Improving HOG with Image Segmentation: Application to Human Detection

Conference Paper · January 2012

DOI: 10.1007/978-3-642-33140-4_16

David Vázquez
Autonomous University of Barcelona
75 PUBLICATIONS **2,720** CITATIONS

SEE PROFILE

Antonio M. López
Autonomous University of Barcelona
218 PUBLICATIONS **5,980** CITATIONS

SEE PROFILE

David Geronimo
Catchoom
37 PUBLICATIONS **1,477** CITATIONS

SEE PROFILE

T. Gevers
University of Amsterdam
305 PUBLICATIONS **16,746** CITATIONS

SEE PROFILE

Improving HOG with Image Segmentation: Application to Human Detection

Yainuvis Socarrás[1,2], David Vázquez[1,2], Antonio M. López[1,2], David Gerónimo[1,2], and Theo Gevers[1,3]

[1] Computer Vision Center, Universitat Autónoma de Barcelona, Spain
[2] Department of Computer Science, Universitat Autónoma de Barcelona, Spain
[3] Informatics Institute, Faculty of Science, University of Amsterdam, The Netherlands
{ysocarras,antonio,dvazquez,dgeronimo}@cvc.uab.es,
th.gevers@uva.nl

Abstract. In this paper we improve the *histogram of oriented gradients* (HOG), a core descriptor of state-of-the-art object detection, by the use of higher-level information coming from image segmentation. The idea is to re-weight the descriptor while computing it without increasing its size. The benefits of the proposal are two-fold: (i) to improve the performance of the detector by enriching the descriptor information and (ii) take advantage of the information of image segmentation, which in fact is likely to be used in other stages of the detection system such as candidate generation or refinement.

We test our technique in the INRIA person dataset, which was originally developed to test HOG, embedding it in a human detection system. The well-known segmentation method, mean-shift (from smaller to larger super-pixels), and different methods to re-weight the original descriptor (constant, region-luminance, color or texture-dependent) has been evaluated. We achieve performance improvements of 4.47% in detection rate through the use of differences of color between contour pixel neighborhoods as re-weighting function.

1 Introduction

Vision-based human detection is a key component in fields such as advanced driving assistance [14, 8, 12] and video surveillance [21, 17, 27]. Detecting people in images represents a challenging task given their intra-class variability, the diversity of backgrounds and the different image acquisition conditions. Nowadays, even detecting non-occluded standing persons is still a hot topic of research. As can be seen in [14], building a vision-based human detector requires to develop different modules. In this work we want to improve human detection by focusing on *classification*, i.e., on building a classifier that given an image window decides if it contains a person or not.

Nowadays, most successful classification processes for human detection follow the learning-from-examples paradigm [14, 8], where core ingredients are the set of *descriptors* used to represent the humans as well as the learning algorithm itself. Indeed, finding good sets of descriptors for developing a human classifier is a major key for its success. Different sets try to exploit (combinations of) cues as shape and texture [28, 7], even adding motion and depth [9, 27]. Among all possible sets of descriptors, one that is being specially useful for building human detectors (and object detectors in general) is the so-called HOG, i.e., the histograms of oriented gradients. This descriptor was proposed in [6] for building a holistic classifier, using linear support vector machines (linear SVM) as learning algorithm. HOG still remains as a competitive baseline method for comparison with new human classifiers [8, 7]. Although HOG descriptors capture the shape of the humans in a dense way, i.e., the positive weights learnt when using a linear classifier resemble the human silhouette, they are also affected by local noise and texture given that gradient is a local measure. On the other hand, there are works that explicitly exploit the human's shape either holistically [13] or in part-based approaches [22]. In this cases, however, the image pixels out of the silhouette are not taken into account, i.e., there is not such a non-human class.

