A Method to Detect Partially Occluded Humans in Still Images

Deepika R[1], Pavan G S[2]

[1,2]M.tech(Digital Electronics and Communication Engineering) , Dept. of Electronic and Communication Engineering , Tumkur , Karnataka.

Abstract—**This paper describes a general method to detect humans when the person is under partial occlusion in still images. The random subspace method (RSM) is chosen for building a classifier ensemble robust against partial occlusions. In this proposed method which does not require manual labelling of body parts, defining any semantic spatial components, or using additional data coming from motion or stereo. Moreover, the method can be easily extended to other object classes. The experiments are performed on three large datasets: the 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 experimental results show that our detector outperforms state-of-the-art approaches in the presence of partial occlusions, while offering performance and reliability similar to those of the holistic approach on non-occluded data. The datasets used in our experiments have been made publicly available for benchmarking.**

I INTRODUCTION

HUMAN DETECTION is the key part in the systems of human-centred image retrieval, visual surveillance, pedestrian detection, and gait recognition, home automation, robot sensing. 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.

In this paper, we focus on human detection in static images. Note that detecting humans in a static image (i.e., a single image or a video frame) is more challenging than in an image sequence. Given a static image, no motion information can be used to model the background and provide clues for object detection.

Most promising methods of the state-of-the-art rely on discriminative learning paradigms. Along this line, researchers have been mostly working on two different issues: extracting features [5]–[9], and classification through machine learning algorithms [5], [6], [10]–[13]. State-of-the-art approaches canbe divided into two groups: holistic, which rely on detecting the target as a whole, and part-based, which combine the detection of different parts the body . Holistic methods offer robustness with respect to illumination, background and texture changes, whereas part based methods are advantageous for different poses [3]. 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 [3].

As expected, detection in the presence of partial occlusions has sparked significant interest [7], [14]–[18]. 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.

ISSN: 2278 – 7798

Therefore, accurate detection in the presence of partial occlusion is of paramount importance when building driver assistance systems.

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 [7], [15], [16], [18]. Therefore, our aim is to introduce a general method for automatic, accurate and robust detection of human figures in the presence of partial occlusion.

The proposed approach brings several benefits: 1)the approach is generic, therefore applicable to any class of objects; 2) as the random subspace classifiers are trained in the original space, no further feature extraction is required; 3) the detection is done on monocular intensity images,

II RELATED WORK

Dai *et al.* [14] propose a part-based method for face and car detection. The method consists of a set of substructure detectors, each of which is composed of detectors related to the different parts of the object. The disadvantage of this method is that the different parts of the object need to be manually labelled in the training dataset, in particular, eight parts for face detection and seven parts for cars.

Wang *et al.* [7] propose a new scheme to handle occlusions. More concretely, the response at a local level of the histograms of oriented gradients (HOG) [6] descriptor is used to determine whether or not such local region contains a human figure. Then, by segmenting the binary responses over the whole Window, the algorithm infers the possible occlusion. If the segmentation process does not lead to a consistent positive or negative response for the entire window, an upper/lower-body classifier is applied. The drawback of this method is that it makes use of a pre-defined spatial layout that characterizes a pedestrian but not any other object class

Gao *et al.* [17] tackle occlusions by identifying and using in the training process, cells of pixels that belong to the object in the bounding box. This method outputs not just the detection, but also the inferred segmentation. However, this method requires the tedious task of manual labelling all the cells that belong to the object in the training set.

Girshick*et al.* [18] propose an extension of the deformable part-based detector [11] 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, different occlusion reasoning has to be defined.

III PROPOSED SYSTEM

1. Proposal Outline

We propose a system which handles partial occlusions and detects partially occluded people. We use a block based feature vector and this feature vector is classified by a two stage classifier. The first block feature vector is fed to the holistic classifier which classifies the object as not occluded or in the ambiguous range. If the block feature is in the ambiguous range then the feature vector is passed on to the next stage of the classifier which is the ensemble classifiers. The ensemble classifiers apply occlusion inference algorithm and determine if the block feature vector is classified as having partially occluded portions or not. The ensemble is applied only when partial occlusion is suspected by passing through a holistic classifier.

