ADAPTIVE PEOPLE DETECTION BASED ON CROSS-CORRELATION MAXIMIZATION

Alvaro García-Martín, Juan C. SanMiguel

Video Processing and Understanding Lab (VPULab), Universidad Autónoma de Madrid, Spain

ABSTRACT

Applying people detectors to unseen data is challenging since patterns distributions may significantly differ from the ones of the training dataset. In this paper, we propose a framework to adapt people detectors during runtime classification. Such adaptation takes advantage of multiple detectors to identify their best configurations (i.e. detection thresholds) without requiring manually labeled ground truth. We maximize the mutual information of detectors by pair-wise correlating their outputs to obtain a set of hypotheses for the detection thresholds. These hypotheses are later combined by weighted voting to obtain a final decision for the detection threshold of each detector. The proposed approach does not require re-training detectors and uses standard people detector outputs, i.e., bounding boxes, therefore it can employ various types of detectors. The experimental results demonstrate that the proposed approach outperforms state-of-the-art detectors whose optimal configuration is learned from training data.

Index Terms— People detection, Detector adaptation, Pair-wise correlation, Thresholds.

1. INTRODUCTION

People detection is pivotal in many computer vision areas such as video-surveillance, human-computer interaction and mobile robotics. Albeit many approaches are available due to the intensive research carried out in the past years, detection performance still exhibits a strong dependency on the training data used to build detectors [1]. Hence, detector performance significantly decreases when training and testing data have different patterns such as viewpoints, motion, poses, backgrounds, occlusions and people sizes [2].

The adaptation of people detectors is therefore desired to learn specific patterns from unseen data [3], which faces several challenges related to high-dimensionality, modeling intra-class variance and determining subspaces shared by both training and test data [4]. Learning such scene-specific detectors is often formulated as a domain adaptation problem [5]. For example, multi-class classifiers are adapted by computing the proportion of objects in the test data during runtime which is used for multi-class bayesian classification [6]. Data augmentation can be also employed to adapt classifiers to the video-surveillance domain [7]. Moreover, feature learning is also proposed without requiring annotated test data [1][2]. However, all these approaches require to re-train detectors with previously recorded data (labeled or unlabeled), which may not be possible in certain applications such as real-time video-surveillance.

Detector combination emerges as an alternative to counteract limitations of independent detectors. For example, the similarity between training and test data in feature spaces can be exploited to select the best detector from a pool [4]. Multiple detection modalities can be combined to improve final recognition performance [8]. Moreover, detector ranking can be efficiently learned for different test data subsets [9]. Cascades of detectors can be also designed to combine the confidence of heterogeneous detectors [10]. Finally, detector adaptation may be achieved by coupling detection and tracking for single [11] or multiple [12] detectors. However, these approaches share the limitations of domain adaption (detector re-training), impose restrictions on the employed detectors (e.g. high-precision and low-recall [12]) or require the use of tracking [11][12].

Fig. 1. People detection results for Faster R-CNN [13] (sequence *S1-T1-C*, http://www.cvg.reading.ac.uk/PETS2006). Each row corresponds to applying a detection threshold with values 0.25 (row 1), 0.5 (row 2) and 0.75 (row 3). Finding an optimal threshold (framed in green) for all cases is challenging due to the variability of viewpoints, people sizes and occlusions.

In order to enable the application of people detection to unseen data, in this paper we propose a framework to automatically adapt people detectors configuration during runtime classification. Unlike approaches based on domain adaptation and detector combination, we avoid re-training detectors and perform such adaptation by finding the best configuration given an image without manual annotations and the outputs of the employed detectors. In particular, we focus on obtaining the optimal detection threshold which has a strong impact on detector's performance (see Fig. 1). This proposal explores multiple thresholding hypotheses and exploits the correlation among pairs of detectors outputs to determine the best pair of thresholds for each one. Then, such pair-wise hypotheses are combined by weighted voting to obtain the final adapted threshold for each individual detector. The proposed framework only requires threshold-

Work partially supported by the Spanish Government through the HA-Video project, under grant TEC2014-5317-R.

