

Accurate Detection of Human in the Presence of Occlusion

Narayan H M
Associate Professor
Dept of CSE,MSEC

Jeevitha S R
M.Tech 4th sem student
Dept of CSE,MSEC

Abstract – This paper describes a general method to detect humans in presence of occlusion in images. The main aim of this work is to introduce a general method for automatic, accurate and robust detection of human figures in the presence of partial occlusion. Current methods for handling occlusion lack generalization, either because additional information is required (coming from manual annotations of the parts or from other sensors), or they are tied to a specific object class. So to overcome this, proposed a method which does not require manual labeling of body parts, defining any semantic spatial components, or using additional data coming from motion or stereo. The experiments are performed on large datasets like INRIA person dataset, the Daimler Multicue dataset, and a new challenging dataset, called *PobleSec*, in which a considerable number of targets are partially occluded. The different approaches are evaluated at the classification and detection levels for both partially occluded and non-occluded data. The datasets used in our experiments have been made publicly available for benchmarking purposes.

I. INTRODUCTION

HUMAN DETECTION is the key role in many applications related to robot sensing, surveillance, home automation and driver assistance. Detecting humans is a challenging task due to major difficulties coming from the wide variability of the target, such as the shape, clothing or pose; and the external factors, such as the scenario, illumination, and partial occlusions.

State-of-the-art approaches can be divided into two groups: holistic, which rely on detecting the target as a whole, and part-based, which combine the detection of different parts of the body (head, torso, arms, legs, etc.). Holistic methods offer robustness with respect to illumination, background and texture changes, whereas partbased methods are advantageous for different poses. In all cases, the presence of partial occlusions causes a significant degradation of performance, even for part-based methods which are supposed to be robust in that respect.

As expected, detection in the presence of partial occlusions has sparked significant interest. For instance, an accident in which a vehicle hits a pedestrian is likely to occur when the pedestrian is not in full view to the driver, e.g., when it appears from behind a parked car. Captured in a sequence of images, several frames prior to the accident will contain a partially occluded human figure. Therefore, accurate detection in the presence of partial occlusion is of paramount importance when building driver assistance systems.

The propose approach uses random subspace classifiers to learn the different regions of the window and subsequently

find the optimal ensemble through a bespoke selection strategy.

The proposed approach brings several benefits like the approach is generic, therefore applicable to any class of objects. The random subspace classifiers are trained in the original space, no further feature extraction is required. The detection is done on monocular intensity images, unlike other methods

for which stereo and motion information are mandatory and during training, we only require a subset of images with and without partial occlusion; other detection methods require delineation of the occluded area. We also introduce a new real world dataset with occluded pedestrians for testing.

The remainder of this paper is organized as follows.

Section II introduces the existing work. Section III presents the method from a generic point of view. Section IV presents proposed method. Section V presents the main conclusions and future work.

II. EXISITNG SYSTEM

A general approach based on the response of different part detectors and a whole-object segmentation process is introduced. Nevatia *et al.* proposes a method requires a hierarchical objectparts design with eleven components making up the head, the torso and the legs. The edge pixels of the object that positively contribute to the part detectors are extracted and used together with the part detector responses to obtain a joint likelihood of multiple objects. In this joint likelihood an occlusion reasoning is applied. In case of finding any inter-object occlusions, the occluded parts are ignored. The main drawback of this method is that it requires a manual spatial alignment of the objects, which has to be adapted to each object class.

Many detection methods learn object classifiers from a labeled training set. Given a test image, the classifier is applied to the sub-windows with variable sizes at all positions. For detection of objects with partial occlusions, part based representations can be used. For each part, a detector is learned and the part detection responses are combined. The part detectors are typically applied to overlapping windows and the windows are classified independently, hence one local feature may contribute to multiple overlapped responses for one object, see Fig. 1. Some false detections may also occur, as local features may not be discriminative enough. Due to poor image cues or partial occlusions, some object parts may not be detected. Girshick *et*

al. propose an extension of the deformable part-based detector with occlusion handling. Specifically, the method tries to place the different body parts over the window. Then, if some of the parts are not matched, the method tries to fit in their designated place occluding objects learned from the data. The obvious inconvenience of such an approach is the need of learning the objects that occlude the target. Besides, to extend the method to other classes, a different occlusion reasoning has to be defined.