$$E(\mathbf{x}) = \sum_{k \in S} \omega_k g_k(\mathbf{x})$$

ISSN: 2278 – 7798

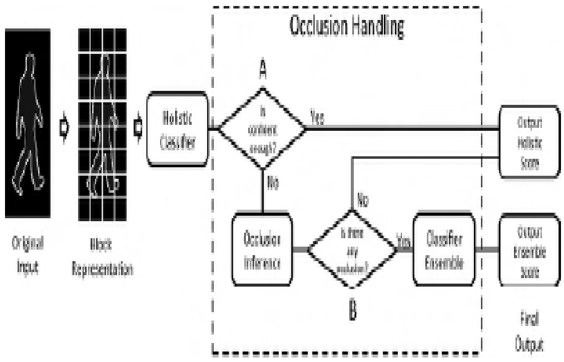

Fig 1.Occlusion handling Scheme.

2. Block Representation

The system uses a block-based representation of the image and is one of the mostpopular representation of feature vectors for detecting humans in images. A block is a subdivided portion of the image as shown below. Blocks can be overlapping also.

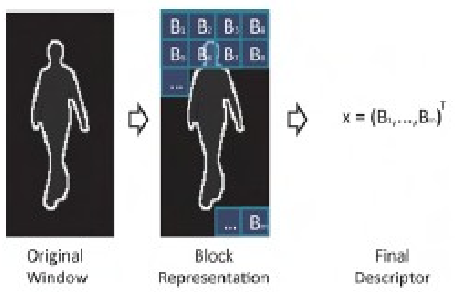

Fig.2 .Block-based representation. From left to right, the original input, then the division into blocks (note that Blocks can overlap), and finally, the feature descriptor

We extract the HOG feature vector for each block in the image and all feature vectors for all blocks in the image are concatenated to form the feature vector which is passed on to the holistic classifier. The higher the value output by the holistic classifier the higher is the confidence level that there is a pedestrian in the image. All values which fall in the ambiguous range of [-2, 1] are passed on to the next stage which is the occlusion inference stage. An SVM can be trained and used as the classifier in this case.

3.Occlusion inference and posterior reasoning

The occlusion inference process is applied when the output from the holistic classifier is ambiguous. In this case, each feature vector from each block is passed onto a local classifier which takes an input from the i^{th}block of the window and provides as output the likelihood that the block belongs to a pedestrian or is a part of the occluding object or background. The algorithm that is applied for classification of each block is shown in algorithm1.

If all the values after segmentation denoted by the symbols $\boldsymbol{s_1}'$ till $\boldsymbol{s_m}'$ are positive it means that there is no occlusion, if all are negative it means they are a part of background, part positive and part negative means that there is an occlusion. The segmentation algorithm used in the system is based on the mean shift algorithm.

Algorithm 1: Pseudo-code for occlusion inference.

Input: $\mathbf{B}_1, \ldots, \mathbf{B}_m$
Output: Found partial occlusion
Procedure:
foreach$i \in 1, \ldots, m$ do
Calculate $l_i(\mathbf{B}_i)$;
$s_i := sign(l_i(\mathbf{B}_i))$;
end
$(s'_1, \ldots, s'_m) := \mathrm{seg}(s_1, \ldots, s_m)$;
if $| \sum s'_I \neq m$ **then**
return true;// There are occluded blocks
else
return false;// Object or Background
end

ISSN: 2278 – 7798

Algorithm 2: Random subspace classifiers pseudo code.

Input: Training dataset $D = \{(\mathbf{x}_j, l_j) | 1 \leq j \leq n\}, T$
Output: $g_1, \ldots, g_T$
Procedure:
$I := \{1, \ldots, m\}$;
$J := \{\emptyset\}$;
$k := 1$;
while k ≤ T **do**
Randomly select a subset $J_k \subset I$ with $J_k \neq \emptyset$;
Given J_k generate the according $(r_1, \ldots, r_m)$;
$(r'_1, \ldots, r'_m)$:=seg$(r_1, \ldots, r_m)$;
Obtain J'_k from $(r'_1, \ldots, r'_m)$;
if $| \sum r'_l \neq m \wedge J'_k \in \mathbb{J}$ **then**
Train g_k in $D_k = \{(P'_k(\mathbf{x}_j), l_j) | 1 \neq j \neq n\}$;
$J := J \cup \{J'_k\}$;
$k := k + 1$;
end

