RIHOG-BoVWs for Rotation-invariant Human Detection

Baozhen Liu1*, Hang Wu2, Yu Zhang1, Wenlong Xu1, Kaiyu Mu1 and Yuhua Yao1

1 China Astronaut Research and Training Center, Beijing, 100094, P.R.China
2 Institute of Medical Support Technology of Academy of System Engineering, Academy of Military Sciences, Tianjin, 300161, China

*liubaozhen91@126.com

Abstract: Rotation-invariant human detection is of vital significance to many applications, including victim detection via a daylight camera on an unmanned aerial vehicle (UAV). With this problem in mind, we propose a powerful rotation-invariant object-level description method, i.e., RIHOG-BoVWs, which is achieved by local-level feature calculation and high-level semantic feature extraction. In local-level feature calculation, we consider Rotation-invariant Histograms of Oriented Gradients (RIHOG) as the local descriptor and project gradient information into the frequency domain, where the rotation-invariant representation vector is constructed. In high-level semantic feature extraction, Bag of Visual Words (BoVWs) model is employed to achieve global description from local features without considering their spatial structures. Experimentally, we first confirm the sufficient discrimination power of the RIHOG-BoVWs on the public Freestyle Motocross dataset, and then demonstrate the high performance of RIHOG-BoVWs on a victim dataset which has varied backgrounds and body postures.

1. Introduction

After disasters, finding human victims is one of the primary goals of search and rescue efforts. It is possible to quickly survey a disaster site from the air to identify humans in need of help. In this paper, we focus on the human detection problem in disaster scenes from a UAV perspective using a daylight camera.

As an important branch of computer vision, the human detection problem has attracted a lot of attention and most research focuses on pedestrian detection [1]. As shown in figure 1, the human body seen from aerial images has more varied postures and is very distinct from pedestrians appearing in upright poses such as standing or walking. The most fundamental limitation of imaging equipment is the varied relative orientations of human bodies in the image. In this paper, we concentrate on the modification of Histograms of Oriented Gradients (HOG) [1], which is the most successful feature in
pedestrian detection, and integrate Rotation-invariant HOG (RIHOG) [3] with Bag of Visual Words (BoVWs) [4] model to achieve rotation-invariant semantic description for human bodies.

The rest of the paper contains an overview of related work in Section 2, a description of HOG in Section 3 and the semantic description method we propose in Section 4. A thorough set of experiments is then presented in Section 5, where the performance of RIHOG-BoVWs is proven, before we finally conclude our work in Section 6.

2. Related work
In computer vision, many researchers seek to find solutions to rotation-invariant object detection. The simplest approach is detecting an image from several discretized orientations using a learning classifier trained by calibrated samples [2, 5]. In addition to this method, rotated object detection is also addressed as a multi-class detection task, where the orientation are discretized artificially and every direction of objects is regarded as one category [6, 7]. Another approach to in-plane rotation is the development of learning procedures. [8] decouples the orientation estimation from object classification and [9] incorporates the rotation-invariance into a discriminatively trained structured Support Vector Machine (SVM). Besides, many researchers pay attention to the rotation-invariant semantic description.

Orientation normalization in Scale-Invariant Feature Transformation (SIFT) [10] is one of the common ways to construct the rotation-invariant feature, which aligns the local coordinate system to the dominant direction at each detected point. Learning from SIFT, [11] first predicts the dominant direction in the histogram bins of HOG and then rotates the image to the estimated orientation to calculate the final HOG feature. The method of orientation normalization relies on the assumption that such a dominant direction is available and robust enough, which restricts its application to arbitrary positions or dense feature computation.

Compared with the Cartesian coordinate system, the polar representation has a unique advantage for achieving rotation-invariance. In polar coordinate systems, [12, 13] propose a rotation-invariant HOG-like feature, which takes the tangent direction as the reference of gradient orientation and then accomplishes the information aggregation on multiple concentric spatial bins. The new feature, i.e., RIFF, is originally proposed for the purpose of keypoint matching and has no sufficient discrimination in object detection. [14] improves RIFF and propose Sector-ring Histograms of Oriented Gradient (SRHOG) method to obtain sufficient discriminations of human bodies. However, RIFF and SRHOG adopt the local coordinate system to determine pixel gradients and have no translation invariance, which limits their detecting efficiency.

