The Combinator: optimal combination of multiple pedestrian detectors

Floris De Smedt*, Kristof Van Beeck*, Tinne Tuytelaars and Toon Goedemé
EAVISE, ESAT-PSI-VISICS,
Campus De Nayer, Katholieke Universiteit Leuven, Belgium
E-mail: {floris.desmedt, kristof.vanbeeck, tinne.tuytelaars, toon.goedeme}@esat.kuleuven.be

Abstract—In recent years, the accuracy of pedestrian detectors significantly improved. Currently, state-of-the-art pedestrian detectors achieve high accuracy results on challenging datasets. As opposed to refining a single detector, in this paper we propose a different approach to further increase the detection accuracy: combining multiple pedestrian detectors. The most straight-forward way to combine pedestrian detectors would be a naive AND or OR combination. Here, we present a novel generic combination framework in which we exploit specific information from each pedestrian detector to determine the optimal combination parameters. Our main motivation for this approach is based on the fact that several pedestrian detection approaches are based on very different techniques (e.g. a different feature pool), and thus an efficient combination should yield higher accuracy results. Indeed, such a combination is far more powerful, and our experiments indicate that specific (that is, cleverly chosen) combinations outperform existing state-of-the-art pedestrian detection results.

I. INTRODUCTION

Pedestrian detection is an active research topic in recent years. Indeed, state-of-the-art algorithms achieve excellent accuracy results on challenging datasets (e.g. INRIA [5], Caltech [11]). As opposed to the optimisation of a single pedestrian detector (e.g. *Roerei* [1]), we propose a different approach to increase the detection accuracy: combining pedestrian detectors. Take for example the left frame of figure 1. Here, the detections of three different pedestrian detectors are visualised. Note that none of them manages to find all pedestrians, and all yield false detections. The optimal combination of all the different detection results we propose in this paper yields the rightmost result, where all pedestrians are detected with minimal error. We therefore address a fundamental question: how should we combine the detection results of multiple pedestrian detetectors to allow for a higher accuracy rate? Seeking such a strong combination rule is not a trivial task, since many design choices are to be considered in this process. This paper tackles these questions, and presents a generic framework to achieve these pedestrian detector combinations. Traditionally, combining multiple pedestrian detectors is performed using a basic *AND* or *OR* rule. Our framework utilises a more profound approach, and exploits information from each pedestrian detector. In a nutshell: our framework combines the detection scores from multiple pedestrian detectors using a weighted sum to calculate the final detection results scores. For that, we propose a measure for the complementarity between different detectors and the confidence of a detector in order to determine the optimal combination to further increase the
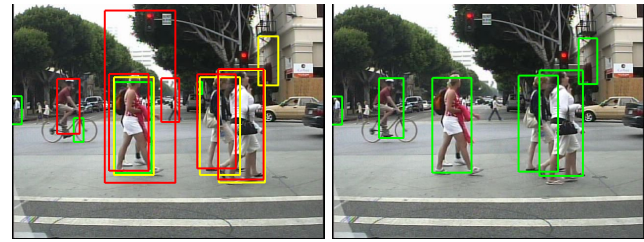

(a) Input image (b) Output of our approach

Fig. 1. An example of how a combination of pedestrian detectors yields higher accuracy results. (Left) Red: LatV4-cc, Green: HOG, Yellow: ChnFtrs (Right) Green: output detection of our approach

accuracy. One could argue that, to determine the optimal weights, a machine learning strategy (e.g. Support Vector Machines) could be utilised. This is difficult since pedestrian detectors do not provide a detection probability score for all positions in the image. To achieve this, the pedestrian detection algorithms themselves should be adapted. However, since many pedestrian detection algorithms use completely different approaches, determining such a probability score in a fair way is an impossible task. Our approach allows to combine the detection results from all detectors *out-of-the-box* and, as we show further on, still manages to find the optimal combination weights. Note that our approach is applicable to arbitrary object detectors: here we use pedestrian detectors as an example application.

Our main contributions are two-fold:

- We propose a generic methodology that allows for an efficient combination of an arbitrary number of pedestrian detectors.

