

RESEARCH ARTICLE

Integral Histogram with Random Projection for Pedestrian Detection

Chang-Hua Liu*, Jian-Kun Lin

Department of Medical Imaging, The 174th Hospital of Chinese PLA, The Succeed Hospital affiliated to Xiamen University, Xiamen, Fujian Province, China

* liuxingc@126.com

Abstract

In this paper, we give a systematic study to report several deep insights into the HOG, one of the most widely used features in the modern computer vision and image processing applications. We first show that, its magnitudes of gradient can be randomly projected with random matrix. To handle over-fitting, an integral histogram based on the differences of randomly selected blocks is proposed. The experiments show that both the random projection and integral histogram outperform the HOG feature obviously. Finally, the two ideas are combined into a new descriptor termed IHRP, which outperforms the HOG feature with less dimensions and higher speed.



OPEN ACCESS

Citation: Liu C-H, Lin J-K (2015) Integral Histogram with Random Projection for Pedestrian Detection. PLoS ONE 10(11): e0142820. doi:10.1371/journal.pone.0142820

Editor: Rongrong Ji, Xiamen University, CHINA

Received: April 15, 2015

Accepted: October 27, 2015

Published: November 16, 2015

Copyright: © 2015 Liu, Lin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The HOG [1] feature has become one of the most popular descriptors in object recognition since being proposed in 2005. The advantage of the HOG feature is that it's not sensitive to small shift and illumination change, so it can encode the edge information of the object efficiently. This is due to its oriented-gradients based, normalized histogram extracted from (overlapped) blocks for description. However, several issues of HOG remain open, i.e.: (1) the feature dimension and the computing cost are high. (2) It performs poor in occlusions. (3) The other region-based information are lack, such as color, shape and texture.

In order to improve the computation efficiency, Zhu et al. [2] allowed the blocks to vary somewhat in size and used the Integral Histogram [3] to calculate the HOG feature quickly. The distinctive blocks are chosen by Adaboost to build a cascaded classifier. The speed of detection is about 70 times quicker than the original HOG feature with the comparable accuracy. Zhang et al. [4] and Pedersoli et al. [5] applied the HOG feature at multiple resolutions in which both the accuracy and efficiency are improved. Wojek et al. [6] used the parallel technique to implement the HOG on GPU, upon which a real-time pedestrian detection system is built.

To overcome occlusion, Dalal [7] used the part-based pedestrian detection by dividing the body into head-shoulder, torso and legs. The HOG features of the three parts are used to build the classifier together. In such a case, a missing part will not result in the detection failure, and therefore the occlusion can be tackled to a certain degree. Wang et al. [8] analyzed the response

of each block to find the occluded blocks by decomposing the output of SVM classifier, and then local binary feature was further include to be a complement of HOG.

To enrich the description of HOG, Dalal [7] combined the HOG feature with the histogram feature based on movement and appearance for pedestrian detection in videos. Schwartz et al. [9] combined the HOG feature with color, texture and edges to create the considerable diversity descriptor. They also used the Partial Least Squares (PLS) to reduced the dimension of descriptor, which has shown the lower error rate than the original HOG descriptor. Watanabe et al. [10] proposed a co-occurrence Histograms of Oriented Gradients, to adapt the contextual statistics into the HOG feature.

Although many improvements of the HOG feature have been proposed, there is few work taking deep insights into the effects of orientation and magnitude of the gradient, which is the fundamental step in building the HOG descriptor. In this work, we will show that, instead of spacing the orientation bins evenly, the gradients can be projected by a random matrix and a regular polyhedron respectively, which surprisingly outperforms the original HOG descriptor with the similar dimension. Our work indicates that the gradient magnitude plays a crucial role in HOG. Besides, dividing the blocks into cells can improve the descriptor accuracy, as it describes more details.

