HSPOG: An Optimized Target Recognition Method Based on Histogram of Spatial Pyramid Oriented Gradients

Shaojun Guo  
*National Innovation of Defense Technology, Academy of Military Sciences PLA China, Beijing 100071, China.*

Feng Liu  
*National Innovation of Defense Technology, Academy of Military Sciences PLA China, Beijing 100071, China.*

Xiaohu Yuan  
*Department of Automation, Tsinghua University, Beijing 100084, China.*

Chunrong Zou  
*National Innovation of Defense Technology, Academy of Military Sciences PLA China, Beijing 100071, China.*

Li Chen  
*National Innovation of Defense Technology, Academy of Military Sciences PLA China, Beijing 100071, China.*

See next page for additional authors

Follow this and additional works at: [https://dc.tsinghuajournals.com/tsinghua-science-and-technology](https://dc.tsinghuajournals.com/tsinghua-science-and-technology)

**Recommended Citation**  
Shaojun Guo, Feng Liu, Xiaohu Yuan, Chunrong Zou, Li Chen, Tongsheng Shen. HSPOG: An Optimized Target Recognition Method Based on Histogram of Spatial Pyramid Oriented Gradients. Tsinghua Science and Technology 2021, 26(4): 475-483.

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in Tsinghua Science and Technology by an authorized editor of Tsinghua University Press: Journals Publishing.
HSPOG: An Optimized Target Recognition Method Based on Histogram of Spatial Pyramid Oriented Gradients

Authors
Shaojun Guo, Feng Liu, Xiaohu Yuan, Chunrong Zou, Li Chen, and Tongsheng Shen

This research article is available in Tsinghua Science and Technology: https://dc.tsinghuajournals.com/tsinghua-science-and-technology/vol26/iss4/4
HSPOG: An Optimized Target Recognition Method Based on Histogram of Spatial Pyramid Oriented Gradients

Shaojun Guo, Feng Liu, Xiaohu Yuan, Chunrong Zou, Li Chen, and Tongsheng Shen*

Abstract: The Histograms of Oriented Gradients (HOG) can produce good results in an image target recognition mission, but it requires the same size of the target images for classification of inputs. In response to this shortcoming, this paper performs spatial pyramid segmentation on target images of any size, gets the pixel size of each image block dynamically, and further calculates and normalizes the gradient of the oriented feature of each block region in each image layer. The new feature is called the Histogram of Spatial Pyramid Oriented Gradients (HSPOG). This approach can obtain stable vectors for images of any size, and increase the target detection rate in the image recognition process significantly. Finally, the article verifies the algorithm using VOC2012 image data and compares the effect of HOG.

Key words: Histograms of Oriented Gradients (HOG); Histogram of Spatial Pyramid Oriented Gradients (HSPOG); object recognition; spatial pyramid segmentation

1 Introduction

Histograms of Oriented Gradients (HOG)\[1\] is widely used in the field of motion recognition\[2\], computer vision\[3\], object classification\[4\], object detection\[5–7\], cultural recognition of language\[8\], and other recognition tasks. However, a fixed input image size/scale is required during the train storage and HOG detector testing\[9\], for an input image that does not meet the requirement needs to be transformed or cropped\[10–13\]. This may cause image distortion or loss of content\[14\], and affect target recognition accuracy. If HOG is employed to train the detector, the length of all vectors with input features should be the same. Therefore, when calculating a feature vector of HOG, fixed-size and fixed-pixel-size images are required. For two target images of different sizes/scales, if you want to determine whether they are the same type, you can compare the two images by regions, not by pixels, and the proportions of the regions in each image should be the same. HOG resizes an image to a fixed size and predefines the size of the cell to calculate the feature maps. Although the pixels and blocks of all input images have the same ratio, the target is always distorted, so the recognition accuracy may be affected.