- We perform thorough experiments, and propose combinations that achieve better than state-of-the-art accuracy results.

Increasing the number of pedestrian detectors evidently increases the calculation time. A compromise needs to be determined between the accuracy and the computational complexity. This paper focusses on accuracy improvement only. However, recent advances concerning speed improvements on existing pedestrian detectors show promising results. Fast multi-core implementations (GPGPU or multi-core CPUs) currently achieve reasonable to excellent processing speeds.

The remainder of this paper is structured as follows. Section II describes related work on this topic. In section III we

---

* F. De Smedt and K. Van Beeck contributed equally to this paper.

IEEE computer society

propose our combination approach and give detailed information on our combination parameters, followed by experimental results in section IV. Finally, in section V we present our conclusions and future work.

II. RELATED WORK

Object detection in general is a challenging task due to the wide variety in both objects and backgrounds. Many computer vision applications (e.g. traffic, surveillance, industrial automation, ...) rely on efficient object detection algorithms. Viola and Jones [19] proposed the use of *Haar features* as weak classifiers for AdaBoost, and combined with their cascaded classifier approach they achieve efficient object detection. Concerning pedestrian detection, Dalal and Triggs proposed the use of Histograms of Oriented Gradients (HOG) [5]. Since then, both the accuracy and speed of object detection algorithms steadily increased. However, even today their work is used as a reference in new benchmarks.

To increase the accuracy, essentially two different approaches were proposed. A first approach is the use of a more complex pedestrian model, of which the work of Felzenszwalb et al. [13] is only one example. They proposed the inclusion of extra parts in the HOG model to increase the accuracy, resulting in their *Deformable Part Models* (DPM). Aside from increasing the complexity of the model, a second approach exploits the use of other features (e.g. color) besides the standard gradient features. Such work is presented in [10], which the authors coined *Integral Channel Features* (furthed referred as ChnFtrs). More recently, based on these features Benenson et al. proposed their *Roerei* detector [1]. They achieve state-of-the-art accuracy results by optimising each individual stage in the detection process. To facilitate benchmark tests, Dollár et al. published an evaluation framework ([11], [12]) in which accuracy results for many state-of-the-art pedestrian detectors are available (currently more than 30), allowing for a fair comparison. Here, we use this framework as a baseline for our combination framework.

Besides an improvement in accuracy, decreasing the calculation time of pedestrian detectors is also an active research topic. Three approaches are commonly used: reducing the algorithmic complexity (e.g. approximate features), using an optimised implementation (e.g. GPUs, [7], [18]) or reducing the search space. Often a combination of all three approaches is utilised [8].

Several algorithmic optimisations have been proposed to speed-up detection. A similar cascaded approach as in [19] was applied to the DPM detector: a cascaded version is presented where initially only part of the model is evaluated to allow for a fast rejection, and the model is enriched later on [14]. Lampert et al. utilise a branch and bound framework to reduce the search space [16]. Dollár et al. proposed the *Fastest Pedestrian Detector in the West (FPDW)* in which only part of the feature pyramid was fully calculated, and intermediate layers were approximated from feature responses nearby [9]. Based on this approach Benenson et al. presented work in which model rescaling is performed: a model is trained for a number of layers, and intermediate models are approximated, thereby eliminating the need to construct a scale-space pyramid [3]. They further exploited this concept on GPU hardware, and

combined with their *stixel world approximation* [2], achieve pedestrian detection at 100Hz in their publicly available *Very-Fast* framework. The authors claim that their *Roerei* detector is compatible with this framework, thus enabling excellent accuracy results with reasonable processing speeds.

Approaches that reduce the dimensionality of the search-space often use object trackers, application specific information or a combination of both, thus increasing the detection throughput. In [4] the authors proposed work in which they utilise ground-plane assumptions to achieve real-time processing speeds based on the DPM detection model. Recently, in [8] a hybrid CPU/GPU implementation combined with a search-space reduction technique (coined the *warping window approach*) and temporal (tracking) information, reached up to 500 detections per second.