Another way to realize the feature’s rotation-invariance is via the Fourier transform, which extracts the rotation-invariant information of an image in frequency domain. [15] represents Circular Fourier-HOG (CHOH) and [16, 17] propose Local Binary Pattern Histogram Fourier Features (LBP-HF). [3] considers the gradient orientation as continuous, and redefines Histograms of Oriented Gradient (HOG) in the frequency space to obtain the new descriptor, i.e., RIHOG. In contrast to the common Fourier-based features, RIHOG focuses not only on the frequency magnitude in the spectrum of the image but also the phase information and the global dependency of those frequency terms, improving the method’s descriptive power in a significant manner.

Although there are a multitude of rotation-invariant features, the unavoidable information loss of their construction weakens their discriminations, and they are not powerful enough to discriminate human bodies. In this paper, we consider RIHOG as the low-level descriptor to catch object’s specialties and generate a set of local features, upon which the Bag of Visual Words (BoVWs) model is then applied to extract the global description. The method is called RIHOG-BoVWs and has stronger discrimination for human detection with rotation-invariance.

3. HOG descriptor
As a gradient based feature, HOG has been paid great attention and applied to human detection extensively [18, 19] since it was proposed in 2005. HOG concentrates on the gradient orientations of
local image patches and employs a histogram binning, spatial aggregation and overlapping local contrast normalization to strengthen the robustness to illumination and shadow.

![Gradient binning and spatial clustering](image)

Figure 2. Illustration of HOG’s calculation. HOG extracts the object’s local features by the orientation histogram and clusters them to the global descriptor in the form of series. Both of two steps are contrary to rotation-invariance.

As shown in figure 2, the calculation of HOG can be summed up into two parts: the low-level feature calculation; and the high-level semantic feature extraction. The former transforms gradient information to orientation histograms, using local image patches and the latter accumulates local histograms to the final descriptor. Apparently, both of these two steps are not rotation-invariant.

4. The proposed method
In this section, we put forward a method to achieve the rotation-invariant global description for human detection. The method is completed from two aspects of low-level feature calculation and high-level semantic feature extraction based on the analysis of HOG in Section 3.

4.1. Rotation-invariant low-level feature
Following the idea of HOG, RIHOG [3] encodes the gradient information in a rotation invariant manner. The calculation of RIHOG can be divided into the following three steps.

4.1.1. Pixel-level features
In HOG, a pixel’s gradient orientation is discretized into the histogram, which changes in a complex manner when the image rotates. In contrast, RIHOG employs the continuous representation of gradient orientation, which means creating an orientation distribution function at each pixel. Taking the pixel with gradient $g$ as an example, the function can be defined as:

$$\rho(\phi) = \|g\| \delta(\phi - \psi(g))$$  \hspace{1cm} (1)

where $\|g\|$ is the gradient magnitude and $\psi(g)$ denotes the gradient orientation that can be any value in $[0, 2\pi)$. Then, project the function into the Fourier space and the Fourier coefficients read

$$f_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \rho(\phi) e^{-im\phi} d\phi = \frac{1}{2\pi} \|g\| e^{-im\psi(g)}$$  \hspace{1cm} (2)

where $m \in \mathbb{C}$ denotes the order of Fourier basis. Equation 2 means that the pixel’s gradient information can be represented by a series of Fourier coefficients. Since gradient magnitude and orientation are real-valued, the terms of $m < 0$ are redundant and can be discarded. When implementing, $m$ has to be limited to $0 \leq m \leq M$, only getting the approximate representation of gradient information. Controlling the value of $M$ is equivalent to low-pass filtering in the frequency domain, which can weaken the aliasing effect in the orientation quantization because that higher frequency terms are more sensitive to noise.