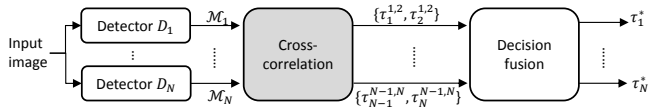


Fig. 2. Proposed framework to adapt N people detectors by finding their best detection thresholds $\tau_1^*, ..., \tau_N^*$. The *cross-correlation* block is extended in Fig. 3 for two detectors.

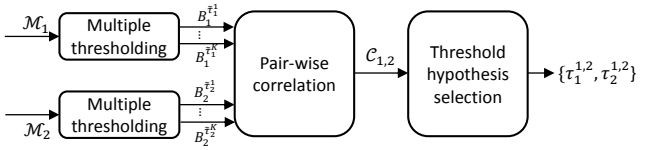


Fig. 3. Block diagram of the cross-correlation between two detectors to obtain adapted detection thresholds $\tau_1^{1,2}$ and $\tau_2^{1,2}$.

based detectors with an output in the form of bounding boxes. Therefore, it can be applied to many recent approaches as demonstrated in the experimental results, where the adaption using sets of people detectors (from two to six) outperform these individual detectors whose threshold is optimally trained in advance.

The rest of the paper is structured as follows. Section 2 overviews the proposed framework whereas Section 3 describes the adaptation based on pair-wise correlations. Section 4 presents the experiments. Finally, Section 5 concludes this paper.

2. DETECTOR ADAPTATION FRAMEWORK

We propose a framework to improve the performance at runtime classification by adapting the detector configuration (see Fig. 2). This proposal is inspired by the *maximization of mutual information* strategy where classifiers are combined assuming that their errors are complementary, being successfully applied for example to detect shadows [14] and skin [15]. We extend such maximization framework to people detection by introducing pair-wise detector correlation and by adapting online their configuration. Note that we are not interested in re-training detectors which may require data not available in real applications or highly-accurate detectors; and may imply high latency [6]. Instead, we consider generic threshold-based detectors pre-trained on standard datasets, thus making this proposal applicable to a wide variety of detectors.

Assuming a set of N independent people detectors $\{D_n\}_{n=1}^N$ applied to an image . Each detector D_n obtains a confidence map $\mathcal{M}_n$ describing the people likelihood for each spatial location (x, y) and scale s in the image. Then, detection candidates are obtained by thresholding this map:

$$\mathcal{T}_n(x,y,s) = \begin{cases} 1 & if \quad \mathcal{M}_n(x,y,s) > \tau_n \\ 0 & otherwise \end{cases}, \tag{1}$$

where $\mathcal{T}_n(x,y,s) = \{0,1\}$ and τ_n is the detection threshold whose values are heuristically set based on the confidence map. These candidates are later combined across scales and can be post-processed by a variety of techniques such as non-maximum suppression [16] and background-people segmentation [17]. The final result is a set $\boldsymbol{B_n^{\tau_n}} = \{b_k\}_{k=1}^{k=K^{\tau_n}}$ with K^{τ_n} detections (i.e.bounding boxes) representing the output of the detector D_n where each detection b_j is described by its position (x, y) and dimensions (w, h). A key parameter in this procedure is the detection threshold τ_n which determines the number of detection candidates. Low (high) values of τ_n generate several (few) detections increasing the false (true) positive rate, three examples of τ_n are shown in Fig. 1. We propose to adapt such detection threshold to the image conditions by exploring similarities with the other independent detectors.

We compare the output of detectors to obtain a set of pairwise correlation scores (*cross-correlation* in Fig. 2). First, we explore the decision space of each detector output by applying multiple thresholds. Then, we correlate these multiple outputs for each pair of detectors (D_n and D_m) to obtain a correlation map $\mathcal{C}_{n,m}$ which measures the output similarity. Finally, we find the configuration with the highest similarity (i.e. highest value in $\mathcal{C}_{n,m}$) to select the best detection threshold for each detector ($\tau_n^{n,m}$ and $\tau_m^{n,m}$, respectively). This stage is extended in Section 3 and Fig. 3.