In this work, we aim to enhance human silhouette orientations, without explicitly computing such silhouettes, but using information not as local as the own gradient magnitude. Thus, we propose to use image segmentation to obtain image segments (or regions) with their corresponding frontiers, in order to later re-weight the HOG descriptor according to this frontier information as well as appearance differences between the segments.

The inspiration for this proposal comes from two sources. The first is the idea of using appearance information, whose importance in object detection has been widely demonstrated [26, 28], and specially the idea of combining cues with the HOG descriptor, e.g., co-occurrence HOG [29], color HOG [24], etc. It has been largely demonstrated by the proposal of different descriptors that appearance is an important cue for object detection. The second source is the increasing trend of using segmentation for both detection and segmentation, which in our case has the potential of highlighting the shape of the human. Segmentation has been used for pixel-based object detection [2, 19], for providing shape-based outputs [15] and even also to generate candidate windows [25], so exploiting global image segmentation is likely to be useful also in other stages of the detection system. In fact, there exists a very related work by Ott et al. [23] which also combines the concepts of HOG descriptor and image segmentation. In such a work, given a window to be classified as human or not, a *soft segmentation* is carried out aimed at separating between foreground (human) and background pixels in order to compute an additional color-based HOG (CHOG) descriptor to be combined with the usual HOG in an augmented descriptor space. In our proposal we do not require to distinguish between foreground and background but rely on a global image segmentation. Besides, in our case we do not augment the original descriptor, thus not increasing the complexity of the classifier.

The outline of the paper is as follows. Sect. 2 describes the proposed algorithm and its parameters (segmentation method, and descriptor re-weighting approaches). The experimental results, including the details of the dataset and detection system, together with discussion, is presented in Sect. 3. Finally, the main conclusions and future work are summarized in Sect. 4.

2 HOG Re-weighting using Global Image Segmentation

The proposed approach consists in re-weighting the HOG descriptor for each one of the cells while it is computed. Basically, HOG consists in an intelligent grouping of gradient information (cells and blocks), as well as well-engineered histograms of gradient orientations (weighting by gradient magnitude, bin interpolation, histogram normalization and outliers clipping are the major steps[1]). The window of interest is covered by overlapping blocks, therefore, the HOG descriptor of the whole window usually ends up being thousand-dimensional. A linear SVM is used for learning the human classifier works in such high dimensional space. Accordingly, the obtained classifier is just a weighted summation running on such number of dimensions.

When computing HOG, the gradient orientation θ_P at a given pixel P is *weighted* by the corresponding magnitude μ_P, i.e., μ_P is accumulated in the histogram bin corresponding to θ_P(of course, taking into account the discretization, Gaussian weighting and interpolation proposed in [6]). Notice that the gradient at P, only encodes local differences in intensity or color, i.e., differences between adjacent pixels. In this paper, we want to incorporate differences based on a wider spatial support into the process in order to assess if they allow to obtain a human detector with higher performance. In particular, we want to weight μ_P by a given ω_P coming up from an image segmentation process, i.e., the vote of θ_P in the histogram will be the re-weighted magnitude λ_P instead of μ_P. Fig. 1 illustrates the idea.

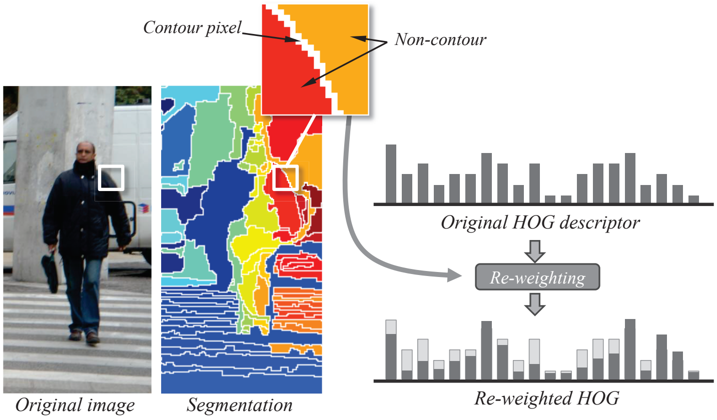


Fig. 1: Re-weighting of the HOG descriptor according to the image segmentation cues.

