Optimized Scale Invariant HOG Descriptors for Object and Human Detection

Akila K¹, Pavithra P²
¹²Assistant Professor, R.M.K College of Engineering &Technology, Chennai, India
pavithracse@rmkce.ac.in

Abstract

This article presents a scale variation approach to identify objects and humans in video sequences using histogram of gradient descriptor. A significant restriction in HOG descriptors is its variations with scale changes and illumination changes, as is frequently the considered case. We recommend unique SIO-HOG descriptors that are figured to be invariant to scale changes. The system associates the benefits of adoptive bin selections and sample resizing in the object recognition process. We analyze the effect of PCA transform based feature selection process on object detection performance, ultimately the finite scale range, adoptive orientation binning in non-overlapping descriptors is all main thing for nominal detection rate. HOG feature vector over complete search window is computationally more exclusive, to acquire more precise set of features with finite Euclidean distance to classify them using KNN classifier. This new approach provides near-perfect ways of separating humans from other objects. The whole object detection system was assessed on a few test samples from real-world data sets and compared beside a publicly available pedestrian detection data base.

Keywords: scale-invariance, HOG descriptor, object detection, PCA transform, spatial features, classification

1. Introduction

Detecting objects in images or vision input is computationally exhaustive tasks since massive quantity of pixels needs to be processed and complex arithmetic operations are considered. Even then, the outcome is still more amount of features that prerequisite to be calculated due to its variable appearance, scale changes and the pose variations. To incorporate robust feature sets like edge orientation histograms [1], SIFT descriptors [2] are computed on a consistently spaced cells to overcome all these trials. We briefly discuss previous work on object detection in section 2, give a detailed overview of our HOG method section 3, describe the detailed description and experimental evaluation of object and human detection process in section 5 and concluded in section 6.

2. Previous Work

The improvements in classifier routines and the increasing accessibility of large range of feature sets to encode local information of objects offers quite a few ways to expand the detection precision. In [3] Haar-like wavelet feature set and space-time differences structures were used to figure an effective region-based detection for moving person detector. In [4] use the mixtures of orientation histograms with threshold gradient to compute feature vector separately for each parts of the object. HOG features have established its discernment for the many object description and detection. HOG feature sets are not
only used to describe the appearance of the object and it is also used to reveal the object shape. This creates them elect and actual technique for object detection. However, it has been used in a few object detection mechanisms it has the incapability to detect objects in various poses. This problem is resolved using discriminatively trained models with the assistance of multi-class SVMs [5]. Survey on Illumination Condition of Video / Image under Heterogeneous Environments for Enhancement. [6] To preserve the uniqueness, sustaining the original scale invariant property mirror invariant descriptors are framed according to the magnitude beneath flip and rotation. [7] Moreover scale invariant based on shape context, other local feature sets such as Continuous wavelets [8] and SIFT [9] have been used as key point-based detection methodologies.

3. HOG descriptor

Histograms of Oriented Gradients are high-level features computed based on image gradient of decomposed patches. The HOG feature vector constructed from each object are used to assess the detection process with limited discrimination. Feature vectors are consequently train the classifier for automated object detection. It has been used in utmost object detection by evaluating well-normalized image gradient orientations in the past decade [10]. Basically, HOG descriptor is estimated based on two dissimilar geometries, rectangular HOG with square spatial cells, and circular HOG partitioning cells in log-polar fashion.

3.1 R-HOG

R-HOG blocks have many likeness measures with scale invariant descriptors but adaptive in nature to local image conditions since RHOG is computed at a single scale without dominant orientation alignment, whereas SIFT’s computed at a scale-invariant key points, with their dominant orientations. [11] RHOG descriptors typically consist of multiple block types with different cell and block sizes to advance the performance.

3.2 C-HOG

Circular block (C-HOG) descriptors are frequently look like shape frameworks but feature vector comprises gradient-weighted orientation cells instead of a single edge orientation cell. CHOG descriptors are having radial bins to advance the performance in the form of centre-surround coding.

3.3. Spatial / Orientation Binning

The usage of orientation histograms for object detection has many pioneers [12], but it spreads constancy only when it is combined with local spatial information and normalization. Independently HOG feature vector calculates edge orientation histogram and are assembled into orientation bins over object regions i.e spatial cells. Cells are rectangular blocks, and the orientation bins are equivalently spread out over 0° – 180° as an unsigned gradient with step size of 20° to 40°. To reduce spatial redundancy, image search window is divided into non overlapping blocks since orientation bins are interpolated.

4. Object Recognition

The over-all architecture of the object detection system is shown in Figure 1. Subsequent the macro block conversion paradigm, [13] the HOG extraction process has been characterised into two phases: bin selection based on intensity variation and re sampling based on appearance changes. The workflow is as follows: The primary stage, the precomputation stage, generates bins and sample size, i.e., rectangular regions and angular bins which generate HOG feature vectors. These extracted HOG get into the next level, the feature selection stage, which analyzes feature sets and selects feature vectors with great variance in a more systematic way. [14] It supports the existence of a most significant feature
and rejects the redundant sets. To end with classification system output relates bounding boxes on object detected windows with described size.