In the HOG feature, the blocks are determined in the fixed size and positions; Based on the rigid global gradient, HOG is hard to capture the contour information of pedestrian precisely, since the training examples may not be aligned well. Inspired by the compressive sensing theory and other variants of HOG [11][12], especially its application on tracking [13], an improvement encoding scheme based on the integral histogram based improvement is proposed. Rather than using the global gradient, the new descriptor is based on the differences among local blocks. In each iteration, two blocks are chosen in the random positions and the difference between the two blocks is measured. In the end, differences all around iterations are combined to form the final descriptor. The experiments show that the integral histogram not only reduces the dimension significantly, but also outperforms the original HOG feature.

As discussed, the random projector and difference of local blocks can improve the performance respectively. Therefore, an Integral Histogram with Random Projection (IHRP) descriptor are proposed. In the novel descriptor, the gradients are projected by a random matrix and then the integral histogram is calculated as the feature. The experiments show that the proposed descriptor outperforms the HOG with a low dimension and a fast speed. It can also perform well in the case of partial occlusions.

The rest of the paper is organized as follows: Related works on pedestrian detection are presented in Section 2. Section 3 describes the algorithms of random projection and the integral histogram. The experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

Related Works

Pedestrian detection is still an active research field in computer vision, new benchmarks [14][15] and novel detection methods [16][17][18] are proposed recently. There also exist some review papers [19][20][21] on pedestrian detection. Generally speaking, a full pedestrian detection system for intelligent vehicle includes the six following module [19]: pre-processing, candidates generation, classification, verification and refinement, tracking and application. Among these modules, classification is the core component, which involves vision and machine learning. Broadly speaking, classification consists of feature extraction and classifier. The former deals with descriptors, while the latter provides algorithms to learn from labeled samples based on the descriptors.

A pedestrian classifier decides if a candidate window contains a pedestrian or not. Popular supervised classifiers used in pedestrian detection are different types of neural networks (e.g. multi-layer perceptron, convolution neural network, deep learning, etc.), AdaBoost variants (e.g. Real AdaBoost, LogitBoost, GentleBoost, MPLBoost, etc.), support vector machine (SVM; linear, kernelized [22], latent [23], etc.), Random Forest [24], et al. In this section, we only briefly review the feature extraction of pedestrian detection.

Roughly speaking, feature for pedestrian detection can be divided into three categories: low-level feature, high-level feature, learning-based feature.

The commonly used low-level feature are color and texture. For example, Oren et al. [25] introduce the basic Haar features, which is extended by Viola and Jones [26] and called Haar-like feature set. Haar feature, AdaBoost, and integral image are three key intergradient for the first real-time face detection system. Levi et al. [27] introduce the Edge Orientation Histogram (EOH) features to compensate for Haar-likes ones, which only rely on gradient information. Dalal and Triggs [7] proposed another gradient-based features, called Histogram of Oriented Gradient, which is the most popular ones nowadays. Local Binary Pattern (LBP) proposed by Ojala et al. [28], is very successful in texture classification. Wang et al. [29] combined the uniform LBP and HOG features, where the LBP histograms are computed based on the cell grid used for HOG features. The computational cost of LBP and HOG is high, so Hinterstoisser et al. [30] proposed the dominant orientation template features, which can be used as an alternative to HOG with similar detection performance, but being about two orders of magnitude faster. People are with different kind of clothes, so color is not distinctive for pedestrian detection. But color self-similarity between different blocks in the scanning window is somewhat powerful.

Here the high-level feature denotes the high order statistical information of low-level feature. Watanabe et al. [31] proposed Co-occurrence HOG (CoHOG), which can be seen as second order of HOG feature. A major problem of CoHoG is of huge dimension, so Hiromoto et al. [32] divide the CoHOG feature into many smaller ones, and a rejection-cascade classification framework is constructed based on the smaller vectors. Tuzel et al. [33] proposed a fast region descriptors for object detection, which is called covariance features. Actually covariance matrices form a Riemannian manifold, not the commonly used vector space. So Tuzel et al. [33] use the LogitBoost classifier with weak classifiers operating in the tangent space of the manifold.