Although integration strategies are applied in other research domains, to the best of our knowledge only few works concerning the optimal integration of multiple object detectors exist. Most work in which multiple object detectors are combined only use naive *AND* or *OR* combination rules. E.g. [6] proposes the combination of a person model (DPM) and a face detector (V&J) to improve the detection results. A more intelligent approach is given in [15], where the authors present a probabilistic framework in which they combine the detection score of an object detector (DPM) with Conditional Random Fields (CRF) to achieve better scene understanding. However, they only use a single object detector. Recent work [17] uses multiple object models, based on the same detector (ChnFtrs), to cope with occlusion (new models are trained for a specific occlusion level). The different detection results are then in a weighted sum fashion combined into a single detection score (with the area of each model taken into account).

Our work significantly differs from all of the previously mentioned works. We propose an approach that aims to combine the detection results of multiple independent object (pedestrian) detectors in the most optimal manner. In essence, our goal is to determine the best possible combination rule to maximally increase the accuracy. We utilise specific information from each individual detector to obtain such an optimal combination rule. Furthermore our framework allows the combination of an arbitrary number of detectors. As shown in section IV, the accuracy of an optimal combination outperforms existing state-of-the-art pedestrian detectors. In the next section we propose our approach and motivate our design choices.

III. APPROACH

As mentioned above, currently only naive pedestrian detector combinations are used. In this work we try to increase the accuracy by combining the detection results of multiple pedestrian detectors in a more profound approach. That is, we propose to combine the detection scores from each individual pedestrian detector using a weighted sum. Our goal is then to find these optimal combination weights, such that they exploit the strenghts of each individual detector. For example, the DPM detector has excellent accuracy for high- to medium-resolution pedestrians, for which the accuracy of the HOG detector is lower; however small pedestrians can still be detected with HOG. Combining such detectors evidently yields
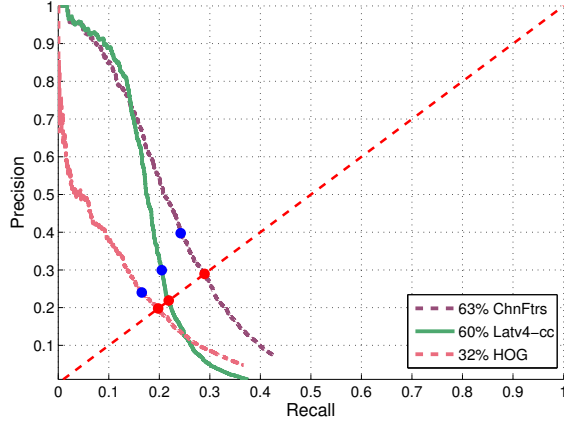
Fig. 2. The optimal thresholds (PR operating points). Blue: obtained using a fixed number of detection windows ($N = 50\%$). Red: using an equal error rate line.

better accuracy. The challenge now lies in the quantification of this information. Therefore, we propose the use of two different measures: *confidence* and *complementarity*.

- **Confidence**: The confidence value indicates how *good* a detector performs. It gives an indication about the probability of a detection by detector i being a correct detection (further indicated as $c_{\text{conf}(i)}$).

- **Complementarity**: Each pedestrian detector uses a specific design methodology (e.g. different feature pools or classifiers). This measure tries to indicate how *different* the detectors are (further referred as $c_{\text{compl}(i)}$).

We use these two measurements in a weighted sum to combine the detection scores from two or more pedestrian detectors (see subsection III-C). Below we explain in detail how each of these measurements are determined. They are calculated beforehand in an offline-stage based on detection results of these detectors on a training set.