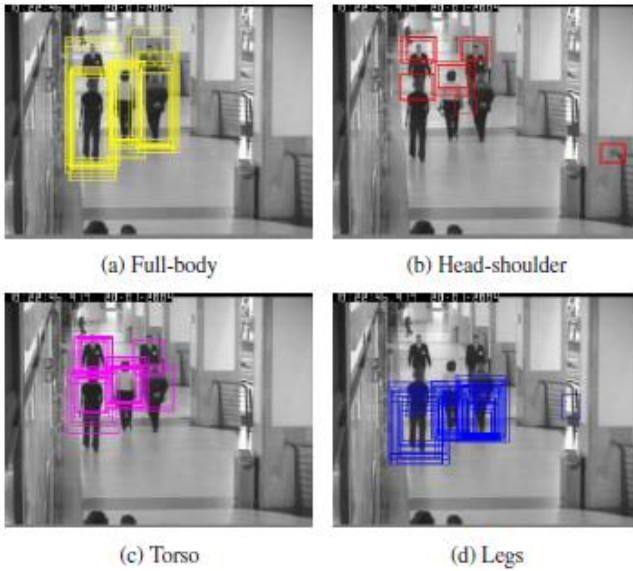


Figure 1. Examples of part detection responses for pedestrians.

Here we propose a method for detecting human figures in still images, which can handle occlusion automatically. Manual annotation or defining specific parts/regions of the window are not needed. Our method is based on an ensemble of random subspace classifiers obtained through a selection process. It is worth mentioning that, as the random subspace classifiers use the original feature space, there is no additional feature extraction cost. Similar to the proposed approach uses a segmentation process to find the unoccluded part of a candidate-window. An ensemble is applied only in uncertain cases. In particular, the proposed method generalizes the inference process presented by extending it to multiple descriptors.

III. PROPOSED SYSTEM

A. Proposal Outline

We present a general method for handling partial occlusions (Fig. 1). In such a design, the window is described by a block-based feature vector. The resulting feature vector is evaluated by the holistic classifier. If the confidence given by the holistic classifier falls into an ambiguous range [Fig. 2], then an occlusion inference process is applied by using the block responses. Finally, if the inference process determines that there is a partial occlusion [Fig. 2], an ensemble classifies

the window. Otherwise, the final output is given by the holistic classifier. Notice that, in order to obtain a more accurate decision, we apply the ensemble only when partial occlusion is suspected. In the following, we explain in detail the components shown in Fig. 2.

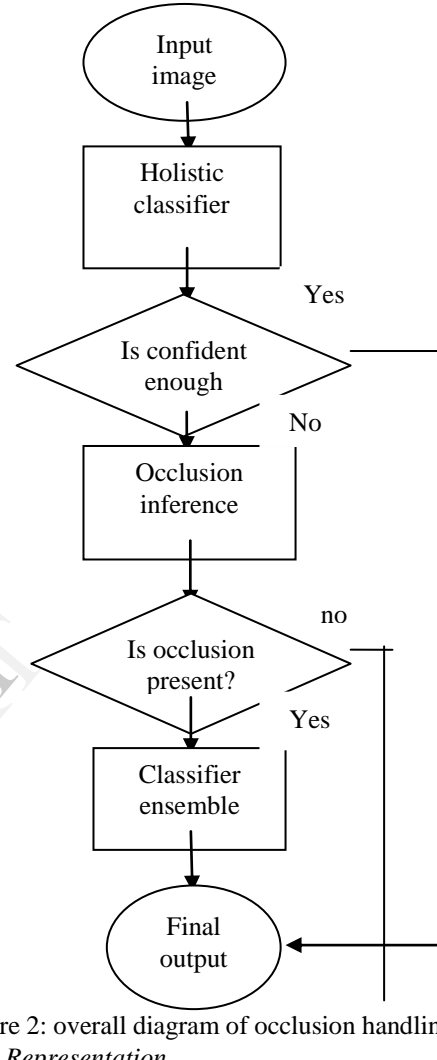


Figure 2: overall diagram of occlusion handling

Block Representation

Our detection system relies on using a block-based representation, one of the most successful descriptor types in use today. A well-known example of such descriptor is the HOG, although many other examples exist. Fig. 2 illustrates the idea of this type of representation, where the window descriptor $\mathbf{x} \in \mathbf{R}^n$ is defined as the concatenation of the features extracted from every predefined block $\mathbf{B}_i, i \in \{1, \dots, m\}$. A block is a fixed subregion of the window as shown in Fig. 2. Our method also allows the blocks to overlap. The descriptor is denoted as