Up to this point, we have hypothesis obtained for each compared pair of detectors (i.e. D_n and D_m) which are combined to obtain a final configuration for each detector (*decision fusion* in Fig. 2). Such hypothesis combination is performed as a traditional mixture of experts via weighted voting [18]:

$$\tau_n^* = \sum_{m=1}^{N} \omega^{n,m} \cdot \tau_n^{n,m} \, (n \neq m), \tag{2}$$

where $\omega^{n,m} \in [0,1]$ is the weight for the hypothesis $\tau_n^{n,m}$ achieved by comparing D_n and D_m and $\sum_{m=1}^{N} \omega^{n,m} = 1 \, (n \neq m)$. Although such ensemble voting may benefit from a previous learning stage [19], currently we assume no prior knowledge about detectors performance so we consider equal weighting $\omega^{n,m} = \frac{1}{N-1}$. It is important to note that we are not combining people detectors, the proposed approach focuses on improving independently each detector by exploring output similarities with other detectors.

3. PAIR-WISE CORRELATION MAXIMIZATION

We obtain the hypotheses $\tau_n^{n,m}$ to find the best threshold of each detector D_n by maximizing the correlation between D_n's output and the rest of the detectors $D_m \, (m \neq n)$. Starting from the confidence maps $\mathcal{M}_n$ such procedure is decomposed into three states (see Fig. 3) described as follows.

3.1. Multiple thresholding

To explore the possible detector outputs, we define a set of L thresholds $\left\{\tilde{\tau}_n^j\right\}_{j=1}^{j=L}$ for each detector D_n whose values are determined by considering L levels between the extreme values of the confidence map $\mathcal{M}_n$ (i.e. minimum and maximum). Then we perform thresholding with multiple values $\tilde{\tau}_n^j$ to obtain a set of outputs as follows:

$$\Omega_n = \{B_n^{\tilde{\tau}_n^j}\}; 1 \leq j \leq L, \tag{3}$$

where each output $B_n^{\tilde{\tau}_n^j}$ is obtained by applying the threshold $\tilde{\tau}_n^j$ to Eq. (1). Note that each detector D_n may have different thresholds $\tilde{\tau}_n^j$ adapted to the range of values in $\mathcal{M}_n(x,y,s)$.

3.2. Pair-wise correlation

We correlate the detector outputs $\{\Omega_n\}_{n=1}^{n=N}$ to estimate their similarity. We compute a correlation map $\mathcal{C}_{n,m}$ for each pair of detectors outputs Ω_n and Ω_m where each element is defined as:

$$\mathcal{C}_{n,m}(i,j) = \rho(B_n^{\tilde{\tau}_n^i}, B_m^{\tilde{\tau}_m^j}), \; (i,j \in \{1, ...L\}) \tag{4}$$

where $\rho(\cdot,\cdot)$ is a function to compute the similarity between the output of detectors. The number of correlation maps $\mathcal{C}_{n,m}$ to be computed for N detectors is $\begin{pmatrix} N \\ 2 \end{pmatrix} = \frac{N!}{2\cdot(N-2)!}$.

We propose to compute $\rho(\cdot,\cdot)$ as a one-class classification problem by applying standard evaluation measures. To compare bounding boxes from two outputs, we use three matching criteria [20]: relative distance $dr \in [0, d_{max}]$ (where d_{max} is the image diagonal divided by each b_j size), cover $co \in [0,1]$ and spatial overlap $ov \in [0,1]$. The criterion dr measures the distance between the bounding box centers of $B_n^{\tilde{\tau}_n^i}$ and $B_m^{\tilde{\tau}_m^j}$ in relation to the size of the bounding boxes in $B_m^{\tilde{\tau}_m^j}$. Similarly to dr, the criteria co and ov employ respectively the percentage of spatial bounding box coverage in $B_m^{\tilde{\tau}_m^j}$ and the intersection-over-union features. A positive match is considered true if $dr \leq 0.5$, $co \geq 0.5$ and $ov \geq 0.5$ as commonly employed in related works [20] which corresponds to a deviation up to 25% of the true object size. Only one $b_k \in B_n^{\tilde{\tau}_n^i}$ is accepted as correct by matching $b_l \in B_n^{\tilde{\tau}_m^j}$ (i.e. true positive), so any additional $b_k \in B_n^{\tilde{\tau}_n^i}$ on the same bounding box is considered as a false positive. Then, we compute Precision and Recall measures from the matching results and obtain the FScore as the final similarity measure $\rho(\cdot,\cdot)$ between $B_n^{\tilde{\tau}_n^i}$ and $B_m^{\tilde{\tau}_m^j}$ as in [21].