**Figure 1.** General architecture of the object detection system

### 4.1. Precomputation stage

In the HOG generation stage, we use an adoptive bin selector and sampling size for robust object detection over poor illumination and fast scale changes due to the two subsequent facts (for details on feature extraction [E]):

- **Simple feature set:** HOG features provide a very basic set of features that is computed efficiently from image patches.
- **Bin size:** HOG feature vector magnitudes are directly depending on input image pixel values. Hence changes in the intensity level will change the HOG description. Feature vector is more sensitive when the number of bins is high.

### 4.2 Feature selection stage

To advance a hypothesis HOG feature set that detects as many objects as possible; the detection rate should be kept high, even if the HOG vectors lied in false positives. For that reason, we carried out principal component analyzes (PCA) to select dominant features, using hierarchical correlation matrix followed by variance computation. This also helps to detect the smallest objects even with the wide range of different poses and appearances.

**Pseudo code:**

```
Step 1: Frame conversion
Step 2: For i = 1 to Total number of frames
    for each frame I
Step 3: select bin & cell size.
    Compute HOG gradient model
Step 4: PCA transform analyzes.
    for m=1: no. of HOG features
        find variance & correlation matrix.
        If correlation level > integer bound
            Include parameter SET!
        end
    end
Step 5: carry out discriminative analyzes.
    set orientation level.
    check symmetric level.
Step 6: If detected HOG model = Human HOG prescription
    Matching count++.
    end
    if Matching count > parameter SET
        Assert HOG optimization and input parameter set evaluations.
Step 7: End
```
In the last stage, the hypotheses are confirmed with computationally efficient classification mechanism. It uses a different set of HOG features extracted from previous stage. Whereas the size of feature vector is slower, the classification performance of KNN classifier is more accurate, since the number of false positives is low.

The HOG feature vectors are extracted over a search window with the size of 64 × 128 pixels and 8x8 cell size; this allows capturing detailed local information. 10 positive samples were taken from the PASCAL VOC 2007 dataset and trade-off between window sizes over detection rate is analysed. The object detector analyses over various resolution images, Scales range from 68×56 to 128×64 pixels, possible cell sizes. As shown in Figure 2 16x16 cell size gives poor detection rate even with larger window size. Since the object samples are quite small, the HOG descriptor is limited to minor details. Thus, we found that low resolution object samples have little effect on the detection performance when the cell size is small.

4.4 Human detection

An image, I, is first partitioned into multiple non overlapping 8 × 8 patches \( P_i \), \( i = 1, \ldots, N \). For each patch, a 9-bin HOG descriptor with the quantization angles \( j \times 20^\circ, j = 0...8 \) is used to signify its local spatial information’s. Though, HOG descriptor deficiencies the symmetric properties as shown in Figure 3, and consequently we distinct the symmetric relationships of HOG descriptors. Figure 4 illustrations that large cell size of an object holds a peak bin angles and its corresponding HOG approximates the gradient direction.

As cited above, image patches be appropriate to the human spatial regions form a symmetric HOG vectors. Thus, the symmetric HOG feature vector can be defined as:
\[
M = \{(X_i, X_j) \mid X_i \in T_1 \land X_j \in T_2 \}
\]

where \( T_1 \)- Train sample, \( T_2 \)- Test sample

Note that the HOG feature vectors need to be large enough to preserve most of the potential symmetric features. Thus, the computational complexity to execute the human detection is not high. The local similarity measurement for \( \{X_i, X_j\} \) on the HOG \((r, \theta)\) space \( P: P(r, \theta) = P(r, \theta) + \text{Sim}(X_i, X_j) \)

4.5 Feature extraction

The extracted features are based on histograms of local gradients (HOG) that are computed from the Non- overlapping image patches in the search window. For each cell size, its histogram values describe the gradient magnitude in particular directions. As stated, former histograms with 9 bins (ranging from 0° to 180°) seem to be select for object detection process. Since the image patches are rectangular and the histogram bins are efficiently computed to build HOG feature vector: To work out integrated samples gradients are extracted from non-overlapping blocks to allows saving time for object detection process.

5. Experimental results

Here KNN classifier is used to assess the performance of object detection by means of gradient-based HOG descriptor with adoptive bin selection scheme. Meanwhile prior information’s are not existing about objects for choosing the class with the major probability. Generally, there are some notable outcomes in object detection rate. The fact that the scale invariant HOG greatly out-performs existing HOG model and that any important modifications in the intensity level before compute HOG gradients harms the true positive samples highlights that ample of the available spatial orientation information is from the fine scales, and the confidence of dropping the sensitivity to illumination and scale changes is increased. Here HOG gradients are calculated at the finest image scale, rectified for adoptive orientation.
binning, and invariant to the blurred spatial object region. On the other hand, at least for human detection, it pays more consideration to asymmetric nature of orientation rather fine symmetric vectors.

