LETTER

Efficient Pedestrian Detection Using Multi-Scale HOG Features with Low Computational Complexity

Soojin KIM[†], *Nonmember and* **Kyeongsoon CHO**[†a)], *Member*

SUMMARY In this paper, an efficient method to reduce computational complexity for pedestrian detection is presented. Since trilinear interpolation is not used, the amount of required operations for histogram of oriented gradient (HOG) feature calculation is significantly reduced. By calculating multi-scale HOG features with integral HOG in a two-stage approach, both high detection rate and speed are achieved in the proposed method.
key words: *pedestrian detection, interpolation, multi-scale HOG, integral HOG, two-stage approach*

1. Introduction

The most critical factors for pedestrian detection in many applications are high detection rate and real-time processing [1]–[3]. Such applications include intelligent vehicle, surveillance, and robotics. Histogram of oriented gradient (HOG) [4] is considered to be the most discriminative feature for pedestrian detection, and trilinear interpolation is one of the most effective techniques for using the HOG feature to improve detection rate. Trilinear interpolation is, however, a major bottleneck of detection speed due to its high computational complexity. Therefore, we present a new method for pedestrian detection without using trilinear interpolation to achieve good results in terms of both detection rate and speed. By calculating multi-scale cells and blocks of HOG feature with integral HOG and applying them in a two-stage manner, both detection rate and speed are improved in our method. The number of stages in the pedestrian detection system is determined by considering detection accuracy and speed. The computational complexity increases sharply as the number of stages increases. In our experiment, a two-stage detection system is considered because good detection accuracy can be achieved with manageable compuational complexity. Note that the proposed method can be extended to the pedestrian detection system with any number of stages.

2. HOG Feature Calculaion

HOG feature is presented in [4] for pedestrian detection, and Figure 1 shows the process of HOG feature calculation. Trilinear interpolation technique, applied at the second step of

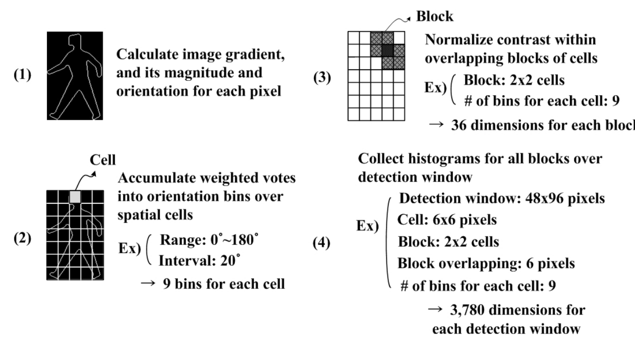

Fig. 1 HOG feature calculation.

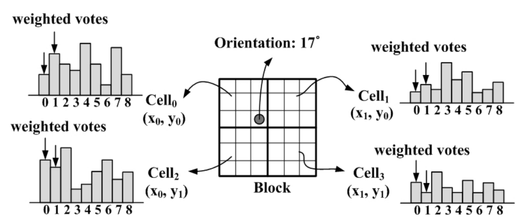

Fig. 2 Example of trilinear interpolation.

HOG feature calculation, is to interpolate weighted votes for gradient magnitude bilinearly between the neighboring bins in both orientation and position.

Figure 2 shows an example of trilinear interpolation when a block consists of 2x2 cells and orientation bins are evenly spaced over 0°~180° (9 bins). When a gradient orientation for a pixel in $Cell_0$ is 17°, the two nearest neighboring bins are determined as 0 and 1. In trilinear interpolation, therefore, weighted votes for gradient magnitude are distributed into eight surrounding bins (bin 0 and 1 for $Cell_0$~$Cell_3$) as shown in the figure, whereas the weighted votes are distributed into only two neighboring bins (bin 0 and 1 for $Cell_0$ only) in linear interpolation. Since only orientation is multiplied by the weight in linear interpolation, a total of two multiplications (one multiplication for each of the two neighboring bins) is required for each pixel. In trilinear interpolation, on the other hand, orientation and both x and y positions of a cell are all multiplied by the weight. Therefore, a total of 24 multiplications (three multiplications for each of the eight surrounding bins) is required for each pixel in trilinear interpolation.