Learning-based feature, also called data-driven feature or feature mining, denotes the algorithms that learn the descriptors as part of the process of learning the pedestrian classifier, which is the realm of artificial neural networks. Enzweiler et al. [34] evaluated the pedestrian classification performance based on a feed-forward neural network with local receptive field (LRF) in the hidden layer. Sermanet et al. [35] proposed a method using unsupervised convolutional sparse auto-encoders to pre-train features, and end-to-end supervised learning to train the a pedestrian classifier.

The Improvement of HOG

In the HOG [1] feature, the orientations are divided into 9 uniform bins, and the histograms of gradient are calculated by weighting the magnitude into corresponding bins according to the orientation. It can be seen as the uniform projecting. Besides, the blocks are overlapped and a global feature is generated finally.

3.1. Random Projection

In this paper, the random projectors are considered firstly. A random matrix $M_{g \times k}$ composed of random numbers is used for projecting the gradient magnitude. It means the gradient directions are randomly divided into uneven bins.

The random matrix is generalized as follows: Firstly an initial matrix $M_{g \times k}$ is generalized, where g represents the number of gradient components and k represents the projection dimension in a block, with $k \ll M$. Each element in the matrix is set to a random number (between -128 and 128), and then a normalization according to column is performed.

Oreifej et.al [11] used the gradient magnitude as an extra dimension and formed a proposed 4D normal. Therefore they can select different bins from different gradient magnitude and achieved the state of the art in recognition accuracy. Inspired by this, we define the 3D normal of the pixel as

$$N = \frac{1}{\sqrt{dx^2 + dy^2 + 1}} \begin{pmatrix} dx \\ dy \\ -1 \end{pmatrix}, \tag{1}$$

where $\begin{pmatrix} dx \\ dy \end{pmatrix}$ is the gradient of the pixel.

The 3D normal can also be projected by the random matrix with $g = 3$. The process of random projector is shown as Algorithm 1.

Algorithm 1: Random Projection Feature

Input: Image $I(x, y)$, the number of gradient components g , the projection dimension in a block k .

Output: Random Project Feature F

1. Generate a matrix $M_{g \times k}$. Each element of the matrix is formed by a random number (from -128 to 128). $g \in \{2, 3\}$, $k = 5, 10, 15 \dots$

2. Calculate each pixel's gradient $\begin{pmatrix} dx \\ dy \end{pmatrix}$, and get the normal vector

$$N(x, y) = \begin{cases} \frac{1}{\sqrt{dx^2 + dy^2}} \begin{pmatrix} dx \\ dy \end{pmatrix} & \text{if } G = 2 \\ \frac{1}{\sqrt{dx^2 + dy^2 + 1}} \begin{pmatrix} dx \\ dy \\ -1 \end{pmatrix} & \text{if } G = 3 \end{cases}$$

3. Multiply normal vector $N_{1 \times G}(x, y)$ and each column of the matrix $M_{g \times k}$, get each pixel's projected value $P_{1 \times k} = (p_1, p_2, \dots, p_i, \dots, p_k)$, if $p_i < 0$ then $p_i = 0$.

4. Refer to [1], vote all pixels' projected value into uneven bin_i ($i = 1, 2, \dots, k$) for each block, and do normalization.

5. Combined each block's bins as the final random projection feature F .

Besides, the regular geometric objects called polyhedrons are considered to quantize the 3D space (seen as 3.1.3).

3.2. Integral Histogram

In traditional HOG, the location and the size of the block are fixed. The template needs to be regenerated in the case of occlusions, global shift and some large offsets of parts. In order to overcome this over-fitting problem, the random blocks of random size and position are selected through several iterations. In this way, the inflexible block sampling of template can be avoided, making features more robust to the above changes.

Algorithm 2: Random Integral Haar Feature

Input: Image $I(x, y)$, a matrix $M_{x \times y}$, Maximum Iterations n , Number of bins m .

Output: Random Integral Haar Feature F .