HOG needs to adjust an image or sub-image to a fixed size/scale in practical applications. This paper introduces spatial pyramid scales to remove HOG’s fixed size limit. We divide an input image into $2^n$ cells on average, so all input images may have the same number of cells. Since the input images have different sizes/scales, cells of different images may have different sizes/scales. For each cell, we calculate a histogram of the oriented gradient, and finally, we calculate the Histogram of Spatial Pyramid Oriented Gradients (HSPOG) of each input image or sub-image, and all HSPOGs of the input images have the same length. In summary, this kind of optimization has two advantages: first, it can prevent image distortion; second, it can guarantee the length consistency of feature vectors and facilitate the
evaluation of the model for recognition. We arrange the article as follows: we discuss what HSPOG is in Section 2. In Section 3 an accelerating method for HSPOG is proposed. In Section 4, the effect of HOG and HSPOG is tested on dataset of VOC2012.

2 HSPOG

HSPOG is an improvement of HOG. For the same target in two images with different sizes, the two target image areas would be divided into \(2^n\) cells, and the relative characteristics between corresponding cells of the two areas would remain unchanged even if the cells have different sizes. The goal of HSPOG is to extract stable target features from different image targets and remove the unique image target size and ratio requirements of HOG. HOG always scales an image target into a fixed size of \(64 \times 128\) pixel and defines the cell size of the target as \(8 \times 8\) pixel. However, we think that the cell size should be dynamically determined by the cell number. So when calculating HSPOG features, we scale a target image dynamically without distorting (see Fig. 1). The scale ratio can be calculated by Eq. (1),

\[
\frac{\text{A\_scale}}{\text{min}(\text{row}, \text{col})}
\]

where A\_scale is the adjustment scale.

To get more target features of pyramid scales, the bigger the value of \(n\) in principle, the higher the accuracy of recognition. However, each increment of \(n\) by 1 will exponentially increase the computation amount, so we make \(2 \leq n \leq 5\) empirically. There is a feature vector for each scale, adding all the features together can form a long descriptor, which includes coarse features on large spatial scales and fine features on small spatial scales. HSPOG is described in Algorithm 1. The length of the small edge of an input image after zooming may be 64, if we specify the cell number along the small edge to be 8, the parameter \(n\) is 3. If the large edge of an image cannot be divided exactly by \(2^n\), we must fill the image by zeros before it can be divided exactly by \(2^n\), as shown in Fig. 2.

The flow chart of HSPOG is shown in Fig. 3. HSPOG can avoid image distortion and information loss. The size of an image cell is calculated dynamically, while the corresponding of HOG is constant. Dynamic cell determination ensures the ratio of the cell proportions the same. HSPOG contains multiple spatial pyramid scales for feature calculation, which enables HSPOG to integrate coarse features at large spatial scales and

### Algorithm 1 HSPOG

**Input:** Target-image of arbitrary size/scale

**Ensure:** Transform the input images \(I\) into \(I_{\text{trans}}\)

1. \(\text{ratio} = \frac{\text{A\_scale}}{\text{min}(\text{row}, \text{col})}\),
2. \(\text{width} = \text{round}(\text{width} \times \text{ratio})\),
3. \(\text{height} = \text{round}(\text{height} \times \text{ratio})\),
4. \(I_{\text{trans}} = \text{interpolation}(I, \text{ratio})\),
5. for \(n = n_1, n_2, \ldots, n_m\) do
6. \(\text{Divide } I_{\text{trans}} \text{ into } 2^n \times 2^n \text{ patches}\)
\[
p = \{p_1, p_2, \ldots, p_{2^n \times 2^n}\}
\]
7. Compute HOGs of \(p\),
\[
\text{feature} = \{\text{HOG}_{p_1}, \text{HOG}_{p_2}, \ldots, \text{HOG}_{p_n}\}
\]

**Output:** HSPOG features of images of arbitrary size/scale

---

![Fig. 1 Cropping or warping to fit a fixed size.](image1)

(a) Fill zeros to image

(b) Zoomed images by ratio

Fig. 2 Zooming method used in HSPOG. If the large edge of an image can not be divided by \(2^n\), it would be filled with zeros along the large edge until it can be divided by \(2^n\) exactly.
fine features at small spatial scales while preserving information integrity.

3 Accelerating HSPOG

The rapid detection and recognition of an image target is key to engineering applications. The article improves the algorithm of HSPOG to accelerate the calculation of features.

When calculating a feature map, HOG divides 180 degrees into 9 oriented blocks commonly, and the degrees between 180 and 360 can be considered as its negative situations. HSPOG not only divides 180 degrees into 9 blocks, but also divides 360 degrees into 18 blocks to obtain better oriented gradient features. In this process, trilinear interpolation\cite{15} is no longer used to calculate the degrees of its adjacent fans.

