Adapting Pedestrian Detection from Synthetic to Far Infrared Images

Yainuvis Socarrás[1,2,*] Sebastian Ramos[1,*] David Vázquez[1] Antonio M. López[1,2] Theo Gevers[3]
[1]Computer Vision Center [2]Dept. Computer Science, Universitat Autònoma de Barcelona
[3]Faculty of Science, University of Amsterdam

{ysocarras,sramosp,dvazquez,antonio}@cvc.uab.es th.gevers@uva.nl

Abstract

We present different techniques to adapt a pedestrian classifier trained with synthetic images and the corresponding automatically generated annotations to operate with far infrared (FIR) images. The information contained in this kind of images allow us to develop a robust pedestrian detector invariant to extreme illumination changes.

1. Introduction

Visual pedestrian detection has been a relevant topic in the computer vision community during the last decade, as presented in the comprehensive state-of-the-art review of [3]. Most of the research done on this task has been focused on the classification stage, *i.e.* given a candidate window decide if it contains a pedestrian or not. To this end, different approaches aim to design a robust pedestrian detector by analysing information like color-, infrared-, and multimodal-stereo [5]. In our case, we focus our research on taking advantage of far infrared information to develop a pedestrian detector able to work during day and night robust to extreme illumination changes.

Additionally, to avoid the need of expensive and tedious manual annotations for the training of the pedestrian detector, we use synthetic images to automatically collect the training examples with their pixelwise ground truth as presented in [6]. However, in this case the detector experiences a drop of performance because of the dataset bias [4] between the synthetic images and the FIR ones. Therefore, in order to avoid this problem, it is required to adapt the detector trained with synthetic data to operate in the real-world scenarios. Such adaptation can be done by using domain adaptation (DA) techniques [1, 7, 9, 8].

In this paper, we explore different methods to adapt a model by training with synthetic images and a relatively low number of FIR ones. For achieving the adaptation, we used a classical pedestrian detector composed by histograms of oriented gradients (HOG) and local binary patterns (LBP) with linear SVM (Lin-SVM) [10] and we take advantage of the complementarity between synthetic and real-world data. Our work is an extension to the approach developed by Vázquez *et al*. [8], where it is investigated the adaptation of a holistic pedestrian model trained with virtual-world samples to operate in visible spectrum real-world images.

2. Domain adaptation methods

We explore four methods for the adaptation of pretrained pedestrian classifiers. These methods are based on different behaviors that an oracle ($\mathcal{O}$) can have regarding to the maximum number of target domain pedestrians ($n_{t.p.}$) that is allowed to provide for training. In particular, the adaptation process consists of the steps presented in Algorithm 1.

Algorithm 1 Domain adaptation process

1. Learning a pedestrian classifier, $\mathrm{C}_{\mathcal{V}}$, by using the synthetic samples.
2. $\mathrm{C}_{\mathcal{V}}$ asks $\mathcal{O}$ the label of *difficult* samples from the target data.
3. Such samples jointly with the synthetic ones will be used for retraining.

For the first of these behaviors, $\mathcal{O}$ annotates $n_{t.p.}$ pedestrians at *random* (Rnd). The rest of behaviors are based on a sort of *selective sampling* [2].

For the second behavior, *active learning for pedestrians* (Act+), $\mathcal{O}$ annotates $n_{t.p.}$ *difficult-to-detect* pedestrians. Analogously, we term our third behavior as *active learning for background* (Act−) because $\mathcal{O}$ only marks false positives. The idea behind Act− is not to collect the annotated false positives, but the right detections (true positives) as provided by the used pedestrian detector. In other words, in this case, the bounding box (BB) annotations of the $n_{t.p.}$ real-world pedestrians are provided by the pedestrian detector itself. Finally, we term as Act± the fourth behavior since it is a combination of Act+ and Act−. In this case we allow to collect $2n_{t.p.}$ real-world pedestrians because just $n_{t.p.}$ are manually annotated with BBs, which is the task we want to avoid.

Finally, we define the difficult cases for $\mathrm{C}_{\mathcal{V}}$. Given a real-world sample $\mathrm{s}_{\mathcal{R}}$, if $\mathrm{C}_{\mathcal{V}}(\mathrm{s}_{\mathcal{R}}) > Thr$, then $\mathrm{s}_{\mathcal{R}}$ is classified as pedestrian. Thr is the classification thresh-

[*]Both authors contributed equally.

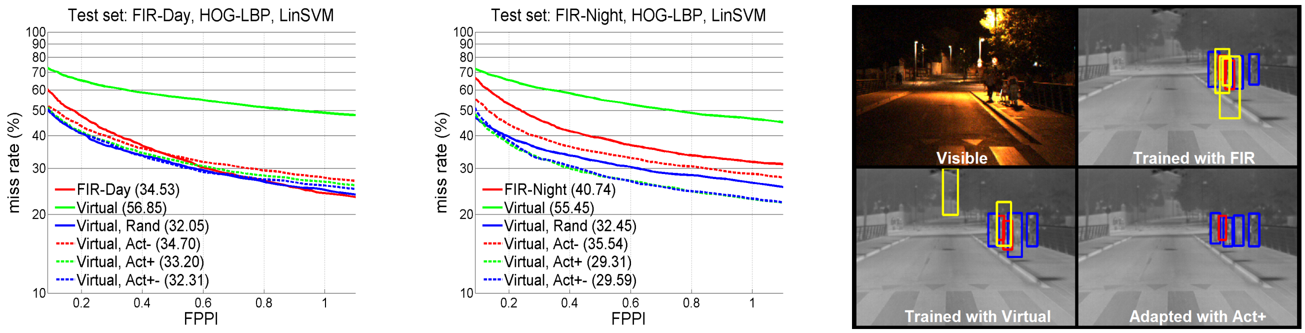

Figure 1: Miss rate vs FPPI curves during day (Plot Left) and night time (Plot Right). The number in parenthesis is the average miss rate (%) of the respective curve. The image shows an example of the proposal at night time. The yellow, red and blue boxes indicates false positive, false negative and true positive detections, respectively.