1. Calculate gradient of each pixel, where $G(x, y)$ as gradient magnitude, $\Theta(x, y)$ as gradient direction.

2. Build matrix $M_{x \times y}$ for each interval, vote each pixel's gradient magnitude according to $\Theta(x, y)$.

3. Calculate each matrix $M_{x \times y}$'s integral graph, denoted by $M_{(x+1) \times (y+1)}$.

4. **for** $i = 1, 2, \dots, n$

for $j = 1, 2, \dots, m$ **do**

 Randomly generated two rectangles A and B , and calculate the sum of gradients for each rectangle respectively, denoted as sum_A, sum_B ,

If $sum_B < sum_A$

$sum_A = sum_B$;

 put sum_A into F .

end for

end for

Normalize for each $M_{(x+1) \times (y+1)}$.

Return final random integral haar feature F .

}

In addition, the differences between local blocks are more powerful than the global block to describe the information in object recognition. It can better describe the difference between the histograms. There are many ways to describe the difference between two local blocks. Empirically, the performance using $\text{MIN}(A, B)$ is superior to the others (A and B represents two different histogram respectively). Since the minimum of two histograms can better capture the main information of two blocks. The process of Integral Histogram is given as Algorithm 2.

The Experimental Results

In this section, the random projection and integral histogram are evaluated respectively, and the proposed IHRP descriptor is compared with the HOG. The evaluation dataset is the entire INRIA dataset [1]. The size of each sample is 64×128 . 2,416 of the images were selected as the positive training examples while 12,180 as the negative ones. All algorithms are tested with 1,126 positive examples and 30,000 negative examples.

4.1. Random Projection

Firstly, the random matrix projectors with different dimension (k) are evaluated. The influence of cells is also considered. Inspired by [11] and [12], a 3D normal projection by polyhedron and random matrix are evaluated.

4.1.1 The influence of the dimension of random matrix. The ROC curves of selected k on INRIA dataset [1] are showed as Fig 1, where k is dimension of random matrix. It shows the random projection outperforms the HOG feature with less dimensions. As the value of k increases, the performance will be improved significantly, with a turning point at $k = 10$. Finally the number of random numbers is set to 10, which is similar to the 9 bins as the HOG suggested. But our algorithm is much better in accuracy.

4.1.2 The influence of cells. The performances of whether the cells are used in the same dimension are showed as Fig 2. The 'cell+block $k = 5$ ' means that the gradient orientations are projected into 5 histograms, and the dimension is $4 \times 5 \times 105 = 2,100$ which is the same with 'block $k = 20$ ' (20×105). It's obvious that the cells can improve the accuracy dramatically.

4.1.3 3D normals. In this section, the 3D normals projected by polyhedron and random matrix are showed as Fig 3. In particular, the icosahedron with 12 vertices is used and the vertices are given as: $(0, \pm 1, \pm \varphi)$, $(\pm 1, \pm \varphi, 0)$, and $(\pm \varphi, 0, \pm 1)$, where $1/\varphi = 2/(1 + \sqrt{5})$ is a constant

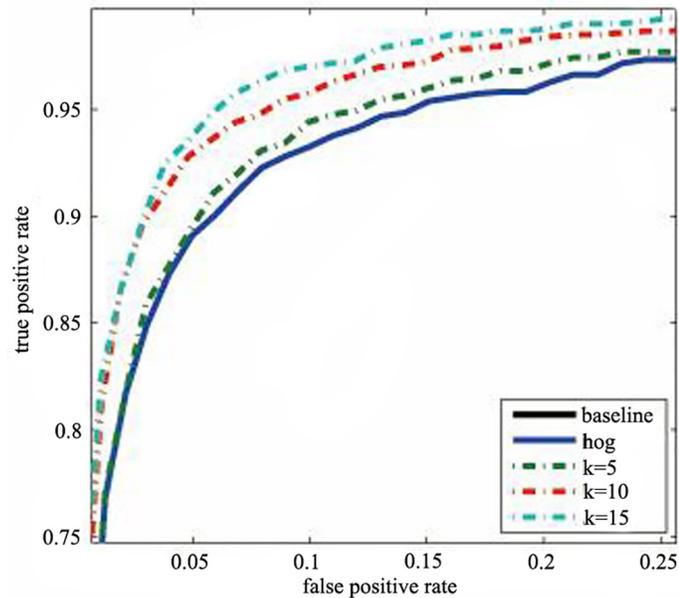


Fig 1. The performance of selected k .