4. Ensemble of Local Classifiers

Given all the blocks of the window, we choose at random only a subset of these blocks during an iteration. An individual classifier is trained using these block features as input. The process is repeated for different subsets of the blocks in the window. For each subset of the blocks we train an individual classifier. The binary image is constructed based on the blocks selected in the window and a segmentation algorithm is applied to the resulting binary image. The result of the segmentation is another binary image from which blocks that have are positive are chosen. We then check if the subspace defined by the new binary image can be defined, if it cannot be defined it implies that a classifier defined by this subspace has already been trained. We move on to the next subspace, else we train a new classifier defined by this subspace. We thus have a set of trained classifiers ranging from g_1 to g_T. The final ensemble is then decided based on the classifier selection. The best N classifiers out of these are chosen and a validation set is used to determine the subset of the classifiers that work well when combined together. The combined decision of the ensemble is defined as where the individual classifiers g_1 to $g_S$are weighted by a weight w_1 to $w_S$respectively .The algorithm for the subspace classifiers is given below.

III RESULTS

An Image is taken as input and the image is shown below in fig.3.1

Fig 3.1.this is an input image

Fig 3.2.this is an output image

Fig3.2 shows the final output where the human present in the image is represented by yellow rectangular block even when the human body is partially occluded by exterior object(vehicles in this case).

References

[1] M. Enzweiler and D. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, 2009.

[2] D. Ger´onimo, A. M. L´opez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEETrans.*

ISSN: 2278 – 7798

Pattern Anal. Mach. Intell., vol. 32, no. 7, pp. 1239–1258, Jul. 2010.
[3] P. Doll´ar, 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.
[4] 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.
[5] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, Jul. 2004.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, San Diego, CA, USA, 2005, pp. 886–893.
[7] 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.
[8] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 1030–1037.
[9] Y. Pang, H. Yan, Y. Yuan, and K. Wang, "Robust CoHOG feature extraction in human-centered image/video management system," *IEEESyst. Man Cybern. B, Cybern.*, vol. 42, no. 2, pp. 458–468, Apr. 2012
[10] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. CVPR*, Anchorage, Alaska, USA, 2008, pp. 1–8.
[11] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. CVPR*, Anchorage, AK, USA, 2008, pp. 1–8.
[12] Y. Xu, X. Cao, and H. Qiao, "An efficient tree classifier ensemble-based approach for pedestrian detection," *IEEE Syst. Man Cybern. B, Cybern.*, vol. 41, no. 1, pp. 107–117, Feb. 2011.
[13] 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.
[14] S. Dai, M. Yang, Y. Wu, and A. Katsaggelos, "Detector ensemble," in *Proc. CVPR*, Minneapolis, Minnesota, USA, 2007, pp. 1–8.
[15] 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.
[16] 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.
[17] 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.
[18] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester, "Object detection with grammar models," in *Proc. NIPS*, Granada, Spain, 2011, pp. 442–450.
[19] 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.
[20] J. Marín, D. V´azquez, D. Ger´onimo, and A. M. L´opez, "Learning appearance in virtual scenarios for pedestrian detection," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 137–144.
[21] 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.
[22] 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.
[23] L. I. Kuncheva, J. J. Rodriguez, C. O. Plumpton, D. E. J. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 531–542, Feb. 2010.
[24] D. Partridge and W. B. Yates, "Engineering multiversion neural-net
systems," *Neural Comput.*, vol. 8, no. 4, pp. 869–893, May. 1995.
[25] T. Ojala, M. Pietik¨ainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
[26] T. Ahonen, A. Hadid, and M. Pietik¨ainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. PatternAnal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
[27] Y. Yu, J. Zhang, Y. Huang, S. Zheng, W. Ren, K. Huang, and T. Tan, "Object detection by context and boosted HOG-LBP," in *Proc. PASCALVisual Object Challenge Workshop, Proc. Eur. Conf. Comput. Vision*, Crete, Greece, 2011.
[28] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May. 2002.

ISSN: 2278 – 7798