement other features for object detection. Experiments show that DEF can help to remove false alarms especially on background with random pattern or rigid man-made shapes.

References

- [1] B. Wu and Ram Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. ICCV 2005.
- [2] A. Torralba, K. P. Murphy and William T. Freeman. Sharing Features: Efficient Boosting Procedures for Multi-class Object Detection. CVPR 2004, Vol. 2, page(s): 762-769.
- [3] S. Agarwal, A. Awan, and D. Roth. Learning to Detect Objects in Images via a Sparse, Part-Based Representation. IEEE TRANSACTIONS ON PAMI, VOL. 26, NO. 11, NOVEMBER 2004.
- [4] S. M. Bileschi, B. Leung and R. M. Rifkin. Towards Component-based Car Detection. 2004 ECCV Workshop on Statistical Learning and Computer Vision.
- [5] H. Schneiderman and T. Kanade. Object Detection Using the Statistics of Parts. International Journal of Computer Vision, 2002.
- [6] A. Mohan, C. Papageorgiou, and T. Poggio. Example-Based Object Detection in Images by Components. IEEE TRANSACTIONS ON PAMI, VOL. 23, NO. 4, APRIL 2001.
- [7] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. CVPR 2001.
- [8] Alexander Kuranov, Rainer Lienhart, and Vadim Pisarevsky. An Empirical Analysis of Boosting Algorithms for Rapid Objects with an Extended Set of Haar-like Features. Intel Technical Report MRL-TR-July02-01, 2002.
- [9] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. CVPR 2005, Vol. 1, page(s): 886- 893.
- [10] Q. Zhu, S. Avidan, M. C. Yeh, and Kwang-Ting Cheng. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. CVPR, 2006.
- [11] O. Tuzel, F. Porikli, and P. Meer. Human Detection via Classification on Riemannian Manifolds. CVPR 2007.
- [12] H. Grabner and H. Bischof. Online Boosting and Vision. CVPR, 2006.
- [13] R. Xiao, H. Zhu, H. Sun and X. Tang. Dynamic Cascade for Face Detection. ICCV 2007.