[1] The computation of the HOG has many details and we refer the reader to [6] for a comprehensive explanation.

2.1 Proposed Algorithm

Given an input image I and HOG parameters ϕ, our proposal can be summarized as follows:

1. Compute the image gradient of I, i.e., δ_I, using the corresponding parameters in ϕ.
2. Compute a global image segmentation of I using method $\mathcal{S}$, i.e., $\mathcal{S}(I) = \mathcal{S}_I$. Let $\mathcal{S}_I$ be the segmented image in which it is easy to distinguish the resulting segments and segment frontiers/contours.
3. Use δ_I, $\mathcal{S}_I$ and ϕ to compute the modified HOG of all desired windows of I. This means to proceed like for standard HOG but rather than weighting each orientation θ_P by its corresponding magnitude μ_P (in the histogram voting), we weight it by λ_P, where $\lambda_P = \omega_P * \mu_P$ and $\omega_P = \mathcal{W}(\mathcal{S}_I(P))$.

$\mathcal{W}$ is the weighting function of each pixel, which will be detailed in the next subsections. As an example, setting $\mathcal{W} = 1$ means that each pixel is not altered, thus getting the original HOG descriptor. Note that δ_I and $\mathcal{S}_I$ are computed at once over the whole I, i.e., they are not computed in a per window basis (except if I is a window).

2.2 Image Segmentation

Providing a spatial partition of an image, i.e., a segmentation, remains as an active topic in Computer Vision. Some of the open issues are the type of descriptors to use, the similarity criteria to merge and split regions or joint pixels, the combination of top-down and bottom-up approaches, etc. Indeed, there is a plethora of proposals in the literature for image segmentation task. Not surprisingly, it is one of the most difficult tasks of the popular PASCAL challenge [10].

In the context of human detection a relevant issue is real-time. Thus, after some initial tests, we discarded some possibilities as using [1] method for gradient computation, as well as other more sophisticated image segmentation techniques (graph-cuts, top-down/bottom-up fusion, etc.) [3, 18, 2, 1]. We did not consider basic methods as K-Means and watershed because parametrization can become a hard task, e.g., provide a good K for K-Means or appropriate markers for watershed was difficult. Instead, we relied on mean-shift algorithm applied to the CIE Luv color space because its computation is fairly fast and the parametrization is done in a relatively simple way. Moreover, we have chosen CIE Luv because the Euclidean distance between two colors in this space is strongly correlated with the human visual perception [30]. For instance, Fig. 2 shows the result of segmenting an image using mean-shift algorithm [4] with different parameters ($\mathit{\Gamma}$), computed in CIE Luv color space. Here we do not claim CIE Luv to be the most suited colorspace, therefore, in the future we want to consider other color spaces as well.

We compute the segmentation of image I with mean-shift algorithm $\mathcal{S}$ according to a set of different parameters that can be defined as $\mathit{\Gamma}$, so that $\mathcal{S}^{\mathit{\Gamma}}(I) = \mathcal{S}_I$. Such $\mathit{\Gamma}$ represents the bandwidth parameter of the mean-shift algorithm, such parameter takes into account the segmentation spatial radius, segmentation feature space radius and minimum segment area [4]. We tuned $\mathit{\Gamma}$ values according to the proposed in [4] in

