SIViP (2017) 11:1181–1188
DOI 10.1007/s11760-017-1073-z

ORIGINAL PAPER

Hierarchical detection of persons in groups

Álvaro García-Martín[1] · Ricardo Sánchez-Matilla[1] · José M. Martínez[1]

Received: 2 October 2015 / Revised: 12 January 2017 / Accepted: 10 February 2017 / Published online: 21 February 2017

Abstract In this paper, we address one of the most typical problems of person detection: scenarios with the presence of groups of persons. In this kind of scenarios, traditional person detectors have difficulties as they have to deal with several simultaneous occlusions. In order to try to solve this problem, we propose the use of two different hierarchies. The first one consists of a hierarchy of persons, i.e., the use of the detection of different persons belonging to a group in order to refine the individual's detections. The second one consists of a hierarchy of parts, i.e., the use of different combinations of body parts in order to refine the final detections. Experimental results over several video sequences show that the proposed hierarchies significantly improve the results with respect to different approaches from the state of the art.

Keywords Person detection · Hierarchy of persons in groups (HPG) · Hierarchy of body parts (HBP) · Hierarchical detector in groups (HDG)

1 Introduction

The detection of persons in video sequences is one of the most difficult and interesting challenges that we have to deal with in video analysis tasks. The main issue lies in the complexity for modeling persons due to their great variability in physical appearance, poses, movements and interactions with other persons and objects. This complexity grows in real-world settings, such as shopping malls, streets, railway stations, etc., since there are large groups of persons and many occlusions among them. In the method proposed in this paper, named hierarchical detector in groups (HDG), we use an individual person detector from the state of the art but enhanced with two new hierarchical structures both for the number of persons in the group and structures of body parts that constitute each person. Consequently, our method provides greater extensibility without requiring any specific group model training.

A canonical person detection approach from the state of the art [7] can be divided into three main stages. The first stage is in charge of the extraction of initial hypotheses or candidates to be a person, usually addressed via background subtraction, a simple and powerful technique but with significant limitations in complex (e.g., crowded) scenarios. Other techniques such as exhaustive search are more robust to rotation, scale changes and variety of poses, even in complex environments, but they add complexity and increase the probability of false-positive detections, besides being rather more computationally expensive. The second stage deals with the matching, evaluating the similarity between the initial candidates and a model designed and trained according to certain key parameters such as movement, size, shape, etc. Finally, the classification or verification stage decides if the candidates correspond or not to a person.

Starting from the individual person detector proposed by Felzenszwalb et al. [3], we incorporate, following [9,16,17], the idea of creating a hierarchy based on the physical location of individuals or different objects in order to associate them in pairs, triplets or larger groups. From [5], we also consider different combinations of parts and let them not to score in the center of the person but at any point which may suit better. From [6], we have used the idea of considering different body parts configurations, but we have included an additional hierarchy of body parts taking into account not only different body parts configurations but also the relationships between

✉ Álvaro García-Martín
alvaro.garcia@uam.es

[1] Universidad Autonoma de Madrid, Madrid, Spain

the occluding and occluded body parts between different persons.

Some works such as [1,12,13,19], propose the use of tracking information in order to deal with the occlusion problems. Other works such as [8,11,14,18] are more related with our work, as they try to find a solution to the problem of detecting individuals located in crowded scenes with lack of visibility of its parts. Nevertheless, in [11,14,18] the authors try to solve this key issue by creating a model and, thus, training it on the patterns formed when two persons are very close, or even overlapping. On the other side, [8] requires an individual person detector and a posterior global occlusion optimization process without using any kind of group structure reasoning. In our algorithm, we use the same model of an individual as in [3], but we use hierarchical structures both for the number of persons forming the group and body parts that forms each person, without modifying the original model. Consequently, our algorithm has wider and simpler configuration settings that provide greater flexibility without requiring any specific training on groups models because we are able to define any kind of group. Therefore, while [8] deals with crowds but without taking advantage of the group structure and [11,14,18] only deal with couple detection, our approach deals with any possible group configuration around each person (pairs, triplets or larger groups) without any kind of additional model training. We solve the problem in real scenarios, where the range of possible different occlusions is much bigger and, therefore, the complexity of the group model and its training increase exponentially.

The rest of the paper keeps the following structure: in Sect. 2, the modifications proposed over the base algorithm are described; in Sect. 3, the evaluation of the proposed approach is described; and finally, in Sect. 4 we summarize the conclusions and the future work.