Table 1: Recognition Rates of Dense and Custom Descriptor

| DATA set-PASCAL VOC 2007 data base | Block decomposition | Eigenvectors selected. |
|-----------------------------------|---------------------|------------------------|
| EHOG model -PCA (Carl Vondricket et a.2013) | Overlapping model | 28                     |
| SIO-HOG model -PCA | Non overlapping model | 22                     |

Table 2: Average detection rate performance of proposed HOG class in the PASCAL VOC 2007 dataset.

| DATA set-PASCAL VOC 2007 data base | False detection rate | Average detection rate performance |
|-----------------------------------|----------------------|-----------------------------------|
| EHOG model (Carl Vondricket et a.2013) | 13%                  | 16 classes                        |
| SIO-HOG model                     | 10%                  | 19 classes                        |

Table 3: Human detection rate analyzes in the INRIA pedestrian dataset.

| DATA set-INRIA pedestrian data base | False detection rate | positive detection rate |
|-------------------------------------|----------------------|-------------------------|
| HOG model (M. Kachouane et a.2012)  | 17%                  | 86%                     |
| SIO-HOG model                       | 12%                  | 92%                     |
Figure 2. Detection rate analyzes of HOG over Scales range from 68×56 to 128×64 pixels.

Figure 3. Symmetric properties of HOG model for human and object discrimination.
We proposed novel SIO-HOG descriptors, which is mined by transforming HOG descriptors to be invariant to scale variation and illumination changes. Different appearance variation and illumination changes have been estimated to demonstrate the robustness of the HOG descriptor configuration. We analyze the performance metrics of our HOG model in both object and human detection and proved it is superior to the existing HOG descriptors through experiment with the PASCAL VOC 2007 dataset and INRIA pedestrian data sets. Furthermore, we exhibited the efficiency of effective feature reduction method. Meanwhile we analyzed the object detection and recognition rates with real time world data we substantiated that the presented object detection system marks an enhancement compared to systems solely based on HOG orientation with constant bin selection. Additional researches also will emphasis on the implementation of human action recognition system to show robustness of proposed HOG model over human detection rate per frame.

References
[1] Freeman W T and. Roth M 1995 Orientation histograms for hand gesture recognition. Intl. Workshop on Automatic Faceand Gesture-Recognition, IEEE Computer Society.
[2] Lowe D G 2004 Distinctive image features from scale-invariant key points. IJCV.
[3] Viola, M. Jones J. and Snow D 2003 Detecting pedestrians using patterns of motion and appearance. pp.734–741.
[4] Mikolajczyk K Schmid C Zisserman A 2004 Human detection based on a probabilisticassembly of robust part Detectors. 8th ECCV, pp. 69–81.
[5] K.Akila, S.Chitrakala, S.Vaishnavi 2016 Survey on Illumination Condition of VideoImage under Heterogeneous Environments for Enhancement ICACCS 2016 IEEE
[6] Felzenswalb P F Girshick R B McAllester D 2010 Object detection with discriminativelytrained part based models IEEE on Pattern Analysis and Machine Intelligence.
[7] Xiaojie Guo and Xiaochun Cao 2012 MIFT: A framework for feature descriptors to be mirror reflection invariant Image and Vision Computing. Pp. 546–556.
[8] Reddy Vijaya Kumar Kiran Kumar, Object Detection by 2-D Continuous Wavelet Transform ComputationalScience and Computational Intelligence IEEE.
[9] Zhao Wan-Lei Chong-Wah Ngo 2013 Flip-invariant SIFT for copy and object detection IEEE Transactions on Image Processing pp. 980-991.
[10] Mikolajczyk K Schmid. C 2004 Scale and affine invariant interest point detectors IJCV pp.63 86.
[11] Kanezaki Asako Yusuke Mukuta, and Tatsuya Harada 2014 Mirror reflection invariant HOG descriptors for object detection Image Processing ICIP IEEE Conference.
[12] Pedersoli Marco TimmTuytelaars 2014 A scalable 3d hog model for fast object detection and viewpoint estimation 3D Vision 2nd International Conference -IEEE.
[13] Chayeb A 2014 HOG based multi-object detection for urban navigation IntelligentTransportation Systems
[14] Chen Lu Panfeng Huang, and Jia Cai 2015 Object detection for non-cooperative targets using HOG-based proposals Robotics and Biomimetics.
[15] Akila K Chitrakala S 2014 Discriminative Human Action Recognition using HOI Descriptor and Key Poses ICSEMR
[16] Singh Shekhar Gupta S C 2016 Human object detection by HoG features PDGC IEEE conference
[17] Akila K Chitrakala S 2019 Highly refined human action recognition model to handleintraclass variability & interclass similarity International Journal of Multimedia