doi:10.1371/journal.pone.0142820.g001

called the golden ratio. The quantization of 3D space by random projector is shown as Fig 3 ($k = 10$ 3d, $k = 15$ 3d). It shows that the 2D gradient combined with $k = 15$ perform best. The second one is the 3D normals with $k = 10$. The 3D normals projected by polyhedron (icosahedrons 3d) perform worse than the random projector, but still outperforms HOG feature. It shows that the random projectors with the proper dimension can achieved the better accuracy whether using 2D gradient or 3D normals.

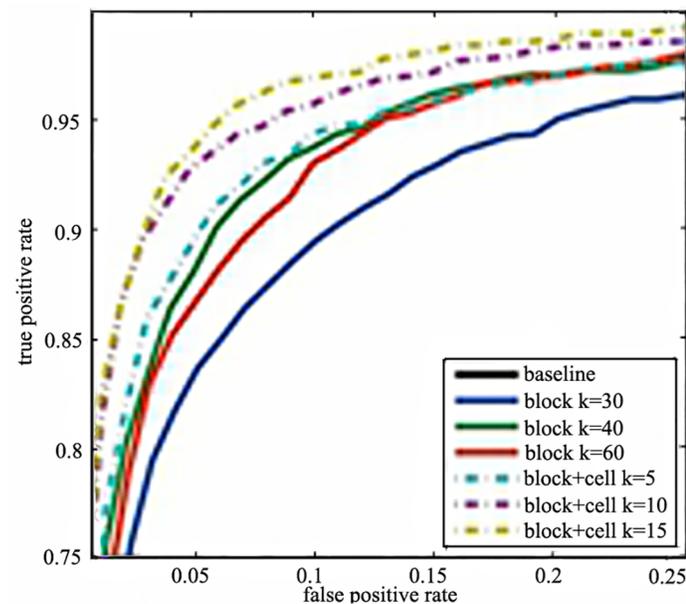


Fig 2. The performance of whether using cells on INRIA dataset.

doi:10.1371/journal.pone.0142820.g002

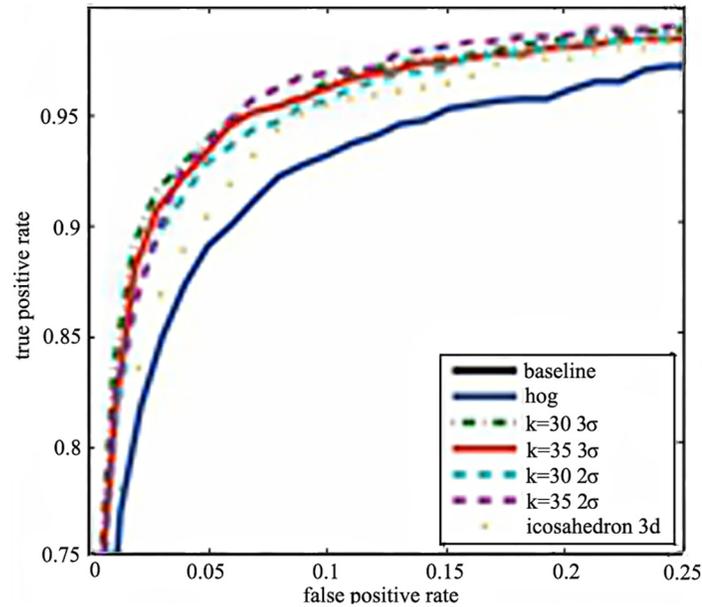


Fig 3. The performance of 3D normals projected by polyhedron and random matrix.

doi:10.1371/journal.pone.0142820.g003

4.2. Integral Histogram

The performances of the selected number (n) of the random block pairs with cells are showed as Fig 4. In general, the performances will be better when n increase. The performance is similar to the HOG feature when $n = 25$. And it grows rapidly when n increases from 25 to 50 but not obviously when n is larger than 50. So n is selected as 50 and 100 respectively.