Since each detector has specific tuning parameters (e.g. thresholds and non-maximum suppression) we need to determine an unbiased way to compute the confidence and complementarity coefficients over all detectors. To achieve this, we select a fair operating point (or detection threshold) on the Precision-Recall (PR) curve for each detector. We determined the optimal threshold using an *equal number of detection windows*. For each detector, we take the N highest scoring detection windows, calculate the precision and recall and retrieve the threshold for this point from the PR curve, giving the optimal operating point for this detector. Here, N is a fixed value for all detectors, determined as a percentage of the number of ground truth detections. This is visualised in figure 2 for three pedestrian detectors: *HOG*, *Channel Features (ChnFtrs)* and the cascaded *Deformable Part Models (LatV4-CC)* where the blue points indicate the optimal threshold setpoints (N is chosen as 50% of the number of ground truth detections). In this paper we use these three detectors as an example to illustrate our combination approach. An alternative method to retrieve the optimal operating points could be the use of a line determined by equal error rate. The intersection

with this line and the PR curve gives the optimal thresholds (visualised by the dotted red line and dots).

In the next subsections we give an overview of how each of the coefficients defined above are determined. Note that all further calculations in the subsections below use this optimal threshold, thus ignoring detections with a lower detection score. We then explain how this information is used to combine the detection output of multiple detectors to achieve a higher accuracy. Finally, this section concludes with a validation of our approach, showing that it manages to reach the most optimal solution.

A. Confidence coefficient

The confidence coefficient $c_{\text{conf}(i)}$ gives an indication of the detection accuracy of detector i. This information is independent of other detectors in the framework, and is based on the accuracy that a detector achieves on a specific dataset. Recent pedestrian detectors evidently reach a higher confidence coefficient. Several statistics can be used for this measure, e.g. the average precision (AP). We propose to use the area from the rectangle through the origin and the operating point on the PR curve for each detector i. Table I gives the confidence coefficients for our example detectors.

B. Complementarity coefficient

Different detectors use different design methodologies and feature pools. Therefore, each pedestrian detector reacts differently to a specific image patch. Combining pedestrian detectors that are complementary w.r.t. each other thus could yield better detection results. Our complementarity coefficient $c_{\text{compl}(i)}$ tries to indicate how *different* these pedestrian detectors react, and thus how complementary they are. When multiple detectors with very different detection approaches yield a detection at the same image location, the chance of that being a true detection increases significantly (much more than when e.g. multiple detections from rather redundant pedestrian detectors using the same approach are found).

As an example, the frame in figure 3 visualises three detector outputs (see caption for color coding). These detectors give significantly different detections. E.g. only HOG manages to find the small pedestrian on the left, and some detectors generate different false positives. Some locations are covered by more than one detector, indicating a higher probability that these are correct detections. If combined efficiently, an optimal accuracy is achieved. To determine a complementarity coefficient for each detector, we first calculate the pairwise complementarity score $w_{i,j}$ between two detectors i and j.

TABLE I. CONFIDENCE COEFFICIENTS FOR OUR THREE EXAMPLE DETECTORS.

	HOG	LatV4-CC	ChnFtrs
Confidence coefficient	0.040	0.061	0.096

TABLE II. COMPLEMENTARITY MATRIX FOR THREE DETECTORS

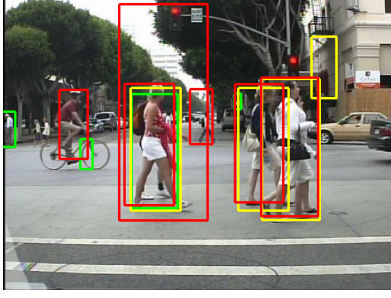|	HOG	LatV4-CC	ChnFtrs
HOG	0	0.4515	0.3641
LatV4-CC	0.4081	0	0.3481
ChnFtrs	0.3555	0.3811	0

Fig. 3. Example detections on Caltech frame. Red: LatV4-CC, Green: HOG, Yellow: ChnFtrs.

Fig. 4. The complementarity function $c_{\text{compl}(A)}$ for two detectors.

This is done as follows. We compare the detection performance over the Caltech dataset. For each frame, each detector is pairwisely compared. The number of detections from a specific detector which are not covered by the other detector (using the 50% intersection criterion of Dollár [11]) is determined. These are then summed over all frames, and divided by the total number of detections for that detector. For example (on one frame), in figure 3 *ChnFtrs* has four detections of which three are covered by *LatV4-CC*, thus $w_{ChnFtrs,LatV4-CC} = 25\%$. *HOG* covers one detection from *ChnFtrs* resulting in $w_{ChnFtrs,HOG} = 75\%$. If done for each detector pair, this results in a square complementarity matrix, visualised for three detectors in table II. Note that no annotation data is used; the fact that a detection is correct or not is irrelevant for the complementarity coefficient. This information is already included in our confidence coefficient. The complementarity coefficients aims to indicate how much extra information each specific detector introduces in the case of overlapping detections.