Thus the final correlation map $\mathcal{C}_{n,m}$ between two detectors is defined as the FScores F:

$$\mathcal{C}_{n,m} = \begin{bmatrix} F\left(B_n^{\tilde{\tau}_n^1}, B_m^{\tilde{\tau}_m^1}\right) & \dots & F\left(B_n^{\tilde{\tau}_n^1}, B_m^{\tilde{\tau}_m^L}\right) \\ \dots & F\left(B_n^{\tilde{\tau}_n^i}, B_m^{\tilde{\tau}_m^j}\right) & \dots \\ F\left(B_n^{\tilde{\tau}_n^L}, B_m^{\tilde{\tau}_m^1}\right) & \dots & F\left(B_n^{\tilde{\tau}_n^L}, B_m^{\tilde{\tau}_m^L}\right) \end{bmatrix}, \tag{5}$$

where $i, j = \{1, ..., L\}$. Fig. 4 shows three examples for $\mathcal{C}_{n,m}$ and the output of two detectors for two threshold values.

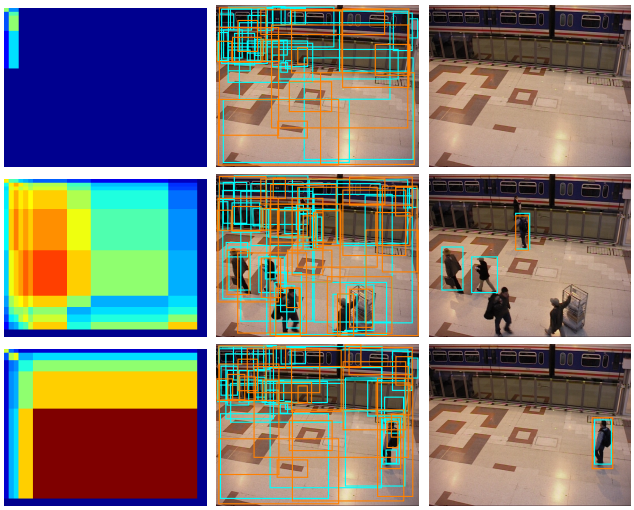
3.3. Threshold hypothesis selection

Based on the principle of maximization of mutual information, we assume that two independent detectors, albeit designed for the same purpose (to detect persons), would be highly correlated when many bounding boxes are matched and therefore, a high level true positive detections is expected. On the other hand, low correlation values would have few matches and therefore, imply an increase in the false positive rate or negative detection rate. Note that there is one exception to this assumption when outputs are empty (i.e. $B_n^{\tilde{\tau}_n^i} = B_m^{\tilde{\tau}_m^j} = \emptyset$) since both outputs are equal and we cannot compute the FScore. To consider this, we avoid this situation by setting the FScore to zero when these sets are empty.

To select a detection-threshold among the *L* hypotheses for each detector, we apply the maximum a posteriori criterion:

$$\{\tau_n^{n,m}, \tau_m^{n,m}\} = \underset{\tilde{\tau}_n^i, \tilde{\tau}_m^j}{argmax} \left(\rho(B_n^{\tilde{\tau}_n^i}, B_m^{\tilde{\tau}_m^j})\right), \tag{6}$$

which is equivalent to get the maximum value of $\mathcal{C}_{n,m}$. Since the solution to Eq. (6) may not be unique, we may obtain various maximum values $\tau_n^{n,m}$ (see the red area in the bottom-left image in Fig. 4) and the detectors are never totally independent. Therefore, we currently propose three alternatives, select the mean, minimum or maximum value among those thresholds $\tau_n^{n,m}$ maximizing $\mathcal{C}_{n,m}$.