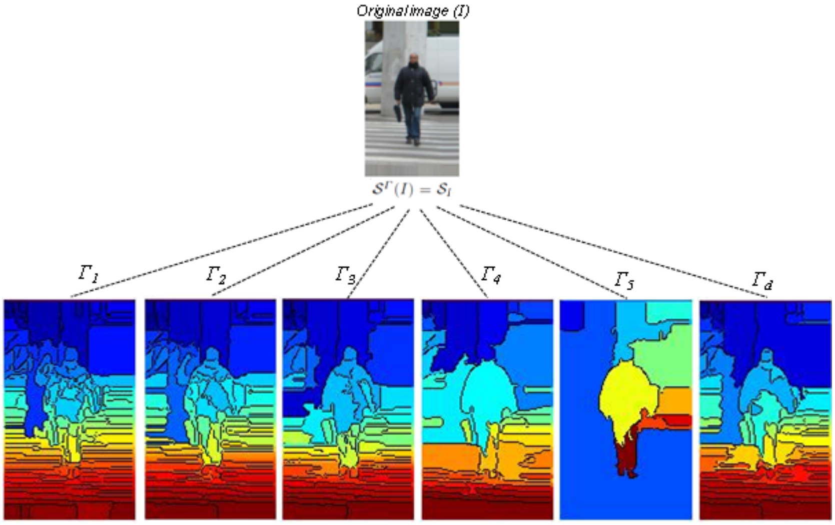


Fig. 2: Mean-shift image segmentation with different parameters (Γ).

order to obtain different number of segments and, therefore, homogeneous regions of dissimilar sizes, i.e., segmentation results that varies from smaller super-pixels to larger super-pixels. In Fig. 2 is illustrated the idea, different segmentations of an image where the number and size of segments varies according to the Γ used, the case of Γ_d shows the resulting image segmented with the default mean-shift parameters.

2.3 Pixel Weighting Functions

As previously mentioned, the simplest case of $\mathcal{W}$ can be defined as $\mathcal{W}(P) = 1$ without considering the pixel position (contour or non-contour), which would make the image segmentation useless since we would be computing the standard HOG. In fact, we are interested in rules of the form $\mathcal{W}(P) = \omega_c$ if P is a contour pixel in $\mathcal{S}_I$ and $\mathcal{W}(P) = 1$ otherwise.

In our case, ω_c of a pixel P depends of the location of P, i.e., is pixel-dependent. Such ω_c is defined as a dissimilarity measure between the neighbour segments of the contour to which P belongs. In particular, we have selected some basic features to establish our dissimilarity measures: color, luminance and texture, we also considered a combination between color and texture. Because of the simple nature of such measures, its computation is done in a simple way.

For each region, the color measure was computed in the CIE Luv colorspace by computing the average of the *u* and *v* components, then we determined the difference between the means by the Euclidean distance. In the case of luminance, the means were calculated in the L component of such CIE Luv colorspace, so the dissimilarity measure was done by the difference between the means. The texture was computed by the Battacharyya distance between the histograms of the adjacent regions, such histograms were calculated using local binary patterns, i.e., LBP values [16]. In the case of the combination between color and texture, each measure is computed separately, as explained above, and the average is calculated.

3 Experimental Results

In this section we evaluate the proposed algorithm in a publicly available dataset.

3.1 Dataset

In order to conduct the mentioned experiments, we make use of the INRIA person dataset [6], which contains color images. This dataset shows a wide range of human variations in pose, clothing, occlusions as well as complex backgrounds. Moreover, the dataset is divided in separated sets of null intersection for training and testing.

The training set contains 2,416 *positive* samples consisting in image windows (original and vertical mirror), each one containing a person framed by certain amount of background. All the positives are of the same size (*canonical detection window, 64x128*), although many of them come from an isotropic down scaling. We term this set of windows as $\mathcal{V}_{+}^{\text{train}}$. For collecting *negative* samples, i.e., image windows that do not contain persons, there are 1,218 person-free images available. We term this set of images as $\mathcal{I}_{-}^{\text{train}}$. The testing set consists of: (1) $\mathcal{I}_{-}^{\text{test}}$: 453 person-free images; (2) $\mathcal{I}_{+}^{\text{test}}$: 288 images containing labeled persons (ground truth); (3) $\mathcal{V}_{+}^{\text{test}}$: 1,126 positives analogous to the ones in $\mathcal{V}_{+}^{\text{train}}$ after cropping and mirroring the ground truth of $\mathcal{I}_{+}^{\text{test}}$.