Extend equation 2 to the image scale and get the Fourier coefficients matrices that encode the image’s gradient information:

$$F_m = \frac{1}{2\pi} \|G\| e^{-im\psi(G)}$$  \hspace{1cm} (3)

where $G = \nabla I$ denotes the gradient field of the image $I$.

To further improve the robustness to local deformation, soft binning, like linear interpolation in HOG, is necessary and can be realized here by spatial convolutions upon the Fourier coefficients matrices. The convolution kernel must be isotropic and triangle or Gaussian kernel is optional.
Additionally, the local contrast-normalization can also be implemented based on convolutions. Therefore, the Fourier coefficients of the matrices are updated to

\[ P_m = \frac{F_w * K_1}{\sqrt{|G|^2 * K_2}} \]  

(4)

where \( K_1 \) and \( K_2 \) is the convolution kernel for the soft binning and the local contrast normalization, respectively. \( P_m \in \mathbb{C}^{2 \times 2} \) is the final pixel-level feature matrix, and its size is as same as the raw image. Figure 3 shows an example.

\[ \text{Figure 3. The example of gradient projection in Fourier space.} \]

The example of local region partition, from which we find that the Fourier basis \( e^{ik\varphi} \) can be considered as the weight distribution term upon pixels within the circular ring.

\[ \text{Figure 4. The example of local region partitions.} \]

4.1.2. Local aggregation Local aggregation is important to the feature’s property. According to [20], the optimal choice in the angular part of the circular ring is the Fourier basis and the function for computing local features are ready:

\[ S_{j,k}(r, \varphi) = \Lambda(r - r_j, \sigma)e^{ik\varphi} \]  

(5)

where \( k \) is the degree of Fourier basis, \( r_j \) denotes the middle radius of the \( j \)th circular ring and \( \Lambda(x, \sigma) = \max\left(\frac{x}{\sigma}, 0\right) \) defines a triangular function of width \( 2\sigma \). Figure 4 presents an example of the local region partition, from which we find that the Fourier basis \( e^{ik\varphi} \) can be considered as the weight distribution term upon pixels within the circular ring.

Subsequently, local aggregation can be completed via the convolution operation between \( P_m \) and \( S_{j,k} \), i.e.,

\[ H_{j,k,m} = S_{j,k} * P_m \]  

(6)

\( H_{j,k,m} \) denotes the regional features in one image patch. When implementing the process, \( k \) and \( m \) can be chosen and combined freely for the independence between gradient information and spatial region.
4.1.3. The selection of rotation-invariant terms According to the aforementioned derivations, every feature from equation 6 can be described as a separable form:

\[ h = A e^{-imz + j\omega} \]

where \( A \) denotes the magnitude of the spectrum while \( e^{-imz + j\omega} \) represents the frequency and phase. It is well-known that the power spectrum, i.e., \( A \), will not change when the image rotates, but it is far from enough to describe the characteristics of one object category. So, we need analyze the \( h \) and create more rotation-invariant features.

Taking the pixel \( p \) as an example, when the image rotates by angle \( \theta \), its gradient orientation \( \alpha \) will change to \( \alpha' = \alpha + \theta \). Similarly, all pixels have the same transformation law. Moreover, the absolute positions of pixels are also changed, while the relative positions between pixels remain invariant, which is performed as the translation of phase in the Fourier space. Consequently, the regional feature after rotation can be represented as:

\[ h_r = A e^{-imz + j(p \omega)} = A e^{-imz + j(p \omega - \theta)} \]  

Compared with the feature before rotation, we discover that \( h_r \) and the image rotation’s influence on the regional feature depends on the values of \( k \) and \( m \). When \( k = m \), \( h_r \) remains the same as \( h \), which can be explained by the impact of image rotation compensated by the change of the pixels’ weights. Both real and imaginary parts of \( h \) are put into the final descriptor.