During the combination of the detection results, this complementarity matrix is used to calculate the complementarity coefficients of overlapping detections as follows. For overlapping detections, we first extract the corresponding square submatrix (containing only the *relevant* detectors - those that account for one of the detections) from the total complementarity matrix. Next we calculate a single *average complementarity coefficient* C_i for each detector i involved in this overlapping detection, using the individual pairwise complementarity scores $w_{i,j}$:

$$C_i = \frac{\sum_{j=1}^{n} w_{i,j}}{n-1} \quad (1)$$

Where n indicates the number of relevant detectors involved in the overlap. Intuitively, if $C_i = 0$ this detector is totally redundant and does not yield extra information, if $C_i = 1$ it is perfectly complementary w.r.t. the other relevant detectors. In a final step we use these averaged complementarity scores C_i from all individual relevant detectors to determine the final complementarity coefficient $c_{compl(i)}$ for a single detector i as follows. As a simplified case, take for example a combination of two detectors A and B which have overlapping detections, thus $n = 2$. If one of both detectors is completely complementary ($C_i = 1$) it yields valuable information and thus we set $c_{compl(i)} = 1$, independent of the other average complementarity coefficient. If both are equally complementary ($C_A = C_B$) they are given the same score
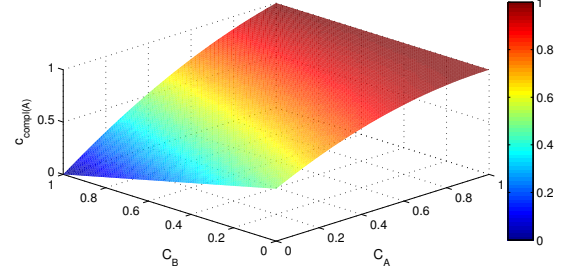
$c_{\text{compl}(A)} = c_{\text{compl}(B)}$. In the extreme case where both are completely redundant ($C_A = 0, C_B = 0$), this score equals $c_{\text{compl}(A)} = c_{\text{compl}(B)} = 1/2$. Intermediate values are calculated based on both average complementarity coefficients. For one of both complementarity coefficients (A) this yields:

$$c_{\text{compl}(A)} = \frac{1}{2}\left[1 + (C_A - C_B) + C_A C_B + C_A(1 - C_A)\right] \quad (2)$$

For clarification, the complementarity function $c_{\text{compl}(A)}$ for $n = 2$ is visualised in figure 4. For more than two detectors, this complementarity function generalises as follows:

$$c_{\text{compl}(i)} = \frac{1}{n}\left[1 + \sum_{\substack{j=1 \\ j \neq i}}^{n}(C_i - C_j) + \sum_{\substack{j=1 \\ j \neq i}}^{n}(C_i C_j) \right.$$
$$\left. + (\sum_{j=1}^{n} C_j - 1)(1 - C_i) + \prod_{j=1}^{n}(1 - C_j)\right] \quad (3)$$

Where C_i are the individual average complementarity coefficients and n indicates the number of relevant detectors involved in the overlap. These n complementarity coefficients thus summarise how each individual detector involved in this overlap should be weighted as such to maximally exploit the information potential of each specific detector.

To illustrate, suppose that an overlapping detection with both *HOG* and *ChnFtrs* is found ($n = 2$). First, we extract the 2×2 complementarity submatrix. Next, using eq. 1, we determine the average complementarity coefficient for each detector ($C_{HOG} = 0.45, C_{ChnFtrs} = 0.36$). Finally, we calculate the complementarity coefficient for each detector using eq. 2, yielding $c_{compl(HOG)} = 0.685$ and $c_{compl(ChnFtrs)} = 0.675$.