$$\mathbf{x} = (\mathbf{B}_1, \dots, \mathbf{B}_m)^T.$$

The feature vector \mathbf{x} is passed to a holistic classifier H

$$H: \mathbf{R}^n \rightarrow (-\infty, +\infty)$$

$$\mathbf{x} \rightarrow H(\mathbf{x}) \quad (1)$$

where the feature space dimension, n , is $n = m \cdot q$, being q the number of features per block. The higher the value returned by the function H , the higher the confidence that there is a pedestrian in the given window. Note that the function H can

be any classifier that returns a continuous-valued output, for example, a hyperplane learnt with an SVM.

Occlusion Inference and Posterior Reasoning

In order to detect if there is a partially occluded human figure in the image. First, we determine whether the score of the holistic classifier is ambiguous. For example, the response from an SVM classifier can be perceived as ambiguous if it is close to 0. When the output is ambiguous, an occlusion inference process is applied. This is based on the responses obtained from the features computed in each block.

In particular, for every block B_i , $i \in \{1, \dots, m\}$ we define a local classifier h_i

$$h_i : \mathbf{R}^q \rightarrow (-\infty, +\infty)$$

$$B_i \rightarrow h(B_i) \quad (2)$$

where the classifier h_i takes as input the i -th block B_i of the window, and provides as output the likelihood that the block B_i is part of the pedestrian or, otherwise, is part of an occluding object or background.

Ensemble of Local Classifiers

In general, partial occlusions can vary considerably in terms of shape and size; hence a flexible model is needed. We propose an adapted random subspace method (RSM) for this task. In particular, we propose to use classifiers trained on random locally distributed blocks; the collection of such classifiers is subsequently browsed to find an optimal combination. Our adapted RSM is introduced below (Fig. 3).

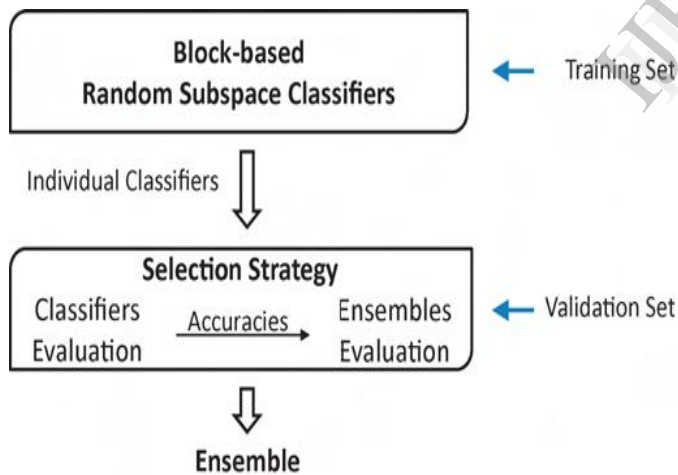


Figure 3: Training of the adapted random subspace method for handling partial occlusion.

Human Detection with Occlusion Handling

In the previous section, we presented a general method to handle partial occlusions for object detection. In order to illustrate and validate our approach, in this section we describe in detail a particular instantiation of our method for the class of humans. In order to apply our method to pedestrians, we make use of both linear SVMs and HOG descriptors, which have been proven to provide excellent results for this object class. In addition to HOG descriptor, we also test our system

using the combination of the HOG and the local binary pattern (LBP) descriptor, which has recently been proposed in for human detection. In the following we explain very briefly each of these components. Given a training dataset D , the linear SVM finds the optimal hyperplane that divides the space between positive and negative samples. Thus, given a new input $\mathbf{x} \in \mathbf{R}^n$, the decision function of the holistic classifier can be defined as

$$H(\mathbf{x}) = \beta + \mathbf{w}^T \cdot \mathbf{x}$$

where \mathbf{w} is the weighting vector, and β is the constant bias of the learnt hyperplane. Motivated by its success, we also propose to use the linear SVM as the learning algorithm for the base classifiers