2 Hierarchical detector in groups (HDG)

The original detector [3] defines an individual person model. We propose two new additional different hierarchies in order to deal not only with individuals but also with groups of persons. The approach has been split in several stages: hierarchy of persons in groups, hierarchy of body parts, calculation of confidence maps, bounding box detection and post-processing.

2.1 Hierarchy of persons in groups (HPG)

The purpose of the first hierarchy is to improve the detection of persons, who are hardly detected in the original algorithm because they are hidden by the other person, using the information of the least occluded person who defines the group. This hierarchy is defined by two types of persons: the Main Person (MP) and the Secondary Person (SP). The MP corresponds to the original individual person model [3], i.e., the least occluded person from which the rest of persons can be detected. The SP corresponds to an additional person of the group, partially occluded by the MP and/or another person (included or not in this group). Each SP_i is defined by the relative position of each possible SP_i, $(i = 1 \ldots, I)$, in relation to the MP. We define this relative position or anchor-shift with a two-dimensional vector $(\Delta x_i, \Delta y_i)$. The objective of using a hierarchy of persons is to gather all the pieces of information about every pair who belongs to the group of the MP in the geometric center of the MP. Thus, the persons who are hardly detected can improve their scores when using the information of other person who belongs to the pair. The system is fully flexible: we could define any search area of the SP. For example, following the state of the art in "double-person" (or couples) detectors [14,18], we consider possible occlusions even higher than 80%. After testing different anchor-shifts configurations (see Sect. 3.1), in this work, we have defined a set of anchor-shifts of nine positions $(I = 9)$ with $|\Delta x| \leq 6$ and $|\Delta y| \leq 6$ with step $s = 6$ (i.e., $\Delta x = -6, 0, +6$, $\Delta y = -6, 0, +6$). Therefore, we are able to detect up to nine different kinds of pairs around each person, the absence of displacement $(\Delta x = \Delta y = 0)$ and the eight neighbors (see Fig. 1). The anchor-shifts are placed in the horizontal axis at -6 and $+6$, as these values correspond to a displacement in which at least half body of the SP is occluded by the MP (assuming a full body person model). The anchor-shifts are placed in the vertical axis at -6 and $+6$, as these values correspond to the size of the head/shoulders in which at least eighty percent of the SP is occluded by the MP (assuming a full body person model). From the above, a range of search for SP from MP is defined that can be observed in Fig. 1. We also take into consideration the absence of displacement, or anchor-shift $(\Delta x = \Delta y = 0)$, to recognize a person who does not belong to any pair, i.e., the absence of a Secondary Person and therefore a single person or the original detector [3].

2.2 Hierarchy of body parts (HBP)

This extension tries to use the most relevant body parts of persons in presence of groups. Consequently, we have developed another hierarchy in relation to the person body parts configuration. The system is completely configurable: any combination is possible among all the possible ones that a person model of N parts allows. Following the previous assumption that the SP is more occluded than the MP, we have decided to use different sets of body parts for the MP and the SP. We have created five different settings for the MP where we discard the lower parts progressively, since they are the most occluded in presence of groups. All configurations use the original root body part both on the MP and the SP (see

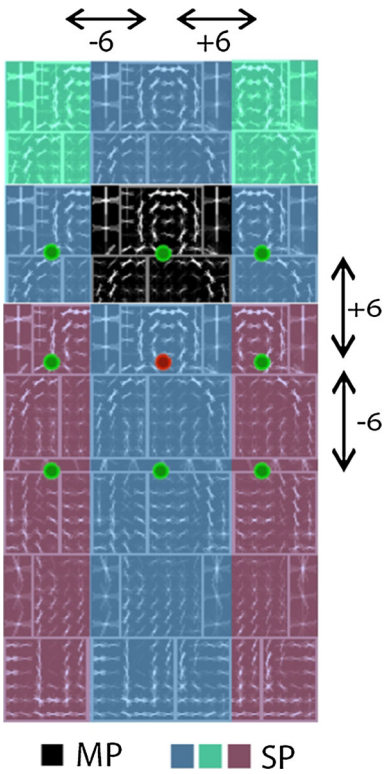

Fig. 1 Visual examples of MP and nine possible positions of the SP_i, ($i = 1, \ldots, I$). The figure only shows eight examples of SP_i: from $(-6, 6)$ to $(6, -6)$. The $(0, 0)$ position corresponds to a single person, the MP