4.3. Integral Histogram with Random projector

The combination of integral histogram with random projector is also considered. In particular, the cells are used to better describe the detail. The performances are showed in Fig 5. It shows that our

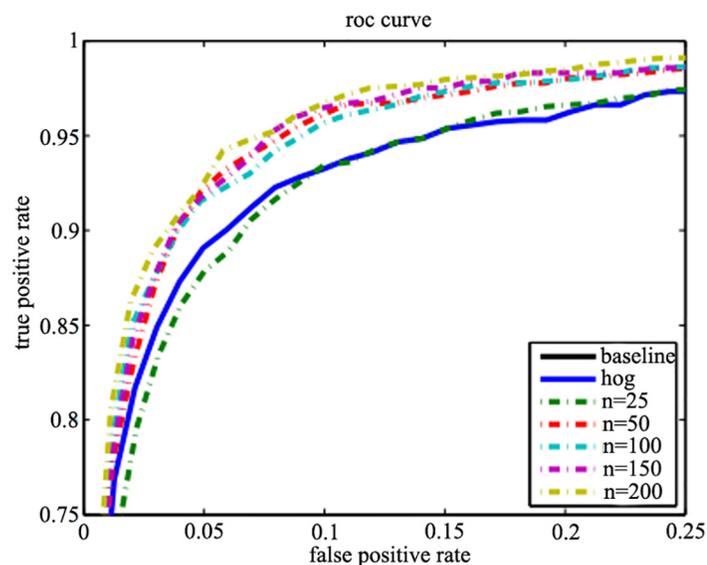


Fig 4. The performance of integral histogram with different number of block pairs n .

doi:10.1371/journal.pone.0142820.g004

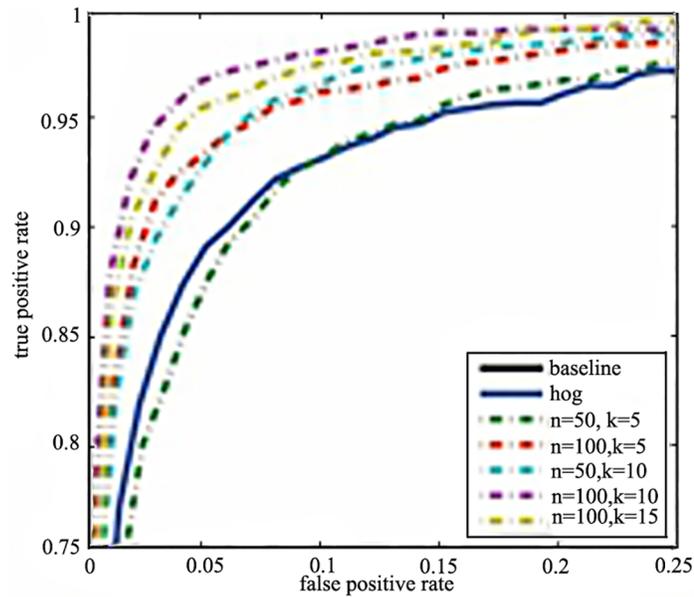


Fig 5. The performance of the proposed integral histogram with random projector (IHRP) with different n and k .

doi:10.1371/journal.pone.0142820.g005

proposed descriptor outperforms the HOG with a low dimension. It performs equally strong to the HOG descriptor with a quarter of HOG’s dimension when $n = 50$ and $k = 5$ ($50 * 5 * 4 = 1000$ dimensions). As the dimension increases to half of HOG when $n = 50$ and $k = 10$ ($50 * 10 * 4 = 2000$ dimensions), it outperforms much than the HOG at 2 times speed. Our method reaches the best performance when $n = 100$ and $k = 10$ ($100 * 10 * 4 = 4000$ dimensions). In the best case, the number of dimensions is similar to that of the HOG, but the accuracy and the efficiency is much better.