There are also some dependencies between the image information of different circular rings. Consider the following examples, \( h_1 = A e^{-imz_1 + j\omega}$ and \( h_2 = A e^{-imz_2 + j\omega} \). When the proper values of Fourier order are chosen to satisfy \( k_1 - m_1 = k_2 - m_2 \), the two features are subject to the same impacts from image rotation. Under this condition, we can construct the union feature \( h_{12} = h_1 + h_2 \) to preserve the relative phase information, which is rotation invariant, too.

In conclusion, to achieve the rotation-invariant description, RIHOG consists of three types of image formations, which include the power spectrum of gradients in frequency domain, the relative spatial structure of pixels in the same annulus, and the relative spatial structure of pixels in different annuli.

4.2. Rotation-invariant high-level semantic feature

In the view of the discussions, RIHOG is more suitable to be the local descriptor and it is necessary to seek a rotation-invariant spatial clustering method. BoVWs model provides a powerful tool. In this paper, we focus primarily on the feasibility of the BoVWs model in rotation-invariant human detection, and we pay no attention on its optimization. Thus, we adopt the classical framework, under which the rotation-invariant global description can be divided into the following three steps:

**Step 1:** Sample the image densely using RIHOG as the local descriptor. Let every RIHOG be a \( D \)-dimensional vector, so the local feature set can be described as \( X = [x_1, x_2, \ldots, x_k] \) \( X \in \mathbb{R}^{D \times K} \), where \( K \) is the number of local features in one image.

**Step 2:** Mix the RIHOG features of training images and generate the codebook via \( K \)-means clustering algorithm. Consider \( N \) training images, and the \( X_{train} = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{D \times NK} \) denotes the RIHOG feature pool. Firstly, \( M \) basis vectors are selected randomly in the pool as the original clustering centers, each of which represents the different feature space. Then, assign each RIHOG feature to one feature space according the minimum Euclidean distance principle and re-determine the clustering centers of every space. Repeat the steps until the maximum iterations or the demanded cumulative approximation error \( \sum_{t=1}^{T} \| x_t - b \|^2 \). The final clustering centers are visual words, which compose the codebook \( B = [b_1, b_2, \ldots, b_M] \in \mathbb{R}^{D \times M} \) and are stored offline.
Step 3: Redescribe the image with the codebook. Firstly, for each local feature contained in a image, select the most similar visual word $d_i = \arg \min_j \| x_i - b_j \|_2$. Then, create a sparse histogram $H \in \mathbb{R}^{M \times K}$ over the visual words set $D = \{ d_1, d_2, \ldots, d_K \}$ and consider it as the final coding result after normalization.

After completing the feature description of all training samples, the classification model is trained via the selected method. When detecting, the test image is also described by the above processes and then decided by the classifier.

5. Experiments

To facilitate the comparison with other research and prove the effectiveness of the proposed method, two datasets are used in our experiments: one is the Freestyle motocross public dataset; and the other is our victim dataset.

The Freestyle motocross dataset was collected by Villamizar et al [8] and has been used in [21]. In Freestyle motocross dataset, there are two image sets for testing: one containing motorbikes without rotations; and another one corresponding to motorbikes with various in-plane rotations. Our victim dataset is established by simulating the perspective of a camera mounted on a UAV during the victim detection task. All images are collected with a quadrotor, in the genuine environment, under uncontrolled illuminations and shadows. We simulate the typical realities and difficulties of the rescue scenario as much as possible, forcing the dataset to have a large diversity of body postures, which include multiple in-plane rotations and various out-of-plane rotations.

5.1. Experiment setup

Our experiments contains four parts. First, using Freestyle motocross public dataset, the feasibility of RIHOG-BoVWs for complicated object detection is confirmed through an orientation-specific motorbike detection. Second, while still using Freestyle motocross public dataset, we conduct rotated motorbike detection tests to demonstrate the effectiveness of our method for rotated object detection. Third, using our victim dataset, we compare the RIHOG-BoVWs with other features and prove its superiority for rotation-invariant human detection. Finally, the processing time of different methods for victim detection is analyzed.