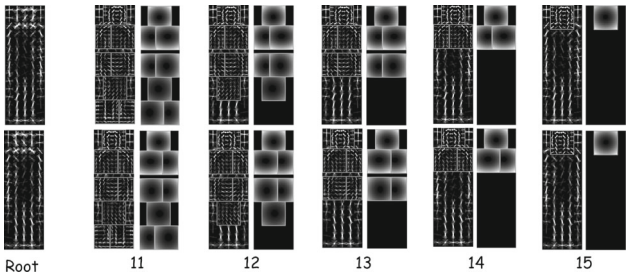

Fig. 2 Visual examples of body part configurations

Fig. 2). In Fig. 2, we can also see the five-model configuration used for the MP. In particular, the original detector corresponds to the use of only the first configuration, named "11" [3]. On the other hand, for each SP_i we use the same five configurations, but there are three variants which depend on the corresponding horizontal anchor-shift $(\Delta x_i, \Delta y_i)$: if $\Delta x > 0$, it corresponds to a horizontal shift to the right and assigns a model of SP with the same combination of body parts as the MP but only with the visible parts placed at the right side of the longitudinal axis of the person; if $\Delta x < 0$, it corresponds, consequently, to a SP with the visible parts placed at the left side of the longitudinal axis; finally, if $\Delta x = 0$, it corresponds to an only vertical displacement and it assigns a model of SP with up to three body parts: head and (two) shoulders parts. All different combinations of hierarchy of body parts have been tested with similar global results in the experimental dataset (see Sect. 3.2).

2.3 Calculation of confidence maps

Felzenszwalb et al. [3] scans exhaustively the entire image at multiple scales, gathering all the scores obtained from the N body parts and root (BP_n) at the geometric center (x, y) of the model for a given level or scale l, in a confidence map $C(x, y, l)$. Thus our MP follows the same process:

$$C_{MP}(x, y, l) = \sum_{n=0}^{N} \mathrm{BP}_n(x, y, l) \tag{1}$$

However, to implement the proposed hierarchy of persons in groups and detect individuals who are part of the group, the scores of the SP have to be redefined according with their relative position with respect to the MP. In our system, a high score indicates a high probability that the pixel corresponds to a detection of a second person for certain level l and a certain anchor-shift i. All the body parts, including the MP and the SP ones, are accumulated on the geometric center of the MP; therefore, each possible SP_i score is accumulated in the confidence map $C_{\mathrm{SP}}(x, y, l, i)$:

$$C_{\mathrm{SP}}(x, y, l, i) = \sum_{n=0}^{N} \mathrm{BP}_n(x', y', l) \tag{2}$$

$$(x', y') = (x + \Delta x_i, y + \Delta y_i) \tag{3}$$

where i indicate the anchor-shift displacement of the SP in relation to the MP.

Since we have also introduced a hierarchy of body parts, the scores of the MP and SP are obtained through two different configuration models (different sets of body parts) and, therefore, their confidence maps values present different ranges and, thus, they cannot be directly combined. Following [6], we are able to estimate the probability density function of each configuration model, normalize the corresponding confidence maps values $\tilde{C}_{\mathrm{MP/SP}}$ ($0 < \tilde{C}_{\mathrm{MP/SP}} < 1$) and, afterward, combine them. At this point, our algorithm continues like the original method. Calculating the maximum of all confidence maps for each level, we will obtain a unique map for each scale and anchor-shift $\mathrm{score}(x, y, l, i)$:

$$\mathrm{score}(x, y, l, i) = \tilde{C}_{\mathrm{MP}}(x, y, l) + \tilde{C}_{\mathrm{SP}}(x, y, l, i) \tag{4}$$

2.4 Bounding box detection

In our case, each final detection indicates that at least a pair has been detected as part of the group and, consequently, the point $p(x, y, l, i)$ is the center pixel of the MP of the

Table 1 HPG results over sequence S1L1-1

Anchor-shift configuration	*A*	*B*	*C*	*D*	*E*	*F*
$\|\Delta x\|$	≤6	≤6	≤12	≤12	≤12	≤12
$\|\Delta y\|$	≤6	≤6	≤6	≤6	≤12	≤12
s	3	6	3	6	3	6
I	25	9	45	15	81	25
lHPG	0.729	0.726	0.728	0.726	0.726	0.725

Anchor-shift configurations include groups from 9 to 81 persons (I) according to the maximum displacement in each direction, $\max(|\Delta x|)$ and $\max(|\Delta y|)$ and the step s between two consecutive SP_i

group. The MP bounding box is generated following the original algorithm, i.e., a bounding box is generated around the detected center (x, y) with the corresponding re-scaled size of the basic model at the detected scale l. Then, using the bounding box of the MP as reference, we extrapolate the bounding box of each SP_i according to the detected anchor-shift i.