In HSPOG, we use bilinear interpolation instead of trilinear interpolation to calculate the weights of each pixel to neighboring cells, which is a fast calculating method\cite{16}. For example, the similarity between 85 and 80 degrees is 0.75, and the similarity between 100 and 85 degrees is 0.25. So, the best position of 85 degrees may be determined as the fifth block. We illustrate the processing in Fig. 4. Feature map of each cell will be normalized four times because it may be contained in four blocks. When all pixel features are added to HSPOG together for an image target, we will discard the features of boundary pixels because they cannot be normalized four times. For an image target, assuming $n_{\text{row}}$ is the number of cells in the row and $n_{\text{col}}$ is the number of cells in the column, the length of the image target HSPOG is $(n_{\text{row}} - 2) \times (n_{\text{col}} - 2) \times 32$. The 32 features in the previous equation contains 9 containers (see Fig. 4a), 18 containers (see Fig. 4b), 4 texture features, and 1 truncated feature, fast HSPOG is shown in Algorithm 2. The acceleration algorithm is an approximate simplification of the original algorithm, but the attenuation is negligible.

The spatial pyramid scales shown in Fig. 5 will be used when computing the HSPOG for an image. The scales play important roles in conventional methods, e.g., the Scale-Invariant Feature Transform (SIFT) vectors are also collected at multiple scales\cite{7,17}. So, we also compute the HSPOG at multiple scales, in the processing, scales include $4 \times 4$, $8 \times 8$, and $16 \times 16$. All of these pyramid scales make HSPOG more potent during the processing of object recognition missions.

Fig. 4 Two types of bins ($Z_{i}$ is the $i$-th bin).
Algorithm 2  Fast HSPOG  

Input:  Image patches $I_{\text{patch}}$

1.  for $j=1$ to $2^n \times 2^n$ do
2.      for $k=1$ to $9$ do
3.          Compute HOG and get 9 matrixes (size=$2^n \times 2^n$) for 9 bins,
4.      end
5.    for $i=1$ to $18$ do
6.          Compute histogram of gradients and get 18 matrixes (size=$2^n \times 2^n$) for 18 bins,
7.    end
8.  for $q=1$ to $4$ do
9.      Compute 4 texture feature compensation $2^n \times 2^n$ matrixes,
10.  end
11.  Compute the truncation feature map for $2^n \times 2^n$ patches.
12.  Normalize the $2^n \times 2^n \times 32$ matrix and abandon the edge feature cells.
13.  Return: HSPOG features of images of arbitrary size/scale

4 Experiment  

In this article, we test HOG and HSPOG on VOC2012 image set, which contains 18 kinds of objects. For each type of object, we use samples whose filenames are written in class_train.txt and Support Vector Machine (SVM) to train a classifier. The recognition rate and false alarm rate are employed as the HOG and HSPOG assessment parameters,

$$\text{Re\_rate} = \frac{\text{Target\_recog}}{\text{Target\_sum}} \times 100\%,$$

$$\text{False\_alarm} = \frac{F\_recog}{\text{Test\_sum}} \times 100\%$$  \hspace{1cm} (2)$$

where Re\_rate is the recognition rate, False\_alarm is the false alarm rate, Target\_recog is the number of rightly distinguished objects, Target\_sum is the total number of objects, $F\_recog$ is the number of wrongly distinguished objects, and Test\_sum is the total number of images used in the test.

The target recognition curves on VOC2012 are shown in Figs. 6–8. Higher the recognition rate and lower false-positive rate would make a better situation for recognition tasks.