(a) (b) (c)

Fig. 4. Correlation and results between two detectors based on Faster R-CNN [13] using VGG (cyan) and ZF (orange) models. Three examples (rows) are shown where columns are (a) correlation map $\mathcal{C}_{1,2}$ with color code: (blue) $0 \leq \mathcal{C}_{1,2} \leq 1$ (red); and obtained bounding boxes with thresholds (b) $\tilde{\tau}_1^1 = \tilde{\tau}_2^1 = 0$ and (c) $\tilde{\tau}_1^{36} = \tilde{\tau}_2^{36} = 0.9$. The full set of $L = 40$ thresholds $\left\{\tilde{\tau}_n^j\right\}_{j=1}^{j=L=40}$.

4. EXPERIMENTAL RESULTS

4.1. Setup

We evaluate the proposed approach, Adaptive people Detection by maximizing Correlation (ADC), using the People detection benchmark repository (PDbm[1]) [22]. It has 19 sequences with ground-truth annotations for traditional indoor and outdoor scenarios in computer vision applications: video surveillance, smart cities, etc.

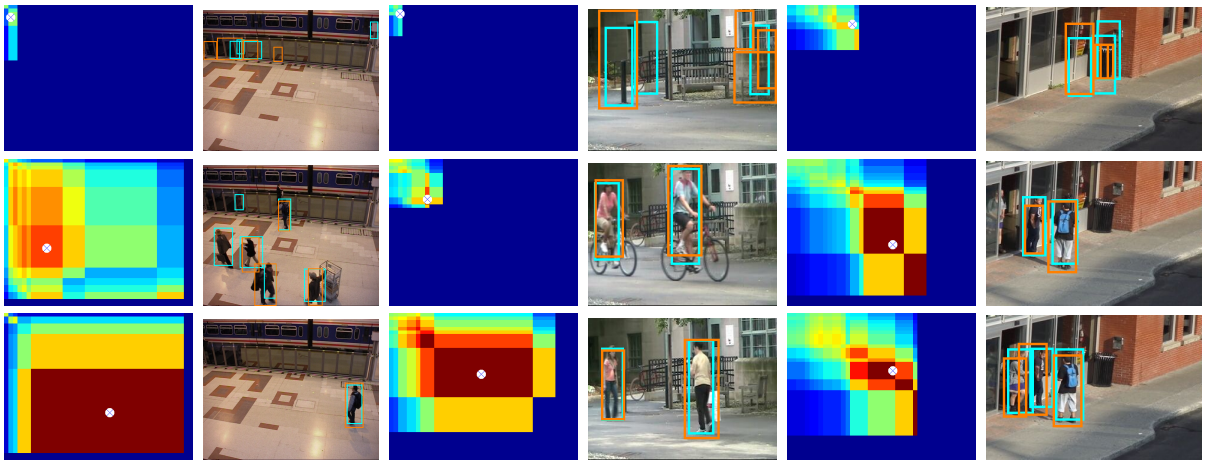
For each frame, we quantify detection performance by Precision, Recall and FScore metrics [21]. We report the mean FScore for all tested images as the final performance value.

We apply ADC to six people detectors using publicly available implementations. We use two versions for DPM [16] (Inria and Pascal models), ACF [23] (Inria and Caltech models) and Faster R-CNN [13] (VGG and ZF models). We evaluate five sets with incremental size to test the effect of successively adding detectors to the final result: ADC2 (DPM-I, DPM-P), ADC3 (DPM-I, DPM-P, ACF-I), ADC4 (DPM-I, DPM-P, ACF-I, ACF-C), ADC5 (DPM-I, DPM-P, ACF-I, ACF-C, FRCNN-VGG) and ADC6 (DPM-I, DPM-P, ACF-I, ACF-C, FRCNN-VGG, FRCNN-ZF). Note that we do not try to improve the performance of each algorithm. Instead, we focus on automatically threshold adaptation during runtime classification in order to get the best possible results for each algorithm.

4.2. Results

Table 1 shows the average results after adapting two and six detectors, ADC2 and ADC6 respectively, with different number of thresholds $L = \{5, 10, 20.40\}$ and strategies to select a threshold $\tau_n^{n,m}$ from those values maximizing $\mathcal{C}_{n,m}$ (*mean*, *minimum* or *maximum*). In both cases, the results show that the performance increases progressively with the number of thresholds. In addition, the *minimum*

[1] http://www-vpu.eps.uam.es/PDbm/, last accessed February 2017.