C. Combining detection results

Using the coefficients determined above, the actual combination is performed as follows. On each input image, all detectors are run separately. A first step consists of the normalisation of all detection scores, since the scores significantly differ for each detector. This is simply achieved using the standard score approach (subtract average and divide by standard deviation). Then an offset is added (the minimal score over all detections) to ensure that all detection scores are positive. Next, we determine where and which pedestrian detectors have an
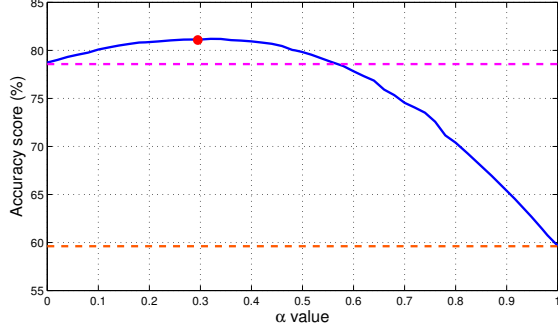
Fig. 5. The accuracy score (%) in function of the α value. Dotted lines indicate the accuracy of both individual detectors (*HOG*: Orange, *ChnFtrs*: Magenta). The red dot is the combination rule we propose, derived independently from this experiment.

Fig. 6. Precision-Recall curve of our combination approach for *ChnFtrs + LatV4-cc*, compared with the standard *AND* and *OR* combinations.

overlapping detection (using the 50% intersection criterion). In this case a final detection score is determined based on the output scores S_i of the n overlapping detections, the confidence coefficient and the complementarity coefficient of each detector i that yielded a detection there, calculated as mentioned above, using a weighted sum:

$$S_{\text{final}} = \sum_{i=1}^{n} c_{\text{conf}(i)} c_{\text{compl}(i)} S_i \qquad (4)$$

For the final bounding box we return the weighted average (using the individual detection scores) over the overlapping bounding boxes. For all non-overlapping detections we multiply the detection score with the confidence value of that detector and the complementarity coefficient, calculated as if this detection overlapped with all detectors. Our confidence and complementary coefficients are chosen in such a way that multiple detections from complementary detectors with high confidence return high scores, whereas redundant detectors with low confidence evidently output lower total detection scores. In section IV our experiments show that our approach achieves very good accuracy results.

D. Validation of our approach

In this section we validate our combination rule, and show that our weighted sum approach using the confidence coefficient and complementarity coefficient as defined above reaches the most optimal solution (i.e. the combination with the highest accuracy). This is done as follows. Take for example the combination of two detectors A and B. Our combination rule then becomes:

$$S_{\text{final}} = c_{\text{conf}(A)} c_{\text{compl}(A)} S_A + c_{\text{conf}(B)} c_{\text{compl}(B)} S_B \qquad (5)$$

Since only the relative weights are important, this can be reformulated as:

$$S'_{\text{final}} = \alpha S_A + (1 - \alpha) S_B \qquad (6)$$

Where:

$$\alpha = \frac{c_{\text{conf}(A)} c_{\text{compl}(A)}}{c_{\text{conf}(A)} c_{\text{compl}(A)} + c_{\text{conf}(B)} c_{\text{compl}(B)}} \qquad (7)$$
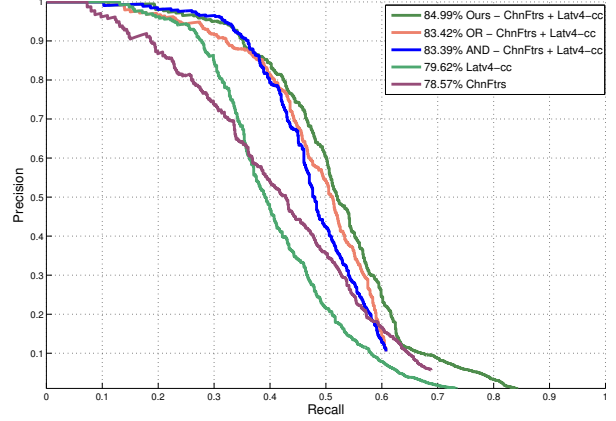
Thus, if we let α vary from zero to one, all possible relative combinations are evaluated.