3.2 Training

We use the standard training procedure for the INRIA dataset [6, 5]. First, we collect random negative windows from the images in $\mathcal{I}_{-}^{\text{train}}$ (10 windows per image to have 12,180 negatives) and down scale them to the size of the canonical detection window; let us call this set of windows $\mathcal{V}_{-}^{\text{train}}$. Then, given the sets $\mathcal{V}_{+}^{\text{train}}$ and $\mathcal{V}_{-}^{\text{train}}$, we compute the HOG of such labeled windows on top of the desired color space, and train a human classifier using the linear SVM. Finally, we run the corresponding human detector on $\mathcal{I}_{-}^{\text{train}}$ in order to follow the recommended *bootstrapping* technique, i.e., to append the set $\mathcal{V}_{-}^{\text{train}}$ with *hard negative windows* and re-train the human classifier. We apply two bootstrapping iterations.

3.3 Testing

In order to perform multi-scale human detection we use the *pyramidal sliding window* strategy as proposed in Dalal's PhD [5]. The original image is resized by a scaling factor s^i to obtain the image corresponding to the pyramid level i. Then, given a pyramid level, we shift the search window along the horizontal and vertical directions with a given stride. The smaller the scalling factor and window stride, the finer the sliding window search, so a better detection performance is expected. However, this is to the expense of a higher processing time. While Dalal [5] sets scaling to 1.2 and window stride to (8,8) pixels, in our experiments we found that a 1.05 scaling factor and (4,4) pixels stride provides a better tradeoff between processing time and performance. Additionally we perform anti-aliasing operations [11] which improve performance around a 9% in the INRIA dataset with respect to the original proposal in [5], which, in fact, makes more challenging to improve the standard HOG results.

Here, in order to compute the weighted HOG, we apply the selected image segmentation algorithm (i.e., $\mathcal{S}$) to each level of the pyramid. Thus, we obtain a sort of multi-scale segmentation of the original image (Fig. 3).

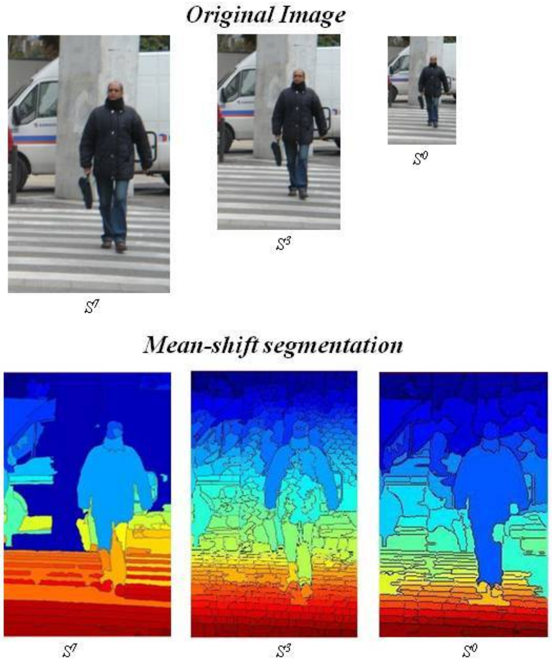


Fig. 3: Pyramid-segmentation. The scale of the slice in the pyramid affects the segmentation, similarly as it affects the HOG descriptor.

Since in multi-scale human detection a single person can be detected several times at slightly different positions and scales but a unique detection per human is desired, multiple overlapped detections shall be grouped by a clustering (*non-maximum-suppression*) procedure. In this case, we rely on the iterative confidence clustering approach of Laptev [20], which is a simpler and faster technique than Dalal's proposal and yields similar results.