(a) (b) (c)

Fig. 5. Correlation and threshold selection results between pairs of detectors. Each column pair shows an example of the selected thresholds (cross-marked) in the correlation map (left column) and the corresponding obtained bounding boxes (right column). Column pairs correspond to (a) Faster R-CNN [13] using VGG (cyan) and ZF (orange) models, (b) DPM[16] using Inria (cyan) and Pascal (orange) models and (c) ACF [23] using Inria (cyan) and Caltech (orange) models.

Table 1. Average FScore of adapted detectors for different strategies to select a threshold $\tau_n^{n,m}$ from those values maximizing $\mathcal{C}_{n,m}$ obtained with various threshold with $L = 5, 10, 20$ and 40. Bold indicates best result for (a) ADC2 and (b) ADC6.

(a) ADC2

	L thresholds			
Strategy	5	10	20	40
Mean	33.2	35.1	35.9	**36.3**
Minimum	33.4	34.9	35.7	35.9
Maximum	33.2	35.0	35.7	36.0

(b) ADC6

	L thresholds			
Strategy	5	10	20	40
Mean	39.8	41.6	42.4	**42.7**
Minimum	38.8	39.9	39.6	39.0
Maximum	40.1	41.7	42.2	42.2

Table 2. Average FScore of the five ADC combinations from ADC2 to ADC6. Percentage increase (%Δ) calculated for each detector with respect to the obtained performance just after their inclusion in the combination (in bold), from ADC2 to ADC5 respectively.

	ADC combinations								
	ADC2	ADC3	%Δ	ADC4	%Δ	ADC5	%Δ	ADC6	%Δ
DPM-I [16]	**37.1**	37.3	0.5	37.7	1.6	38.3	3.2	38.5	3.8
DPM-P [16]	**35.4**	35.9	1.4	36.2	2.3	36.9	4.2	37.1	4.8
ACF-I [23]	-	**38.3**	-	38.6	0.8	39.3	2.6	39.6	3.4
ACF-C [23]	-	-	-	**40.0**	-	41.7	4.3	42.2	5.5
FRCNN-VGG [13]	-	-	-	-	-	**51.6**	-	51.7	0.2
FRCNN-ZF [13]	-	-	-	-	-	-	-	47.2	-

strategy obtains in general the worst results and the *mean* strategy obtains slightly better results than the *maximum* one. Fig. 5 shows examples of correlation and threshold selection results between pairs of detectors.

Table 2 shows one example of successively adding detectors to the final configuration from 2 detectors to 6 (from ADC2 to ADC6). In general, the results show that the greater number of detectors the higher performance. For example, the DPM-I increases progressively the performance from 37.1 (ADC2) to 38.5 (ADC6).

Table 3 shows the comparative results of our approach (ADC6, all the six detectors independently of the order or their inclusion) versus two different Fixed Thresholding approaches (FT_{PDbm} and FT_{VOC12}). The FT_{PDbm} approach is the ideal case, the optimal threshold is determined according to the chosen evaluation dataset

Table 3. Comparison in terms of average FScore between two fixed thresholding approaches and the ADC6. Percentage increase (%Δ^{PDbm} and %Δ^{VOC12}) calculated with respect to the fixed thresholding approaches, FT_{PDbm} and FT_{VOC12} respectively.

	Fixed threshold		Proposed threshold adaptation		
	FT_{PDbm}	FT_{VOC12}	ADC6	%Δ^{PDbm}	%Δ^{VOC12}
DPM-I [16]	33.9	29.9	38.5	13.6	28.8
DPM-P [16]	32.9	31.3	37.1	12.8	18.5
ACF-I [23]	35.2	32.1	39.6	12.5	23.4
ACF-C [23]	36.6	35.2	42.2	15.3	19.9
FRCNN-VGG [13]	50.1	46.0	51.7	3.2	12.4
FRCNN-ZF [13]	44.2	41.2	47.2	6.8	14.6
Average Improvement	-	-	-	10.7	19.6

(PDbm [22]) and the FT_{VOC12} is a more realistic approach, where the optimal threshold is previously learnt with the training dataset VOC2012 (Visual Object Classes Challenge 2012 [24]). The results show clearly that the use of our adaptive threshold approach ADC6 significantly improves the results of any of the individual detectors using a fixed threshold (10.7 and 19.6% average improvement with respect to FT_{PDbm} and FT_{VOC12} respectively).