2.5 Post-processing

Felzenszwalb et al. [3] avoids multiple detections of the same person using a non-maximum suppression (NMS) process: bounding boxes with an overlap higher than 50% are deleted, remaining only the one with higher score. It has been necessary to modify the NMS process, as with the original one we could lose most of the detections, taking into account that in this kind of videos persons are very close and the overlap is often greater than 50%. Again, according to the state of the art in "double-person" (or couples) detectors [14,18], we consider possible occlusions even higher than 80%. Therefore, we have divided the NMS process in two stages: firstly, we delete all the detections in which the head is minimally overlapped (more than 0%). Therefore, if we have two overlapped head detections, it is because they are probably two detections of the same person. Afterward, we perform the normal NMS process but allowing a greater overlapping than in the original algorithm. During the algorithm design process, we found out that the tolerance value with good performance is around 90% of overlapping. Any possible overlapping between 50 and 99% has been tested with similar global results (see Sect. 3.3).

3 Evaluation

We evaluate our approach on 10 challenging, publicly available video sequences with a ground truth [13] that includes all detections, even when persons are strongly occluded. The first 8 sequences: S1L1 (1 and 2), S1L2 (1 and 2), S2L1, S2L2, S2L3 and S3MF1, are from the PETS2009 database [4] and the last 2 sequences: Campus and Crossing, are from the TUD database [1]. Note that the PETS scenarios include higher complexity in terms of number of persons and occlusions than the TUD dataset, which is traditionally used for couple detection [14,18]. Following the experimental evaluation of [6], we classify the sequences according to the degree of occupation of the scene (low, medium or high). The evaluation metrics are the Precision–Recall curves and the area under the curve (AUC-PR), which is used to condense the algorithm performance in a single value.

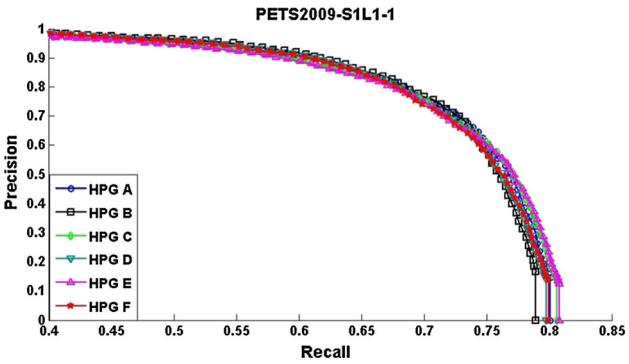

Fig. 3 Precision–Recall curves of S1L1-1 sequence using different anchor-shifts configurations

Firstly, we present exhaustive results of each stage of the approach (HPG, HBP and post-processing) over the sequence S1L1-1 (medium complexity). After that, we present the final results and a comparison with the state of the art.

3.1 Hierarchy of persons in groups (HPG)

In this section, we evaluate the first stage of the proposed approach, HPG, analyzing the results using different anchor-shifts configurations. The HBP and post-processing settings have been fixed as the final approach (HBP with configurations 12, 13 and 14, post-processing of 90% NMS); in particular, according to the maximum displacement in each direction, $\max(|\Delta x|)$ and $\max(|\Delta y|)$ and the step s between two consecutive SP_i. In Table 1 and Fig. 3, we can see six of the most representative anchor-shift configurations (named from A to F), including groups from 9 to 81 members. In general, the results show similar performance around 0.72–0.73 AUC. For efficiency and performance reasons, the

Table 2 HBP results over sequence S1L1-1

Configuration	11	12	13	14	15	11–15	12–14	12,13	12,14	13,14
HBP	0.690	0.725	0.727	0.729	0.684	0.719	0.726	0.725	0.727	0.728