From the experimental results (see Figs. 6–8), we can find that HSPOG has significantly improved over HOG, and the effect of the multi-classification on 18 targets is much better than HOG. The experimental curves show that HSPOG improves the recognition accuracy of image targets at four levers compared to HOG. At the first lever, HSPOG raises the accuracy by more than 20%, including the recognition of airplane, bird, bottle, bus, pottedplant, and chair. At the second lever, HSPOG raises the accuracy by more than 10%, including the recognition of boat, cow, dining table, and dog. At the third lever, HSPOG improves the accuracy by more than 5%, including the recognition of bicycle, car, horse, motorbike, and sheep. At the fourth lever HSPOG improves the accuracy by less than 5% or performs almost like HOG does, including the recognition of cat and person. From the four levers and object types, it is obvious that HSPOG can better recognize artificial targets than HOG. Nevertheless, although the recognition rate of targets (such as horse and cow) is better than HOG, there are always some false alarms in identifying certain targets. Increasing the number of angle bins and image cells may make a breakthrough. To test this hypothesis, a recognition operator is trained on ship targets. Some examples of the training ship samples are shown in Fig. 9. In the training process of the recognition operator, we take $2 \leq n \leq 7$, and decompose the 180 degrees into 18 bins and the 360 degrees into 18 bins to calculate the HSPOG. We also increase the image patches and make more bins to compute the feature maps for HOG.

In this paper, we use more than 1600 images of positive and negative sea-ships downloaded from the...
Shaojun Guo et al.: HSPOG: An Optimized Target Recognition Method Based on Histogram of . . .

Fig. 6 Curves of recognition rate and false alarm of experiments about (a) airplane, (b) bicycle, (c) bird, (d) boat, (e) bottle, and (f) bus.

Internet to obtain a classifier. Figure 10 is a comparison chart of the ship recognition rate, false alarm rate, and error rate between HSPOG and HOG.

From Fig. 10, it is easy to find that the ship target detection rate and accuracy of HSPOG are higher than that of HOG by 2.5% and 1.2%, respectively, and the false alarm rate and the error rate of HSPOG are lower than that of HOG by 1.3% and 1%, respectively. It finds that the performance of HSPOG is better than that of HOG while we evaluate them by accuracy rate, false alarm, and error rate. Under the condition of appropriately increasing the complexity of training the recognition operator, it can significantly improve the recognition rate of ship targets, and at the same time
Fig. 7 Curves recognition rate of and false alarm of experiments about (a) car, (b) chair, (c) cow, (d) dog, (e) horse, and (f) motobike.

it can greatly make the false alarm or error rate much lower than HOG.

5 Conclusion

The HSPOG descriptor proposed in this paper can improve the recognition rate of HOG in target recognition. The using of spatial pyramid scales in HSPOG will make the information of an image keep well without missing, so that HOG features can be calculated on different spatial scales of an image and information can be fully utilized.

In this work, HSPOG only needs to adjust two parameters, one is the scaling ratio, the other is the image cell number factor $n$. These two factors are not difficult
Shaojun Guo et al.: HSPOG: An Optimized Target Recognition Method Based on Histogram of . . .

Fig. 8 Curves of false alarm and recognition rate of experiments about (a) person, (b) potted plant, (c) sheep, (d) sofa, (e) cat, and (f) dinningtable.

to adjust. The main considerations we have discussed are how to avoid distorting the image and make more complete use of the data through the spatial pyramid scales. The HSPOG can be summarized as the following:

1. Zoom the images or sub-images of arbitrary size/scale.
2. Divide the zoomed images into \(2^n \times 2^n\) cells.
3. Compute the \((2^n - 2) \times (2^n - 2) \times 32\) descriptor matrix for each input image and label it.
4. Convert the feature maps of samples into vectors.

From the experiment of ship target recognition, we can find that increasing quality of the HSPOG may make the recognition rate much better than before, and it is expected to be applied in some application scenarios.
Fig. 9 Part of positive and negative training samples. (a), (b), and (c) are samples of positive images and (d), (e), and (f) are samples of negative images.

Fig. 10 Comparison of HSPOG and HOG after increasing feature quantity (Each point represents 50 sample images and the total number of testing images is 1684).

Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (No. 51802348).