In the evaluation, whether a prediction is consider true or false depends on if the ratio of the intersection over the union of the bounding box and the ground truth exceeds 0.5. Multiple predictions of the same ground truth object are considered to be false positive except one as true. For all of our experiments, we report equal error rate (EER) and show precision-recall (PR) curves.

In addition, for all detection tests, we employ linear SVM as the classification method with its value optimized by 5-fold cross-validation.

5.2. Experiment 1: orientation-specific object detection on Freestyle motocross dataset

The first test dataset in Freestyle motocross has 69 images containing 78 motorbikes without explicit planar rotations. This dataset is used to verify the feasibility of RIHOG-BoVWs for object detection. In addition to our proposed method, we apply RIFF, RIHOG and HOG as the global descriptors to make comparisons. Additionally, the popular classification strategy, Dense SIFT plus BoVWs, is being tested in the detection task too. Relative to the standard SIFT, the Dense SIFT adopts a dense sample and cancels the dominant gradient direction, which make it very similar to HOG and is not rotation-invariant. So the method of the Dense SIFT plus BoVWs can also be viewed as the application of BoVWs model on HOG. The learning is done using 500 positive samples and 2000 negative samples selected randomly from the Caltech motorbike dataset [22].
Results of the first experiment are shown in figure 5. RIHOG-BoVWs achieves 0.7949 EER, which is comparable with HOG and is much better than other methods. Comparing to [8], their best result is 0.9103 EER, which is much higher than our results. In addition to the non-linear classification method, their results are achieved after training with 800 images that are generated by adding affine transformations from the most closely related images to the target motorbike model according to the author’s doctoral thesis [23].

5.3. Experiment 2: rotation-invariant object detection on Freestyle motocross dataset
Regarding the second motorbike dataset of Freestyle motocross, there are 100 images of 128 motorbikes with challenging rotations. Like [21], in our experiment, the training set is created by taking the first 40 images and the remaining 60 images are considered as the test dataset.

In rotation-invariant motorbike detection, we also use RIFF, RIHOG and HOG as the global descriptors to compare with. For HOG, when detecting, the image detection is recalled 15 times with a step of 24 degrees to respond to objects with different orientations.

Detection performance of the second experiment is shown in figure 6. Among all the methods, RIHOG-BoVWs achieves the best performance, which demonstrates its superiority for rotation-invariant object detection. In [21], the result is 0.90 EER, which is better than our method. However, their approach is based on complex harmonic basis and equivariant filters. On the contrary, we focus on the rotation-invariant description and employ a linear classification method, which is more intuitive and concise.

5.4. Experiment 3: rotation-invariant human detection on our victim dataset
Finally, we conduct the experiment on our victim dataset. Apart from the methods in Section 5.3, we also compare the HogLbp[24], DPM[25], and ACF (Aggregate Channel Features) [26], which have revealed high-performance results in pedestrian detection. As a note, we recall the image detection 15 times with a step of 24 degrees to realize the desired victim detection when using HOG or other pedestrian detectors. Our victim dataset consists of 158 images with 58 training images and 100 testing images, where bounding boxes have been manually labeled.
Detection results are shown in figure 7. From the comparisons, we find that RIHOG-BoVWs is more suitable for victim detection from areal views. Figure 8 shows instances of detection results and presents the robustness of the RIHOG-BoVWs method to various rotations and complex postures of human bodies.

5.5. Experiment 4: computational time analysis
We test the time associated with different methods when working on victim detection and present them in figure 9 with HOG as the benchmark. Among all methods, RIHOG-BoVWs has the similar computational complexity with ACF and has huge efficiency in victim detection task.

6. Conclusion
In this paper, we presented a rotation-invariant semantic description method, called RIHOG-BoVWs, to investigate rotation-invariant human detection, which is very significant in the victim detection case from a UAV. Learning from HOG, we concentrate on the object’s local characteristics, which are robust against the geometric deformation and illumination variation. In the proposed method, RIHOG is employed upon image patches to get a local feature set and BoVWs model is then used to extract the semantic object-level description. Both steps guarantee the rotation-invariance of the final descriptor. Experimentally, we tested the sufficient discrimination of RIHOG-BoVWs for complicated objects and evaluated its ability in addressing planar rotations. Finally, we have established a victim dataset.
and compared RIHOG-BoVWs with other pedestrian detection methods. Detection results show that RIHOG-BoVWs has the high performance for victim detection with the advantage of being computationally more efficient.