Manuscript received August 28, 2013.
Manuscript revised October 19, 2013.
[†]The authors are with the Department of Electronics and Information Engineering, Hankuk University of Foreign Studies, Yongin-si, 449-791, Republic of Korea.
a) E-mail: kscho@hufs.ac.kr
DOI: 10.1587/transinf.E97.D.366

3. Proposed Method for Pedestrian Detection Using HOG Feature

We define several parameters of HOG feature as shown in Fig. 3, and the number of required multiplications for interpolation and L2-Hys normalization [4] in HOG feature calculation is presented in Table 1 (Gaussian function is assumed to be pre-calculated). Although trilinear interpolation is used to avoid aliasing for high detection rate, it degrades detection speed siginificantly since a large amount of operations is required as shown in Table 1. Since it is important to achieve both high detection rate and real-time processing in many applications, we propose a new method to accelerate pedestrian detection without notable loss of detection rate. Instead of using trilinear interpolation in HOG feature calculation, we use linear interpolation and utilize multi-scale cells and blocks of HOG feature in order to compensate for the loss of detection rate. By applying two-stage approach to speed up multi-scale HOG feature calculation and adopting integral HOG to avoid redundant computations, we achieved good results in terms of both detection rate and speed. In the conventional methods, the operations of linear or trilinear interpolation and normalization are required for every overlapping block. Therefore, the number of required multiplications per detection window is directly proportional to the number of overlapping blocks in the detection window ($B_x \times B_y$) as shown in Table 1. In the proposed method, on the other hand, the number of required multiplications in linear interpolation ($2\times S_x \times S_y$) is not affected by the number of overlapping blocks. The operation of linear interpolation is required only once at the first stage, and the results of linear interpolation are used to construct integral HOG. Then, the operation of normalization is required for the overlapping blocks. Note that the integral HOG constructed at the first stage is shared in the second stage. Therefore, the required operation for the second stage is a block normalization only since the sizes of cell and block of HOG feature for the second stage are different from those of the first stage. If higher detection accuracy is required, more stages can be added into the detection system with the expense of increased computational efforts.

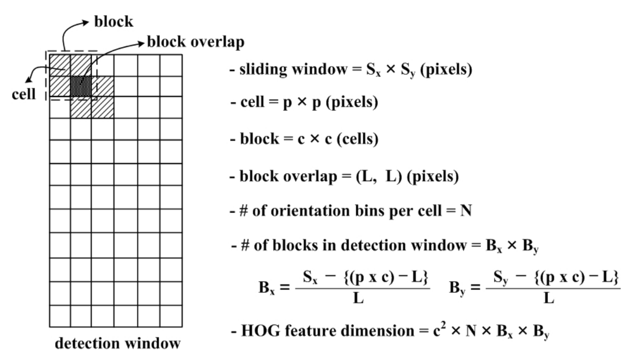

Fig. 3 Parameter definitions for HOG feature.

Table 1 The number of required multiplications per detection window.

Inter-polation	Gaussian weighting	# of multiplications
Linear	O	$\{(4\times p^2\times c^2)+(2\times c^2\times N)+1\}\times B_x\times B_y$
	X	$\{(2\times p^2\times c^2)+(2\times c^2\times N)+1\}\times B_x\times B_y$
Trilinear	O	$[\{32\times p^2\times(c-1)^2\}+\{32\times p^2\times(c-1)\}$ $+(8\times p^2)+(2\times c^2\times N)+1]\times B_x\times B_y$
	X	$[\{24\times p^2\times(c-1)^2\}+\{24\times p^2\times(c-1)\}$ $+(6\times p^2)+(2\times c^2\times N)+1]\times B_x\times B_y$
Linear (proposed)	X	1^{st}: $(2\times S_x\times S_y)+[\{(2\times c^2\times N)+1\}\times B_x\times B_y]$
	X	2^{nd}: $\{(2\times c^2\times N)+1\}\times B_x\times B_y$

As mentioned in [4], detection rate is improved when multiple blocks with different scale cells and blocks of HOG feature are used together. However, the computational cost is greatly increased as the number of types of HOG feature is increased. Therefore, we use two different scales of cells and blocks of HOG feature in order to improve detection rate, and utilized them in two-stage manner in order to reduce computational cost. In our method, HOG feature calculated with p=8 and c=2 (L=8) is used in the first stage and the feature calculated with p=6 and c=3 (L=6) is used in the second stage as shown in Fig. 4.

Two-stage approach has been used to speed up pedestrian detection, in which different types of feature are used for each stage. In this case, however, the amount of operations and computational complexity are significantly increased since the operations for calculating two different features are required. In the proposed method, therefore, we use a single type of feature, i.e., HOG feature only, for both stages. HOG feature with p=8 and c=2, default detector in [4], is used in the first stage in order to discard negative input quickly since the feature has relatively low computational cost. In the second stage to verify the pedestrian candidate precisely, we use HOG feature with p=6 and c=3 since it performs best [4]. Since negative input can be discarded quickly and positive input can be verified twice as shown in Fig. 4, both high detection rate and speed can be achieved in the proposed method. Note that the additional operation for calculating different type of feature is not required in our method since the same type of feature is used in both stages.