All the single body parts configuration versions and five of the most representative combinations of them

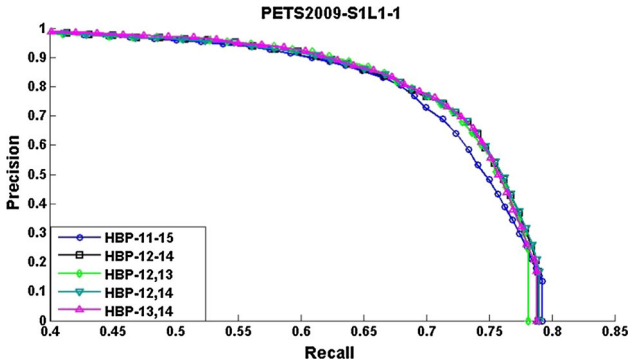

Fig. 4 Precision–Recall curves of S1L1-1 sequence using different body parts configurations from *A* to *F*

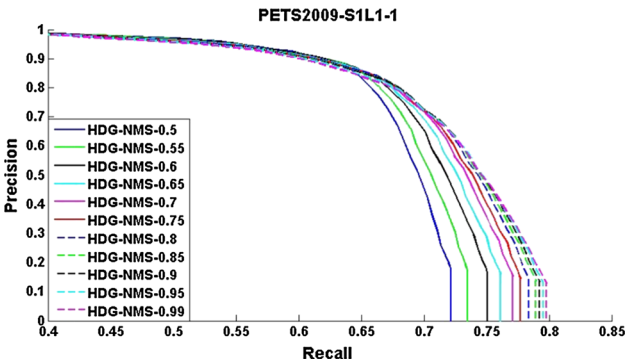

Fig. 5 Precision–Recall curves of S1L1-1 sequence using different non-maximum suppression overlap criteria

selected default setting corresponds to configuration B, previously described in Sect. 2.1, i.e., $I = 9$ with $|\Delta x| \leq 6$ and $|\Delta y| \leq 6$ with step $s = 6$ (see Fig. 1).

3.2 Hierarchy of body parts (HBP)

In this section, we evaluate the second stage of the proposed approach, HBP. The HPG and post-processing settings have been fixed as the default setting (HPG with $I = 9$; $|\Delta x| \leq 6$ and $|\Delta y| \leq 6$ with step $s = 6$, post-processing of 90% NMS). We present the results using different body parts configurations, from the single configuration versions 11, 12, 13, 14 and 15 to five of the most representative combinations of them. In Table 2, we can see the results for each configuration. In general, the results show similar performance around 0.68–0.73 AUC. In particular, in Fig. 4 we can see how the use of any combination including the configurations 12 or 14 and not 11, has almost the same performance. The selected default setting is a HBP with configurations 12, 13 and 14.

3.3 Post-processing

In this section, we evaluate the last stage of the proposed approach, the post-processing or non-maximum suppression overlap criteria. The HGP and HBP settings have been fixed to default configuration (HPG with $I = 9$; $|\Delta x| \leq 6$ and $|\Delta y| \leq 6$ with step $s = 6$, HBP with all configurations 11, 12, 13, 14 and 15). In Table 3 and Fig. 5, we can see eleven different non-maximum suppression overlap criteria (from 50 to 99% allowed overlap). Our approach HDG has been designed in order to support occlusions even higher than 80%. In general, the results show similar performance but in particular any overlap higher than 80% shows almost the same performance around 0.71–0.72 AUC. The selected default value for the non-maximum suppression overlap criteria is 90%.

3.4 Hierarchical detector in groups (HDG)

We have compared the HDG results with 7 person detectors from the state of the art: the original discriminatively trained deformable part-based detector (DTDP) [3], the aggregate channel features (ACF, Inria and Caltech variations) [2], the implicit shape model (ISM) [10], a multi-configurations body part variation of the DTDP in order to deal with groups (MC-DTDP) [6], the faster regions with convolutional neural network version (FRCNN) [15] and the human detection in dense crowds (HDDC) [8].