We calculated for each of these combinations the accuracy score of the resulting combined detector on the test set. Figure 5 gives these results for the combination of *HOG* and *ChnFtrs*. At the extreme values 0 and 1 of α, the accuracy score equals that of both individual detectors (indicated with the dotted lines). An optimal combined accuracy is reached for a specific value of α between these two boundaries. The red dot indicates the value of α calculated using our combination rule (with $c_{\text{compl}(HOG)} = 0.685$, $c_{\text{compl}(ChnFtrs)} = 0.675$). As can be seen, our proposed combination rule manages to find the most optimal combination weights. Note that our weights are calculated based only on the confidence and complementarity measures. These are easily extracted from the detection results, thus avoiding the need to perform an exhaustive search over all possible combinations like we do in this validation experiment.

IV. EXPERIMENTS AND RESULTS

To illustrate the potential of our combination approach, we performed thorough accuracy experiments. Our framework uses the publicly available detection results from the Caltech dataset [11]. This dataset consists of about 250,000 frames of which each 30th frame is used for evaluation (resulting in about 8300 frames). All combination coefficients as mentioned above were first calculated on the trainingset (set00 - set05, 4250 frames). As optimal operating point we use $N = 50\%$ of the number of ground truth detections. Next, our combination rule was executed on the testset (set06 - set10, 4024 frames), using the *reasonable* settings. The experiments indicate that our combination approach achieves excellent accuracy results, and specific combinations achieve better than state-of-the-art detection results.

Besides our approach, for each experiment we also performed the *AND* (only keep overlapping detections) and *OR* (keep all detections) combinations. Figure 6 displays the precision-recall results of our combination rule for *ChnFtrs + LatV4-cc*, compared with these AND and OR results. Our proposed combination approach easily outperforms these naive combination rules. Further note that the recall of our approach
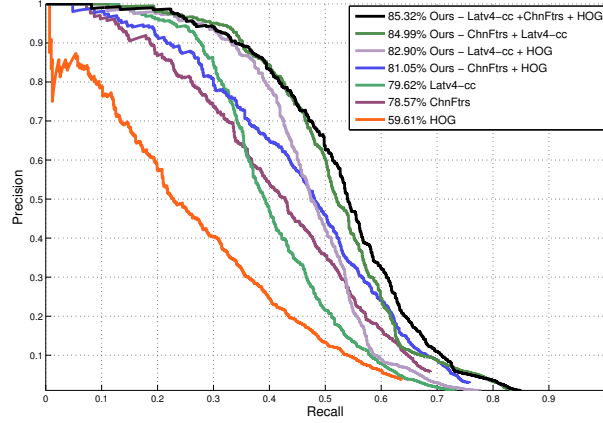
Fig. 7. Precision-Recall curve of all possible combinations of our three benchmark detectors.
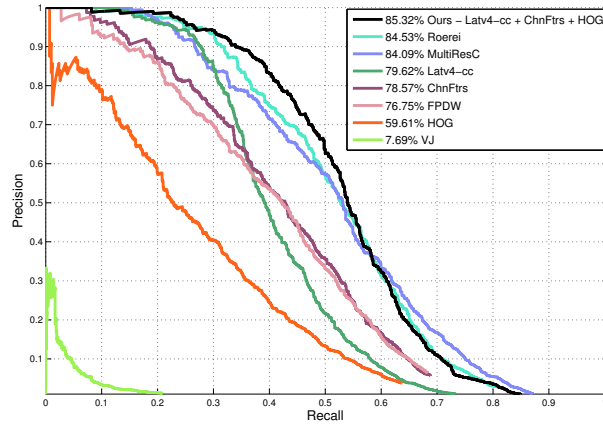

Fig. 8. Accuracy of our combination rule for the three benchmark detectors versus the state-of-the-art pedestrian detectors [11], [12].

is significantly higher. A similar trend is noticed for our other combinations versus the AND and OR combination rules.