In order to further accelerate pedestrian detection by avoiding redundant computations for multi-scale HOG feature calculation, we adopted integral image technique [5] to our method as shown in Fig. 5. Since integral HOG is constructed by pre-accumulating weighted votes for each orientation bin, multi-scale cells and blocks of HOG feature

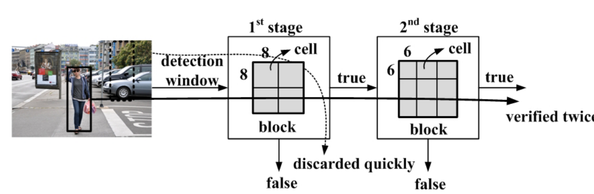

Fig. 4 Two-stage approach for proposed pedetrain detection.

Table 2 Comparison results in terms of detection accuracy and the number of required multiplications per 48x96-pixel detection window.

Inter-polation	Gaussian weighting	Cell size (pxp)	Block size (cxc)	Integral HOG	Detection rate @10^{-3}	Detection rate @10^{-4}	# of multiplications	Detection rate @10^{-3}	Detection rate @10^{-4}	# of multiplications Min.	# of multiplications Max.
Linear	X	6x6	2x2	X	86%	68%	37,905	−7.5%	−13%	x1.41	x2.86
Linear	X	8x8	2x2	X	86%	74%	32,175	−7.5%	−7%	x1.19	x2.43
Linear	X	6x6	3x3	X	88.5%	73%	68,124	−5%	−8%	x2.53	x5.15
Linear	O	6x6	2x2	X	86%	65%	68,145	−7.5%	−16%	x2.53	x5.15
Linear	O	8x8	2x2	X	88.8%	77%	60,335	−4.7%	−4%	x2.24	x4.56
Linear	O	6x6	3x3	X	88.8%	74%	122,556	−4.7%	−7%	x4.55	x9.26
Trilinear	X	6x6	2x2	X	89%	75%	211,785	−4.5%	−6%	x7.87	x16.01
Trilinear	X	8x8	2x2	X	90%	76%	194,095	−3.5%	−5%	x7.21	x14.67
Trilinear	X	6x6	3x3	X	92%	83%	467,292	−1.5%	+2%	x17.36	x35.32
Trilinear	O	6x6	2x2	X	89%	78%	279,825	−4.5%	−3%	x10.39	x21.15
Trilinear	O	8x8	2x2	X	91%	77%	257,455	−2.5%	−4%	x9.56	x19.46
Trilinear	O	6x6	3x3	X	93%	85.1%	618,492	−0.5%	+4.1%	x22.98	x46.75
Linear (proposed)	**X**	**1^{st}: 8x8 2^{nd}: 6x6**	**1^{st}: 2x2 2^{nd}: 3x3**	**O**	**93.5%**	**81%**	**1^{st} only: 13,231 1^{st}+2^{nd}: 26,923**	-	-	-	-

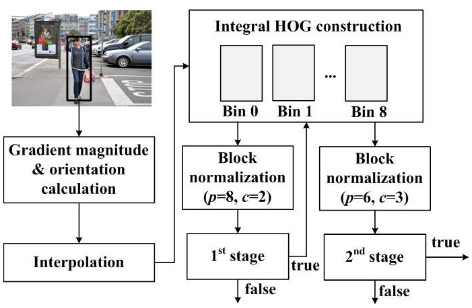

Fig. 5 Multi-scale HOG feature calculation with intergal HOG.

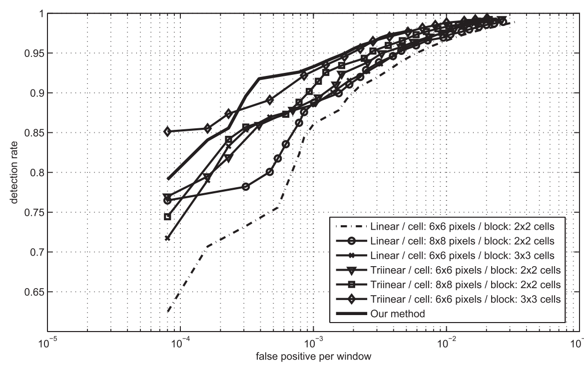

Fig. 6 Comparison of detection rate for several cases.

are computed quickly from integral HOG. By using integral HOG to compute multi-scale HOG features, we reduced the amount of required operations significantly since the operations for calculating gradient magnitudes and orientations, interpolating weighted votes for gradient magnitude, and accumulating the interpolated votes into the orientation bins are all shared in both stages. The additional operation for the second stage is a block normalization only.