3.4 Evaluation

In our experiments we use the widely extended *per image* evaluation procedure[2], which consists in running the detection system in a set of images containing persons and then comparing with the groundtruth for counting how many of such detections are true positives (T^{TP}) and how many are false positives (T^{FP}). If $\mathrm{I}^{\#}$ is the cardinality of $\mathcal{I}^{\mathrm{test}}$ and $\mathrm{H}^{\#}$ the number of labeled persons in $\mathcal{I}_{+}^{\mathrm{test}}$, then we can define the per image detection rate as $\mathrm{DR} = \mathrm{T}^{\mathrm{TP}}/\mathrm{H}^{\#}$ ($\mathrm{DR} \in [0,1]$; per image miss rate $\mathrm{MR} = 1 - \mathrm{DR}$) and the false positives per image as $\mathrm{FPPI} = \mathrm{T}^{\mathrm{FP}}/\mathrm{I}^{\#}$. In order to determine if a detection overlaps sufficiently with a labeled human of $\mathcal{I}_{+}^{\mathrm{test}}$ we follow the so-called PASCAL criterion [7] (also for bootstrapping during training). Now, we can define a curve to compare the algorithms. Note that FPPI can be greater than one, which would mean to have more than one false positive per image).

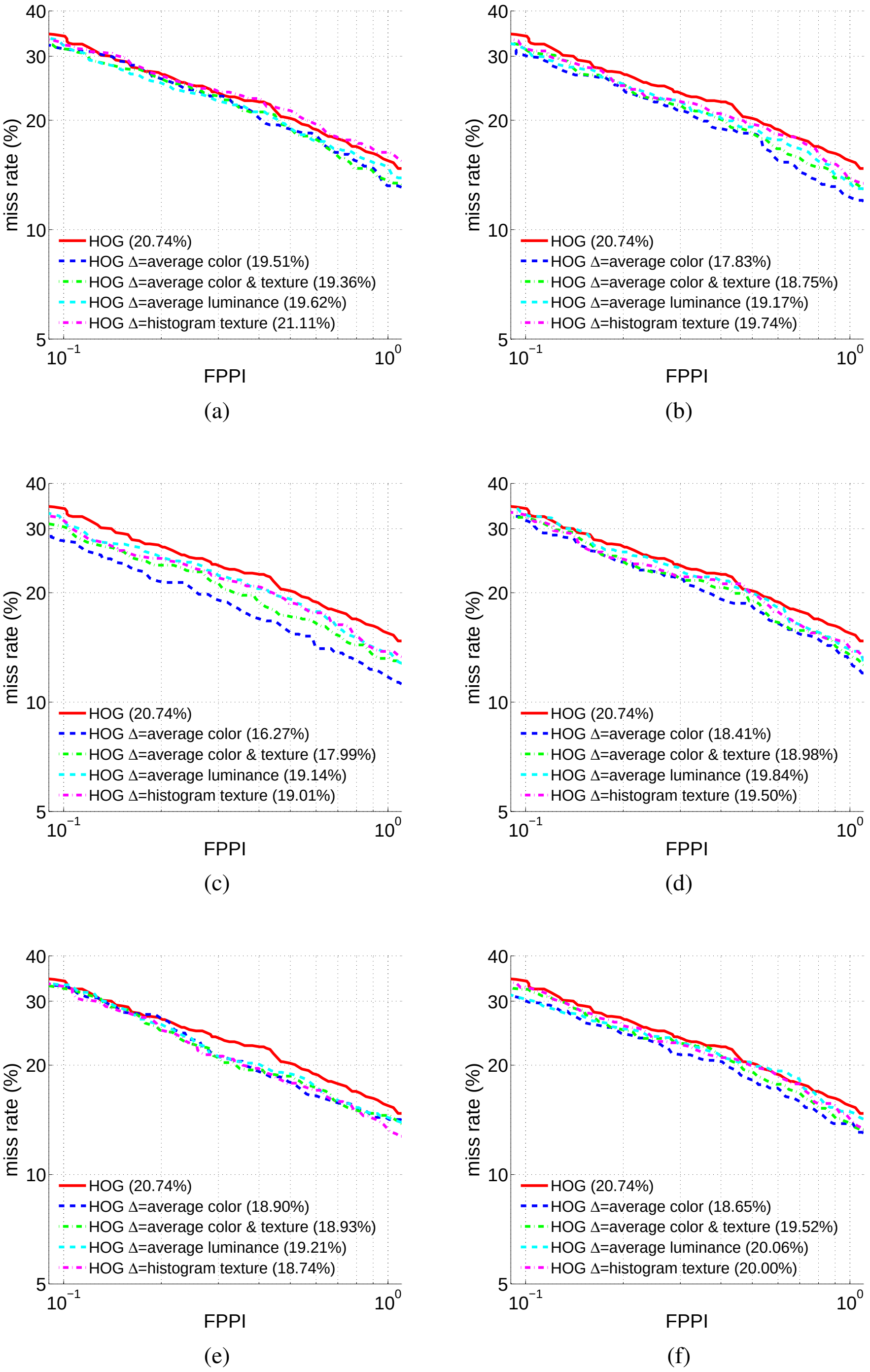
Taking into account $\mathcal{S}$ and $\mathcal{W}$, we have performed experiments in which the weight in the contour pixels is given by different criteria, $\omega_c = \Delta$. Such difference (Δ) is computed considering the variation in color, luminance (gray), texture (LBP) and combination of color and texture means between adjacent regions.