old. Accordingly, in the Act+ case, $\mathcal{O}$ will annotate real-world pedestrians, $\mathbf{s}^{+}_{\mathcal{R}}$, for which $\mathrm{C}_{\mathcal{V}}(\mathbf{s}^{+}_{\mathcal{R}}) \leq Thr$. In the Act− case, those background samples, $\mathbf{s}^{-}_{\mathcal{R}}$, for which $\mathrm{C}_{\mathcal{V}}(\mathbf{s}^{-}_{\mathcal{R}}) > Thr$ must be rejected by $\mathcal{O}$. For the Act± both things hold. In general, selective sampling for SVM focus on samples inside the ambiguity region $[-1, 1]$. However, underlying such an approach is the assumption of a shared train and test domain. Here, due to dataset shift, wrongly classified samples out of the margins can be important to achieve domain adaptation.

3. Experiments and results

We trained a holistic pedestrian classifier based on HOG, LBP and Lin-SVM. We tested such a classifier in our own FIR dataset for pedestrian detection benchmarking in the driver assistance context, our main field of interest. The results were compared with a pedestrian detector whose classifier was trained using real-world images.

In particular, our FIR dataset is composed by two sets of images, named as the *day* and *night* sets, which refers to the moment of the day they were acquired. The first set contains 5990 frames and the second 5081. Table 1 shows the number of frames and bounding boxes of each training and testing sets. The virtual dataset contains 1208 annotated pedestrians. All datasets have 1219 negative frames.

Datasets	FIR	
	Day	Night
Train	3110 (4548)	2198 (4333)
Test	2880 (2304)	2883 (2883)

Table 1: The numbers of frames in FIR dataset are indicated and its corresponding bounding boxes are in parenthesis.

We performed experiments following the DA process described in Algorithm 1, for both *day* and *night* sets in the FIR dataset. Since we aim to avoid as many manual annotations as possible, we set the $n_{t.p.}$ as the 10% of the real-world pedestrians [8] in FIR training data. The $n_{t.p.}$ value is the same for the four behaviors described in Section 2. Additionally, we trained a classifier with the full FIR training set and another one with all the synthetic images.

The obtained results (see Figure 1) revealed that synthetic and FIR images based training give rise to different classifiers for the *day* set, but to similar ones for the *night* set. Applying the proposed domain adaptation techniques using only a 10% of the real data we reduce 24.80 points of average miss rate in the *day* set and 26.14 points in the *night* one. Some improvements in the performance could be obtained by increasing $n_{t.p.}$ until about a 25% of the data, as shown in [8]. However, such improvements are not significant regarding the manual annotation effort if compared with the 10%. This way, the efficiency of our methods is supported, achieving our final objective of building a pedestrian detector robust to extreme illumination changes.

4. Conclusion

We have explored how synthetic images (source domain) can be used to learn appearance-based models for pedestrian detection in FIR images (target domain). To this end, we have tested different techniques to collect a few pedestrian samples from the target domain and to combine them with many samples from the source domain in order to train a domain adapted pedestrian classifier, robust to extreme changes in the environment illumination. These techniques allows to significatively save manual annotation effort while providing pedestrian detectors outperforming the obtained by using standard passive training based on a larger amount of manual annotations.

Acknowledgements

This work is supported by the Spanish MICINN projects TRA2011-29454-C03-01, TIN2011-29494-C03-02 and TIN2011-25606, Yainuvis Socarrás' UAB Grant and Sebastian Ramos' FPI Grant BES-2012-058280.

References

[1] Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. In *IEEE ICCV*, 2011. 1

[2] D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *ML*, 15(2):201–221, 1994. 1

[3] D. Gerónimo, A.M. López, A.D. Sappa, and T. Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE TPAMI*, 2009. 1

[4] A. Khosla, T. Zhou, T. Malisiewicz, A. Efros, and A. Torralba. Undoing the damage of dataset bias. In *ECCV*, pages 117–122, 2012. 1

[5] S. Krotosky and M. Trivedi. On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *IEEE TITS*, 8:619–629, 2007. 1

[6] J. Marin, D. Vázquez, D. Gerónimo, and A.M. López. Learning appearance in virtual scenarios for pedestrian detection. In *IEEE CVPR*, page 137 144, San Francisco, USA, 2010. 1

[7] K. Tang, V. Ramanathan, L. Fei-Fei, and D. Koller. Shifting weights: Adapting object detectors from image to video. In *Advances in Neural Information Processing Systems*, 2012. 1

[8] D. Vázquez, A.M. López, J. Marín, D. Ponsa, and D. Gerónimo. Virtual and real world adaptation for pedestrian detection. *IEEE TPAMI*, 2013. 1, 2

[9] D. Vázquez, A. López, and D. Ponsa. Unsupervised domain adaptation of virtual and real worlds for pedestrian detection. In *Int. Conf. in Pattern Recognition*, Tsukuba Science City, Japan, 2012. 1

[10] X. Wang, T. Han, and S. Yan. An HOG–LBP human detector with partial occlusion handling. In *IEEE ICCV*, Kyoto, Japan, 2009. 1