References

[1] L. M. Surhone, M. T. Tennoe, and S. F. Henssonow, Histogram of oriented gradients, Betascript Publishing, vol. 12, no. 4, pp. 1368–1371, 2010.
[2] B. Liang and L. Zheng, Diffractive phase elements based on two-dimensional artificial dielectrics, presented at the 22th International Conference on Pattern Recognition, Stockholm, Sweden, 2014.
[3] Q. Liu, Z. G. Wu, and J. M. Guo, The conversion of histograms of oriented gradient in different vision-angle and rotation-angle, Control Theory & Applications, vol. 27, no. 9, pp. 1269–1272, 2010.
[4] S. A. Iamsa and P. Horata, Hand written character recognition using histograms of oriented gradient features in deep learning of artificial neural network, presented at the 3th International Conference on IT Convergence and Security, Macao, China, 2013.
[5] Y. W. Pang, Y. Yuan, X. L. Li, and J. Pan, Efficient HOG human detection, Signal Processing, vol. 91, no. 4, pp. 773–781, 2011.
[6] Y. E. Lina, Y. L. Chen, and J. L. Lin, Pedestrian fast detection based on histograms of oriented gradient, Computer Engineering, vol. 36, no. 22, pp. 206–207, 2010.
[7] K. Grauman and T. Darrell, The pyramid match kernel: Discriminative classification with sets of image features, presented at the 10th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Beijing, China, 2005.
[8] N. V. Tavari and A. V. Deorankar, Indian sign language recognition based on histograms of oriented gradient, International Journal of Computer Science & Information Technology, vol. 5, no. 3, pp. 3657–3660, 2014.
[9] H. X. Jia and Y. J. Zhang, Fast human detection by boosting histograms of oriented gradients, presented at the 8th International Conference on Image and Graphics, Tianjin, China, 2007.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, vol. 25, no. 2, pp. 1–8, 2012.
[11] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, presented at the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 2014.
[12] J. Donahue, Y. Jia, and O. Vinay, DeCAF: A deep convolutional activation feature for generic visual recognition, https://arxiv.org/abs/1310.1531, 2013.
[13] R. Girshick, J. Donahue, T. Darrel, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, presented at the 31th IEEE Conference on Computer Vision, Columbia, CA, USA, 2014.
[14] K. He, X. Zhang, and S. Ren, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 37, no. 9, pp. 1904–1916, 2015.
[15] P. C. Hung, Colorimetric calibration in electronic imaging devices using a look-up-table model and interpolations, Journal of Electronic Imaging, vol. 2, no. 1, p. 53, 1993.
[16] P. Felzenszwalb, D. Mcallester, and D. Ramanan, A discriminatively trained, multiscale, deformable part model, presented at the 25th Conference on Computer Vision and Pattern Recognition (CVPR), Alaska, AK, USA, 2008.
[17] J. P. Dong and C. Kim, A hybrid bags-of-feature model for sports scene classification, Journal of Signal Processing Systems, vol. 81, no. 2, pp. 249–263, 2014.
Shaojun Guo received the PhD degree from Navy Aeronautical University, Yantai, China in 2017. He is currently a researcher at National Innovation of Defense Technology, Academy of Military Sciences PLA China. His current research interests include metamaterial design and machine learning algorithms.

Feng Liu received the PhD degree from Navy Aeronautical University, Yantai, China in 2017. He is currently a researcher at National Innovation of Defense Technology, Academy of Military Science PLA China. His current research interests include image processing, target recognition, and machine learning algorithms.

Xiaohu Yuan received the PhD degree from Tsinghua University, Beijing, China in 2018. He is currently an engineer at the Department of Automation, Tsinghua University. His research interests include image processing and quantum information processing.

Chunrong Zou received the PhD degree from National University of Defense Technology in 2016. He has been working as a researcher at National Innovation of Defense Technology, Academy of Military Science PLA China since 2018, where he is mainly involved in the electromagnetic design, characterization of metamaterials, and artificial dielectrics. He has authored/coauthored more than 20 papers in peer-reviewed journals and conference proceedings. His research interests include information and communication engineering and computer application on metamaterial design.

Li Chen received the PhD degree from National Defence University of People’s Liberation Army in 2015. He has been working as a boffin at National Innovation of Defense Technology, Academy of Military Science PLA China since 2018. He is now engaged in scientific and technological innovation management research and has published more than 20 academic articles in domestic newspaper and journals. His main research interest is the application of computational procedures to project management.

Tongsheng Shen received the PhD degree from Beijing University of Aeronautics and Astronautics, Beijing, China in 2002. He is currently a professor at National Institute of Defense Technology Innovation. His research interests include information and communication engineering and computer application.