Figure 7 compares all possible combinations of the three benchmark detectors used in our paper. A combination of two detectors evidently outperforms both corresponding individual detectors. A combination of the three detectors further slightly increases the accuracy.

Finally, figure 8 displays the accuracy of our combination rule for the combination of our three benchmark detectors, compared with the current state-of-the-art pedestrian detectors. As can be seen our combination rule achieves excellent detection results, reaching an accuracy of 85.32% on the challenging Caltech dataset.

V. CONCLUSIONS AND FUTURE WORK

In this paper we presented a generic pedestrian detector combination methodology to further increase the detection accuracy. Our approach allows for the combination of an arbitrary number of pedestrian detectors, and manages to achieve an optimal combination rule. For this we introduced two measures: confidence and complementary. Using these

measures the detection scores of multiple pedestrian detectors are combined in a final detection score.

Our experiments, using three standard pedestrian detectors, indicate that we outperform the traditionally used (naive) *AND* and *OR* approaches, and achieve better than state-of-the-art detection results on the challenging Caltech dataset.

In future work we plan to extend our approach with a dynamic confidence measure (as opposed to a fixed value per detector used in this paper) based on image dependent features (e.g. contrast, height, texturedness,..). This parameterised measure can then be used in two ways: either to further improve the detection accuracy, or in a *select-out* strategy where, during detection, the most appropriate pedestrian detector is selected for a specific image window.

REFERENCES

[1] R. Benenson, M. Mathias, T. Tuytelaars, and L. Van Gool, "Seeking the strongest rigid detector," in *Proc. of CVPR*, 2013.

[2] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Fast stixel computation for fast pedestrian detection," in *Proc. of ECCV*, 2012, pp. 11–20.

[3] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second," in *Proc of. CVPR*, 2012, pp. 2903–2910.

[4] H. Cho, P. Rybski, A. Bar-Hillel, and W. Zhang, "Real-time pedestrian detection with deformable part models," in *IEEE Intelligent Vehicles Symposium*, August 2012.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of CVPR*, vol. 2, June 2005, pp. 886–893.

[6] S. De Beugher, G. Brône, and T. Goedemé, "Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection," in *VISAPP*, 2014.

[7] F. De Smedt, L. Struyf, S. Beckers, J. Vennekens, G. De Samblanx, and T. Goedemé, "Is the game worth the candle? Evaluation of OpenCL for object detection algorithm optimization," in *Proc. of PECCS*, 2012.

[8] F. De Smedt*, K. Van Beeck*, T. Tuytelaars, and T. Goedemé (* equal authorship), "Pedestrian detection at warp speed: Exceeding 500 detections per second," in *CVPR - Embedded Vision Workshop (EVW)*, 2013.

[9] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *Proc. of BMVC*, 2010.

[10] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. of BMVC*, 2009.

[11] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. of CVPR*, June 2009.

[12] ——, "Pedestrian detection: An evaluation of the state of the art," *In IEEE PAMI*, vol. 99, 2011.

[13] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *In IEEE PAMI*, vol. 32, no. 9, pp. 1627–1645, 2010.

[14] P. F. Felzenszwalb, R. B. Girshick, and D. Mcallester, "Cascade object detection with deformable part models," in *Proc. of CVPR*, 2010.

[15] L. Ladický, P. Sturgess, K. Alahari, C. Russell, and P. H. S. Torr, "What, where and how many? Combining object detectors and CRFs," in *ECCV*, 2010, pp. 424–437.

[16] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Efficient subwindow search: A branch and bound framework for object localization," *In IEEE PAMI*, vol. 31, no. 12, pp. 2129–2142, 2009.

[17] M. Mathias, R. Benenson, R. Timofte, and L. Van Gool, "Handling occlusions with franken-classifiers," in *Proc. of ICCV*, 2013.

[18] V. Prisacariu and I. Reid, "fastHOG - a real-time gpu implementation of HOG," Department of Engineering Science, Oxford University, Tech. Rep., 2009.

[19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of CVPR*, vol. 1, 2001, pp. 511–518.