The HOG descriptor was proposed for human detection. Since then, the descriptor has grown in popularity due to its success. These features are now widely used in object recognition and detection. They describe the body shape through a dense extraction of local gradients in the window. Usually, each region of the window is divided into overlapping blocks where each block is composed of cells. A histogram of oriented gradients is computed for each cell. The final descriptor is the concatenation of all the blocks' features in the window. The LBP descriptor proposed first has been successfully used in face recognition and human detection. These features encode texture information. In order to compute the cell-structured LBP descriptor, the window is divided into overlapping cells. Then, each pixel contained in a cell is labelled with the binary number obtained by thresholding its value to its neighbour pixel values. Later, for each cell a histogram is built using all the binary values obtained in the previous step. Finally, the cell-structured LBP is the result of concatenating all the histograms of binary patterns in such window. The HOG-LBP is the concatenation of both descriptors, HOG and LBP. These two descriptors complement each other, as they combine shape and texture information. Besides, this combination has been proven to outperform the original HOG descriptor. Note that in our case, we interpret every cell LBP as a block, thus a block HOG-LBP represents the concatenated block HOG and the cell LBP computed in the same region. Following the formulation the constant bias β can be distributed to each block B_i by using the training data. This technique allows the possibility to rewrite the decision function of the whole linear SVM as a summation of classification results. Then, using this formulation we can define the local classifiers as $h_i(B_i) = \beta_i + \mathbf{w}_i^T$

B_i where \mathbf{w}_i and β_i are the corresponding weights and distributed bias for each block B_i , respectively. By defining the local classifiers this way, no additional training per block is required.

Moreover, when computing the holistic classifier, the local classifiers are implicitly computed, which means that there is no extra cost. In this paper, instead of just using HOG features to infer whether there is a partial occlusion, we extend the process to rely on both, HOG and LBP features.

IV. CONCLUSIONS

In this work, we presented a general approach for human detection in still images with the presence of partial occlusion.

The method was based on a modified random subspace classifier ensemble. The method can be easily extended to other objects, and allows to incorporate other block-based descriptors. Two of the most acclaimed descriptors in the literature of the pedestrian detection—HOG and HOG-LBP were implemented. The linear SVM was used as the base classifier. We evaluated our approach on large datasets, INRIA data is considered a standard benchmark for human detection.

V. REFERENCES

- [1] D. Gerónimo, A. M. L'opez, 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, Jul. 2010.
- [2] P. Dollár, 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, Apr. 2012.
- [3] Z. Lin and L. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 604–618, Apr. 2010.
- [4] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, Jul. 2004.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, San Diego, CA, USA, 2005, pp. 886–893.
- [6] X. Wang, T. Han, and S. Yan, "An HOG–LBP human detector with partial occlusion handling," in *Proc. ICCV*, Kyoto, Japan, 2009, pp. 32–39.
- [7] Y. Xu, D. Xu, S. Lin, T. X. Han, X. Cao, and X. Li, "Detection of sudden pedestrian crossing for driving assistance systems," *IEEE Syst. Man Cybern. B, Cybern.*, vol. 42, no. 3, pp. 729–739, Jun. 2012.
- [8] S. Dai, M. Yang, Y. Wu, and A. Katsaggelos, "Detector ensemble," in *Proc. CVPR*, Minneapolis, Minnesota, USA, 2007, pp. 1–8.
- [9] B. Wu and R. Nevatia, "Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detector responses," *Int. J. Comput. Vision*, vol. 82, no. 2, pp. 185–204, Apr. 2009.
- [10] B. S. M. Enzweiler, A. Eigenstetter and D. M. Gavrila, "Multi-cue pedestrian classification with partial occlusion handling," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 990–997.
- [11] T. Gao, B. Packer, and D. Koller, "A segmentation-aware object detection model with occlusion handling," in *Proc. CVPR*, Colorado Springs, CO, USA, 2011, pp. 1361–1368.
- [12] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester, "Object detection with grammar models," in *Proc. NIPS*, Granada, Spain, 2011, pp. 442–450.
- [13] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
- [14] J. Marín, D. Vázquez, D. Gerónimo, and A. M. L'opez, "Learning appearance in virtual scenarios for pedestrian detection," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 137–144.
- [15] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian Manifold," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
- [16] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *Proc. ICCV*, 2009, pp. 24–31.