Fig. 4 illustrates the performance of the different algorithms. In all cases the result of the standard HOG is included for comparison. Within the legend parentheses, for the different plotted curves, we will indicate missrate at FPPI=10^0 and average area under the curve (A-AUC) between FPPI=10^{-1} and FPPI=10^0. Such FPPI points are relevant to detectors for driver assistance, the application field in which we will focus, taking into account that temporal coherence can help in reducing false positives if they are few per image.

3.5 Discussion

According to the experiments, it is clear that our proposal outperforms the standard HOG, i.e., the contribution of the segmentation cues in the computation of the HOG features seems to be significant for detecting pedestrians in the INRIA dataset. However, we can see how our proposal is sensitive to the segmentation output. A plausible explanation is the following, mean-shift with Γ_1 (Fig. 4a) outputs many small regions, also called super-pixels. Thus, such super-pixels are still too local, $i.e.$, too close to pixels size. In the case of mean-shift with Γ_2(Fig. 4b), the performance of the algorithm is slightly better due to the increased sizes of the segmented regions. On the contrary, Γ_4(Fig. 4d) and Γ_5(Fig. 4e) provides larger super-pixels so the information provided to the classifier is too general, $i.e.$, such contribution is not enough detailed. In the case of mean-shift with Γ_3(Fig. 4c) the resulting super-pixels are larger than mean-shift obtained with Γ_1(Fig. 4a) and Γ_2(Fig. 4b) but smaller than the obtained with Γ_4(Fig. 4d) and Γ_5(Fig. 4e) providing detailed information to the classifier but not too local. In the case of Γ_d(Fig. 4f) our proposal is computed with the default parameters provided by mean-shift method. The obtained results are quite good but not as good as

[2] Through the literature it has been demonstrated [7] that *per image* evaluation is more realistic than *per window* evaluation for assessing object detectors, which consists in classifying cropped examples and counterexamples, so in this paper we only use the former.