Firstly, in order to quantify which part of the improvements have been obtained due to hierarchy of persons in groups and which by the hierarchy of body parts, we have compared the original DTDP and the proposed hierarchy of

Table 3 HDG results with different non-maximum suppression overlap criteria (from 50 to 99% overlap) over sequence S1L1-1

NMS	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.99
HDG	0.677	0.686	0.696	0.702	0.708	0.712	0.716	0.718	0.719	0.719	0.718

Table 4 Comparative results for each hierarchy versus original detector

Sequence	Complexity	DTDP [3]	HBP	**HDG**	$\%\Delta_1$	$\%\Delta_2$	$\%\Delta_1 - \%\Delta_2$
S1L1-1	Medium	0.628	0.685	**0.726**	15.6	9.1	6.5
S1L1-2	Medium	0.734	0.816	**0.848**	15.5	11.2	4.3
S1L2-1	High	0.479	0.568	**0.679**	41.8	18.6	23.2
S1L2-2	High	0.494	0.598	**0.653**	32.2	21.1	11.1
S2L1	Low	0.934	**0.951**	0.949	1.6	1.8	−0.2
S2L2	Medium	0.664	0.764	**0.809**	21.8	15.1	6.7
S2L3	High	0.558	0.661	**0.738**	32.3	18.5	13.8
S3MF1	Low	0.930	**0.942**	**0.942**	1.3	1.3	0.0
Campus	Low	0.765	0.759	**0.791**	3.4	−0.8	4.2
Crossing	Low	0.854	0.854	**0.855**	0.1	0.0	0.1
Average		0.704	0.760	**0.799**	16.6	9.6	7.0

The best results in each sequence are in bold. Percentage increase HDG versus DTDP ($\%\Delta_1$). Percentage increase HBP versus DTDP ($\%\Delta_2$)

Table 5 State-of-the-art person detection performance

Sequence	DTDP [3]	ACF [2]		ISM [10]	MC-DTDP [6]	FRCNN [15]	HDDC [8]	**HDG**	$\%\Delta_3$
		Inria	Caltech						
S1L1-1	0.628	0.640	0.648	0.453	0.660	0.607	0.645	**0.726**	10.0
S1L1-2	0.734	0.686	0.823	0.491	0.794	0.714	0.740	**0.848**	3.0
S1L2-1	0.479	0.447	0.553	0.296	0.560	0.512	0.566	**0.679**	20.0
S1L2-2	0.494	0.519	0.580	0.359	0.568	0.546	0.575	**0.653**	12.6
S2L1	0.934	0.858	0.932	0.779	**0.949**	0.931	0.910	**0.949**	0.0
S2L2	0.664	0.582	0.741	0.552	0.745	0.751	0.717	**0.809**	7.7
S2L3	0.558	0.478	0.601	0.340	0.615	0.568	0.603	**0.738**	20.0
S3MF1	0.930	0.944	0.940	0.820	0.949	**0.968**	0.907	0.942	−2.7
Campus	0.765	0.809	0.751	0.761	0.761	**0.811**	0.759	0.791	−2.5
Crossing	0.854	**0**.880	0.834	0.843	0.854	0.816	0.823	0.855	−2.8
Average	0.704	0.684	0.740	0.569	0.746	0.722	0.725	**0.799**	6.6

Percentage increase HDG versus the best individual detector results in each sequence $\%\Delta_3$. The best results in each sequence are in bold

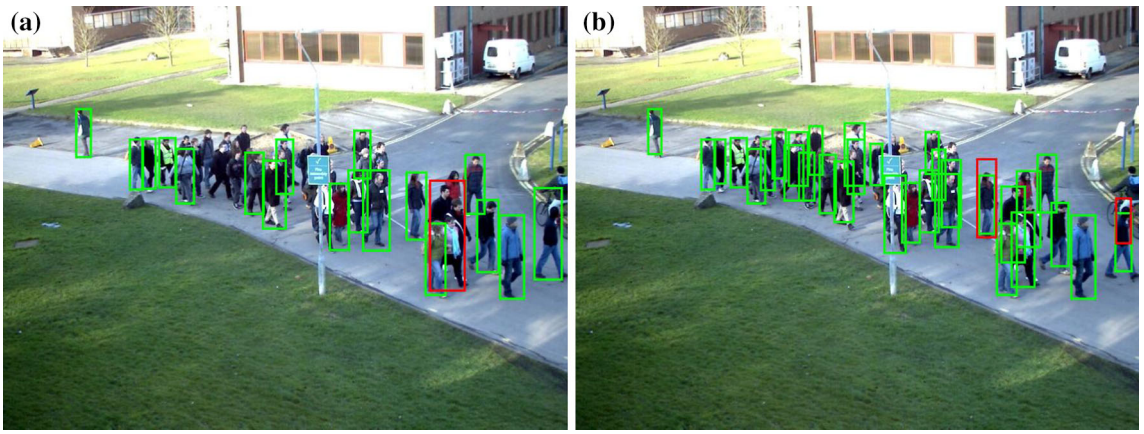

Fig. 6 Visual example of true positive detections (*green color*) and false detections (*red color*): **a** original DTDP detector [3] and **b** HDG detector (color figure online)