5. CONCLUSIONS

We have presented a framework to automatically adapt people detectors during runtime classification. This proposal explores multiple thresholding hypotheses and exploits the correlation among pairs of detector outputs to determine the best threshold. These hypotheses are later combined by weighted voting to obtain a final decision for the detection threshold of each detector. The proposed approach uses standard state of the art detector outputs (bounding boxes), therefore it can employ various types of detectors. This framework allows the automatic threshold adaptation without requiring a re-training process and therefore without requiring manually labeled data. The experimental results demonstrate that any correlation up to six detectors outperforms state-of-the-art detectors, whose thresholds are optimally trained in advance.

As future work, we will explore other threshold selection and fusion alternatives; we will apply this proposal to other detectors.

6. REFERENCES

[1] X. Wang, M. Wang, and W. Li, "Scene-specific pedestrian detection for static video surveillance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 361–374, February 2014.

[2] Z. Xingyu, O. Wanli, W. Meng, and W. Xiaogang, *Deep Learning of Scene-Specific Classifier for Pedestrian Detection*, pp. 472–487, Springer International Publishing, 2014.

[3] T. Kalinke, C. Tzomakas, and W. V Seelen, "A texture-based object detection and an adaptive model-based classification," in *IEEE Intelligent Vehicles Symposium*, 1998, pp. 341–346.

[4] S. Zhang, Q. Zhu, and A. Roy-Chowdhury, "Adaptive algorithm selection, with applications in pedestrian detection," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3768–3772.

[5] K. Kyaw Htike and D. Hogg, "Adapting pedestrian detectors to new domains: A comprehensive review," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 142–158, 2016.

[6] A. Royer and C. H. Lampert, "Classifier adaptation at prediction time," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1401–1409.

[7] A. Dimou and F. Alvarez, "Multi-target detection in cctv footage for tracking applications using deep learning techniques," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3–7.

[8] O. Mees, A. Eitel, and W. Burgard, "Choosing smartly : Adaptive multimodal fusion for object detection in changing environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 151–156.

[9] S. Karaoglu, Y. Liu, and T. Gevers, "Detect2rank: Combining object detectors using learning to rank," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 233–248, Jan 2016.

[10] A. Verma, R. Hebbalaguppe, L. Vig, S. Kumar, and E. Hassan, "Pedestrian detection via mixture of cnn experts and thresholded aggregated channel features," in *IEEE International Conference on Computer Vision (ICCV)*, 2016, pp. 555–563.

[11] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, July 2012.

[12] A. Gaidon, G. Zen, and J. Rodriguez, "Self-learning camera: Autonomous adaption of object detectors to unlabeled video streams," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 1–9.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[14] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–6.

[15] J. C. SanMiguel and S. Suja, "Skin detection by dual maximization of detectors agreement for video monitoring," *Pattern Recognition Letters*, vol. 34, no. 16, pp. 2102–2109, Dec 2013.

[16] P. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[17] A. García-Martín and J. M. Martínez, "Post-processing approaches for improving people detection performance," *Computer Vision and Image Understanding*, vol. 133, pp. 76–89, April 2015.

[18] B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, *Fusion in Computer Vision: Understanding Complex Visual Content*, Springer, 2014.

[19] B. Baruque and E. Corchado, *Fusion methods for unsupervised learning ensembles*, Springer, 2011.

[20] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 878–885.

[21] A. García-Martín and J. M. Martínez, "People detection in surveillance: classification and evaluation," *IET Computer Vision*, vol. 9, no. 5, pp. 779–788, September 2015.

[22] A. García-Martín, B. Alcedo, and J. M. Martínez, "Pdbm: people detection benchmark repository," *Electronics Letters*, vol. 51, no. 7, pp. 559–560, April 2015.

[23] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, January 2014.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.