In order to evaluate the performance in the situation of occlusions, each example in the test set is embedded with the blocks of size $16 * 16$ in random position. The number of blocks can be one, two or three. As it shows in Fig 6, the proposed method outperforms HOG much better

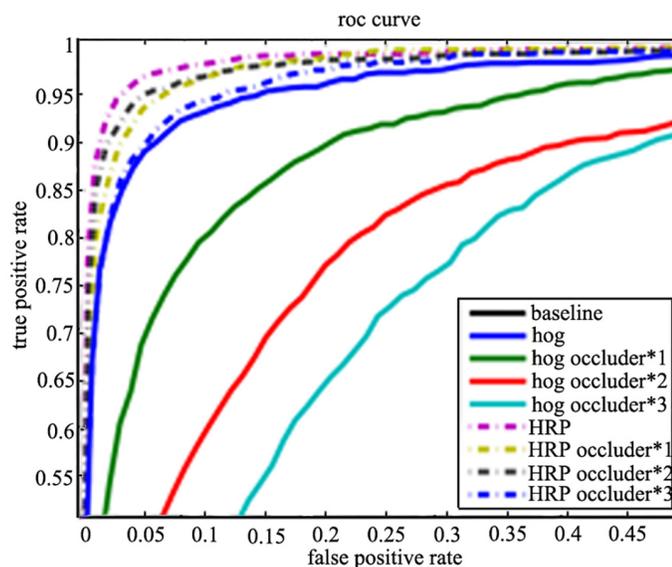


Fig 6. The performance of the proposed IHRP in the case of occlusions.

doi:10.1371/journal.pone.0142820.g006

when $n = 100$ and $k = 10$. Besides, as the number of local blocks pairs grows, the performances will be improved further.

Conclusions

In this paper, the orientation and magnitude of gradient are studied and an integral histogram is proposed. The experiments show that both the random projector and integral histogram outperform the HOG feature obviously. Especially, the orientations can be randomly projected by random matrix and polyhedron. Besides, the local difference descriptor based on the integral histogram is more flexible and robust to the global HOG feature. Finally, the two ideas are combined to form the new feature IHRP, which can outperform the HOG feature with lower dimension and higher speed even in the case of partial occlusions.

In the future, a deep study of local blocks will be investigated, such like using AdaBoost to choose the more distinctive blocks for the templates.

Author Contributions

Performed the experiments: CHL JKL. Analyzed the data: CHL JKL. Contributed reagents/materials/analysis tools: CHL JKL. Wrote the paper: CHL JKL.