4. Experimental Results

We have conducted several experiments in order to demonstrate that our method is efficient in terms of both detection accuracy and speed. In order to evaluate detection quality, we tested our detector on Daimler Pedestrian dataset [6] using linear support vector machine (SVM) [7]. 5,000 positive and 5,000 negative samples are used to train the detectors and 10,000 positive and 12,870 negative samples are randomly selected to test them. As shown in Fig. 6 and Table 2, the detection rate of the proposed detector is about 4% lower than the detector (p=6 and c=3 with Gaussian weighting and trilinear interpolation) at 10^{-4} false positive per window (FPPW). However, our detector achieves the best performance at 10^{-3} FPPW and 3∼16% higher detection rate than others at 10^{-4} FPPW. By utilizing multi-scale cells and blocks of HOG feature and applying them in a two-stage manner, detection rate is kept even though trilinear interpolation is discarded in the proposed method. Since our detector is based on two-stage approach, the number of required multiplications depends on what kind of object is contained in detection window. If the current detection window contains a pedestrian, the first-stage detector determines it as a positive candidate and the detection proceeds to the second stage in order to verify the candidate once more. Otherwise, the detection process is finished right after the first stage. Note that most of detection windows contain backgrounds rather than pedestrians. Therefore, in most cases only 13,231 multiplications are required for our detector. Although detection accuracy of the detector (p=6 and c=3 with Gaussian weighting and trilinear interpolation) is a little higher than ours at 10^{-4} FPPW, the number of required multiplications for interpolation and block normalization is 47 times larger when detection window contains a background and 23 times larger when detection window contains a pedestrian.

By using variable-size blocks with integral HOG and applying them in a cascade-of-rejectors approach, [8] also achieves fast detection time and accuracy comparable to [4].

The cascade in [8] consists of a total of 30 levels and several HOG blocks with variable sizes are used in each level. More than 770 variable-size blocks are used in 30 levels, and four blocks are used in the first level. Therefore, only four blocks are used in the best case when the current detection window contains a background and the first-level detector determines it as a negative candidate (true negative). In the worst case when the detection window contains a background but the detector determines it as a positive candidate (false positive) or when the detector recognizes a pedestrian correctly (true positive), more than 770 HOG blocks are required in [8] while 257 blocks are required in our method. Therefore, the worst-case performance of our method is much better than that of [8]. Note that the worst-case performance is more important than the best-case performance in many cases. For example, when the pedestrian detection system is implemented in hardware, many design decisions are made on the base of the worst-case performance. It is hard to compare the average performance distinctly since detection rate depends on what kind of input image is used in the evaluation of detection rate. Therefore, we considered both the best-case and worst-case performances. We found that our method is faster than [8] when more than 16.6% detection windows in each image are determined as positive candidates (false positive and true positive).

5. Conclusion

We considered multi-scale cells and blocks of HOG feature for pedestrian detection in order to improve detection rate. As a result, we could achieve a high detection rate even though Gaussian weighting and trilinear interpolation are not used in the proposed method. Furthermore, by calculating mutli-scale HOG features with integral HOG and applying them to a two-stage approach, detection speed is also improved since the amount of required operations for HOG feature calculation is significantly reduced. Therefore, the proposed method can be utilized in many applications such as intelligent vehicle, surveillance, and robotics in which both high detection accuracy and fast detection speed are strongly required.

Acknowledgements

This work was supported by Ministry of Trade, Industry and Energy (MOTIE) and IDEC Platform Center (IPC).

References

[1] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.4, pp.743–761, April 2012.
[2] D. Geronimo, A.M. Lopez, A.D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.7, pp.1239–1258, July 2010.
[3] Z. Yang, L. Weiming, M. Chen, and L. Zilong, "Pedestrian detection in complex scene using full binary tree classifiers based on locally assembled binary Haar-like features," Proc. 9th World Congress on Intelligent Control and Automation, Taipei, Taiwan, pp.1180–1184, June 2011.
[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition, San Diego, USA, pp.886–893, June 2005.
[5] P. Viloa and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Hawaii, USA, pp.511–518, Dec. 2001.
[6] M. Enzweiler and D.M. Gavrila, "Monocular pedestrian detection: Survey and experiments," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.12, pp.2179–2195, Dec. 2009.
[7] V.N. Vapnik, Statistical Learning Theory, Wiley & Sons, USA, 1998.
[8] Q. Zhu, S. Avidan, M.C. Yeh, and K.T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, New York, USA, pp.1491–1498, June 2006.