body parts (only HBP with configurations 12, 13 and 14) versus the final HDG results. Thus, we obtain the increase percentage due to each hierarchy. As the results show in Table 4, in general, the hierarchy of body parts (9.6% average increase, $\%\Delta_2$) is more relevant than the hierarchy of persons in groups (7.0% average increase, $\%\Delta_1 - \%\Delta_2$), although in

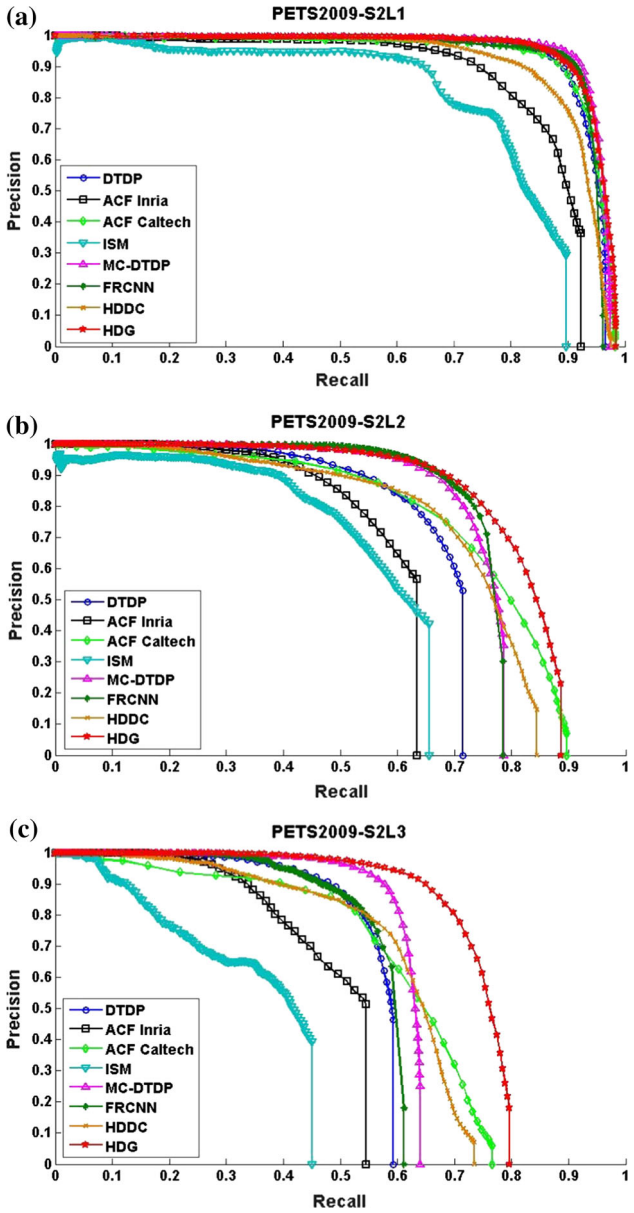

Fig. 7 Precision–Recall curves from three sequences which have different complexity: **a** PETS2009-S2L1—low, **b** PETS2009-S2L2—medium and **c** PETS2009-S2L3—high

the most complex videos the hierarchy of persons in groups has a greater importance, as this kind of scenes are its specific target. In Table 5, we show the AUC of each of the seven algorithms analyzed (Fig. 6). In more complex sequences, like PETS2009-S1L2-1, PETS2009-S1L2-2 or PETS2009-S2L3, the increase is in the range of $12-20\%$. On the other hand, in easier sequences like PETS2009-S2L1, PETS2009-S3MF1, TUD-Campus and TUD-Crossing, we achieve a smaller or even negative improvement. Our algorithm is developed to improve the results in complex scenarios with the presence of groups of persons, where there are large occlusions, and therefore, when these circumstances do not occur, the benefit is limited. Nevertheless, we would like to underline that our algorithm always improves the original performance (16.6% average increase, $\%\Delta_1$). In Fig. 7, we show the Precision–Recall curves from three sequences which have different complexity (low, medium and high, respectively): PETS2009-S2L1, PETS2009-S2L2 and PETS2009-S2L3. In Fig. 6, we show examples of true positive detections (green color) and false detections (red color) of the original DTDP and our HDG detector. Both results have been obtained for the same detection score; while the original detector is only able to detect a few persons with partial occlusions, our HDG detector is able to detect many persons with partial occlusions in presence of groups.