References

1. Dalal N, Triggs B. Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition*, 2005, 1:886–893.
2. Zhu Q, Yeh MC, Cheng KT, Avidan S. Fast human detection using a cascade of histograms of oriented gradients. *Computer Vision and Pattern Recognition*, 2006, 2:1491–1498.
3. Porikli F. Integral histogram: A fast way to extract histograms in cartesian spaces. *Computer Vision and Pattern Recognition*, 2005, 1:829–836.
4. Zhang W, Zelinsky G, Samaras D. Real-time accurate object detection using multiple resolutions, *International Conference on Computer Vision*, 2007, 1–8.
5. Pedersoli M, Gonzalez J, Villanueva J. High-speed human detection using a multiresolution cascade of histograms of oriented gradients. In *Proc. of the 4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, 2009, 48–55.
6. Wojek C, Schiele B. A performance evaluation of single and multi-feature people detection. In *Proc. of DAGM*, 2008, 82–91.
7. Dalal N. Finding people in images and videos. INRIA Ph.D thesis, 2006.
8. Wang X, Han X, Yan S. A HOG-LBP human detector with partial occlusion handling. In *Proc. of ICCV*, 2009.
9. Schwartz W, Kembhavi A, Harwood D. Human detection using partial least squares analysis. In *Proc. of ICCV [C]*, 2009.
10. Wantanbe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for pedestrian detection. In *Proc. of The 3PrdPacific-Rim Symposium on Image and Video Technology*, 2009, 37–47.
11. Oreifej O, Liu Z. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
12. Alexander K, Marszalek M. A spatio-temporal descriptor based on 3D-gradients, *BMVC*, 2008.
13. Zhang KH, Zhang L, Yang MH. Real-time compressive tracking, *ECCV*, 2012, 864–877.
14. Hall D, Perona P. Fine-grained classification of pedestrians in video benchmark and state-of-the-art, *CVPR*, 2015.
15. Kim S, Kweon Y. Multispectral pedestrian detection benchmark dataset and baseline, *CVPR*, 2015.
16. Yang B, Yan J, Lei Z, Li SZ. Convolutional channel features for pedestrian face and edge detection, *arXiv:1504.07339*, 2015.
17. Hattori H, Boddeti VN, Kitani K. Learning scene-specific pedestrian detectors without real data, *CVPR*, 2015.
18. Zhang S, Benenson R, Schiele B. Filtered channel features for pedestrian detection, *CVPR*, 2015.

19. Geronimo D, Lopez AM. Vision-based pedestrian protection systems for intelligent vehicles. 2014, Springer Briefs in Computer Science.
20. Benenson R, Omran M, Hosang J, Schiele B. Ten years of pedestrian detection, what have we learned? *ECCV*, 2014.
21. Dollar P, Wojek C, Schiele B, Perona P. Pedestrian detection: an evaluation of the state-of-the-art. *PAMI*, 2012, 34(4): 743–761.
22. Maji S, Berg A, Malik J. Classification using intersection kernel support vector machines is efficient. *CVPR*, 2008.
23. Felzenszwalb P, Girshick R, McAllester D, Ramanan D. Object detection with discriminatively trained part based models. *IEEE Trans. on Pattern Analysis and, Machine Intelligence*, 2010, 32(9):1627–1645.
24. Tang D, Liu Y, Kim T. Fast pedestrian detection by cascade random forest with dominant orientation. In: *British Machine Vision Conference*. Surrey, UK (2012)
25. Oren M, Papageorgiou C, Sinha P, Osuna E, Poggio T. Pedestrian detection using wavelet templates. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.
26. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *IEEE Conf. on Computer Vision and Pattern Recognition*. 2001.
27. Levi K, Weiss Y. Learning object detection from a small number of examples: the importance of good features. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.
28. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and, Machine Intelligence*, 2002, 24 (7):971–987.
29. Wang X, Han T, Yan S. An HOG-LBP human detector with partial occlusion handling. In: *Int. Conf. on Computer Vision*. Kyoto, 2009.
30. Hinterstoisser S, Lepetit V, Ilic S, Fua P, Navab N. Dominant orientation templates for real-time detection of texture-less objects. In: *IEEE Conf. on Computer Vision and Pattern Recognition*. 2010.
31. Watanabe T, Ito S, Yoki K. Co-occurrence histograms of oriented gradients for pedestrian detection. In: *Pacific-Rim Symposium on Image and Video Technology*. 2009.
32. Hiromoto M, Miyamoto R. Cascade classifier using divided cohog features for rapid pedestrian detection. In: *Int. Conf. on Computer Vision Systems*. 2009.
33. Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on Riemannian manifolds. *IEEE Trans. on Pattern Analysis and, Machine Intelligence*, 2008, 30(10):1–15.
34. Enzweiler M, Gavrila D. Monocular pedestrian detection: Survey and experiments. *IEEE Trans. on Pattern Analysis and, Machine Intelligence*, 2009, 31(12): 2179–2195.
35. Sermanet P, Kavukcuoglu K, Chintala S. Pedestrian detection with unsupervised multi-stage feature learning. *CVPR*, 2013.