Acknowledgments
Supported by the Key Program of Experimental Technique of General Armament Department (2014SY54A1102).

References
[1] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005, pp. 886-893.
[2] M. Andriluka, P. Schnitzspan, J. Meyer, and S. Kohlbrecher, "Vision based victim detection from unmanned aerial vehicles," in ieee/rsj International Conference on Intelligent Robots and Systems, 2010, pp. 1740-1747.
[3] K. Liu, H. Skibbe, T. Schmidt, T. Blein, K. Palme, T. Brox, et al., "Rotation-Invariant HOG Descriptors Using Fourier Analysis in Polar and Spherical Coordinates," International Journal of Computer Vision, vol. 106, pp. 342-364, 2014.
[4] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 524-531.
[5] T. Moranduzzo and F. Melgani, "Detecting Cars in UAV Images With a Catalog-Based Approach," IEEE Transactions on Geoscience & Remote Sensing, vol. 52, pp. 6356-6367, 2014.
[6] D. P. Casasent, "Real-time people and vehicle detection from UAV imagery," Proceedings of SPIE - The International Society for Optical Engineering, vol. 7878, pp. 78780B-78780B-13, 2011.
[7] S. Tuermer, F. Kurz, P. Reinartz, and U. Stilla, "Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps," IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, vol. 6, pp. 2327-2337, 2013.
[8] M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto, and A. Sanfeliu, "Efficient rotation invariant object detection using boosted random ferns," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 1038-1045.
[9] A. Vedaldi, M. Blaschko, and A. Zisserman, "Learning equivariant structured output SVM regressors," in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 959-966.
[10] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.
[11] W. Zhang, X. Sun, K. Fu, C. Wang, and H. Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model," Geoscience and Remote Sensing Letters, IEEE, vol. 11, pp. 74-78, 2014.
[12] G. Takacs, V. Chandrasekhar, S. S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Fast computation of rotation-invariant image features by an approximate radial gradient transform," IEEE Transactions on Image Processing, vol. 22, pp. 2970-2982, 2013.
[13] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Unified real-time tracking and recognition with rotation-invariant fast features," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 934-941.
[14] B. Liu, H. Wu, W. Su, and J. Sun, "Sector-ring HOG for Rotation-Invariant Human Detection," Signal Processing: Image Communication.
[15] H. Skibbe and M. Reisert, "Circular Fourier-HOG features for rotation invariant object detection in biomedical images," Proceedings, vol. 88, pp. 450-453, 2012.
[16] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features," Image Analysis, vol. 5575, pp. 61-70, 2009.
[17] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-Invariant Image and Video Description With Local Binary Pattern Features," IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, vol. 21, pp. 1465-1477, 2012.

[18] S. Tang, M. Andriluka, and B. Schiele, "Detection and tracking of occluded people," International Journal of Computer Vision, vol. 110, pp. 58-69, 2014.

[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, pp. 1627-1645, 2010.

[20] Q. Wang, O. Ronneberger, and H. Burkhardt, "Rotational invariance based on Fourier analysis in polar and spherical coordinates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 1715-1722, 2009.

[21] K. Liu, Q. Wang, W. Driever, and O. Ronneberger, "2d/3d rotation-invariant detection using equivariant filters and kernel weighted mapping," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 917-924.

[22] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, 2003, pp. II-264-II-271 vol. 2.

[23] M. A. Villamizar Vergel, "Efficient approaches for object class detection," 2012.

[24] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in IEEE International Conference on Computer Vision, 2009, pp. 32-39.

[25] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[26] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast Feature Pyramids for Object Detection," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 36, pp. 1532-45, 2014.