4 Conclusions and future work

Our main goal consists of detecting the occluded persons in groups who are usually not detected. To achieve this goal, we have proposed a hierarchy of persons in groups, where the detection of the most visible person could help to detect the occluded ones, and a hierarchy of body parts, which main principle is to use the body parts with most useful information. The algorithm has been evaluated and compared with the state of the art. The results show how our approach has the best results in videos with a higher density of persons or complexity, where there are strong occlusions. In easier sequences, we have also achieved a slight improvement over the original approach. The proposed approach is flexible: we could set the number of persons who conform a group, defining different anchor-shifts, and we could set the body parts which form a person. Therefore, as future work, we will study the behavior of the hierarchy on larger groups of persons, evaluating exhaustively the different anchor-shift ranges and steps. In relation to the hierarchy of body parts, we would like to study other body parts configurations or even to use the person model without root in the SP, since in the presence of strong occlusions it could be counterproductive.

Acknowledgements This work was partially supported by the Spanish Government (HAVideo, TEC2014-53176-R).

References

1. Andriluka, M., Roth, S., Schiele, B.: People-tracking-by-detection and people-detection-by-tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
2. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012)
3. Felzenszwalb, P., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2010)

4. Ferryman, J., Shahrokni, A.: Pets: dataset and challenge. In: Proceeding of PETS-Winter (2009)
5. Garcia-Martin, A., Cavallaro, A., Martinez, J.M.: People-background segmentation with unequal error cost. In: Proceeding of ICIP, pp. 157–160 (2012)
6. Garcia-Martin, A., Evangelio, R.H., Sikora, T.: A multi-configuration part-based person detector. In: International Conference on Signal Processing and Multimedia Applications (SIGMAP). IEEE, pp. 321–328 (2014)
7. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man. Cybern. C (Appl. Rev.) **34**(3), 334–352 (2004)
8. Idrees, H., Soomro, K., Shah, M.: Detecting humans in dense crowds using locally-consistent scale prior and global occlusion reasoning. IEEE Trans. Pattern Anal. Mach. Intell. **37**(10), 1986–1998 (2015)
9. Lee, B., Erdenee, E., Jin, S., Rhee, P.K.: Efficient object detection using convolutional neural network-based hierarchical feature modeling. Signal Image Video Process. **10**(8), 1503–1510 (2016)
10. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. IEEE Comput. Vis. Pattern Recognit. **1**, 878–885 (2005)
11. Li, B., Song, X., Wu, T., Hu, W., Pei, M.: Coupling-and-decoupling: a hierarchical model for occlusion-free object detection. Pattern Recognit. **47**(10), 3254–3264 (2014)
12. Liu, Q., Ma, X., Ou, W., Zhou, Q.: Visual object tracking with online sample selection via lasso regularization. Signal Image Video Process. (2017). doi:10.1007/s11760-016-1035-x
13. Milan, A., Roth, S., Schindler, K.: Continuous energy minimization for multitarget tracking. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 58–72 (2014)
14. Ouyang, W., Zeng, X., Wang, X.: Single-pedestrian detection aided by two-pedestrian detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1875–1889 (2015)
15. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of NIPS (2015)
16. Sadeghi, M., Farhadi, A.: Recognition using visual phrases. In: Proceedings of CVPR, pp. 1745–1752 (2011)
17. Sadovnik, A., Chen, T.: Hierarchical object groups for scene classification. In: Proceedings of ICIP, pp. 1881–1884 (2012)
18. Tang, S., Andriluka, M., Schiele, B.: Detection and tracking of occluded people. Int. J. Comput. Vis. **110**(1), 58–69 (2013)
19. Vázquez, C., Ghazal, M., Amer, A.: Feature-based detection and correction of occlusions and split of video objects. Signal Image Video Process. **3**(1), 13–25 (2009)