(a) (b)

(c) (d)

(e) (f)

Fig. 4: Performance curves of the different parameters (Γ) for the segmentation algorithm mean-shift. The curves (a)-(e) shows the results of our proposal with different segmentation parameters, obtaining from smaller to larger super-pixels. The diagrams (a)-(e) shows the results of the contribution coming from the segmentation with parameters Γ_1-Γ_5, in the case of (f) corresponds to the default parameters (Γ_d) of the mean-shift algorithm.

those obtained with $\mathit{\Gamma}_3$, therefore, it is necessary a validation set to adjust the segmentation parameters. However, in all the cases our proposal outperforms standard HOG descriptor, although mean-shift with $\mathit{\Gamma}_3$(Fig. 4c) achieves the best results.

Regarding the tested $\mathcal{W}$, the best option consists in weighting HOG at contour pixels (ω_c) according to the difference in color between the segments separated by the contour. Overall, using mean-shift ($\mathit{\Gamma}_3$), differences of color between segments for setting ω_c, we have down shifted missrate an average of 4.47% in our area of interest (from FPPI=10^{-1} to FPPI=10^0) compared to our HOG implementation explained in section 3.3.

An interesting further question is whether this improvement is maintained when combining HOG with other descriptors as it is normally done. In order to test this we have reproduced the recent HOG-LBP approach (combining HOG with local binary patterns), presented in [28, 31] with satisfactory results. In addition to the original LBP implementation, we have introduced three improvements with respect to [28] which increase its performance: (i) we use a threshold in the pixel comparisons, which increases the descriptor tolerance to noise; (ii) we do not interpolate the pixels around the compared central one; and (iii) we perform the computation directly in the luminance channel instead of separately computing the histograms in the three color channels. Fig. 5 illustrates the comparison among [28] (Wang's approach), our implementation of HOG-LBP and our proposal. As can be seen, our implementation already outperforms Wang's in 6.02% MR, and including the proposed segmentation-based weighting we further decrease MR to 7.31%, which demonstrates that our proposal is complementary to combining HOG with other cues.

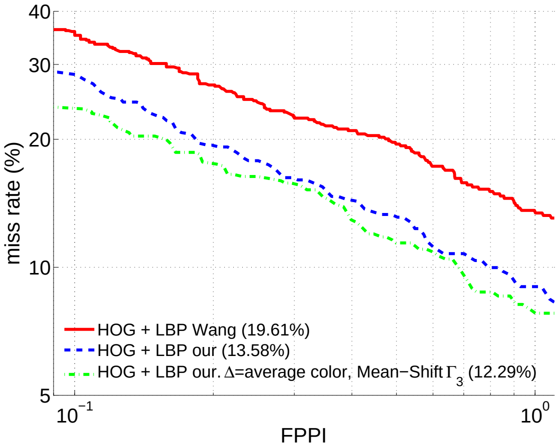


Fig. 5: Comparison between HOG-LBP with and without our proposed algorithm.

4 Conclusions

In this work we have investigated the possibility of improving HOG descriptors in the context of human detection. In particular by weighting HOG with information coming from image segmentation. We have conducted different experiments to clarify what type of segmentation is preferred (from smaller to larger super-pixels) and how such HOG reweighting must be performed. Overall, we have seen that using mean-shift with Γ_3 and differences of color between segments for setting HOG weights at contours, we have down shifted missrate an average of 4.47% in our area of interest (from FPPI=10^{-1} to FPPI=10^0). Furthermore, our proposal is complementary to combining HOG with other descriptors such as LBP, achieving a decrease of 1.29% in MR.

As future work we plan both to analyze the cases in which segmentation-based HOG weight is helping most in order to exploit this a prior information in the design of further detectors. Furthermore, we want to take advantage of the image segmentation for other tasks different than classification, e.g., candidate generation or refinement.

Acknowledgements

This work is supported by Spanish MICINN projects TRA2011-29454-C03-01, TIN2011-29494-C03-02, and Consolider Ingenio 2010: MIPRCV (CSD200700018. Y. Socarrás and D. Vázquez are supported by Universitat Autónoma de Barcelona, Spain.

References

1. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence99(1), 898–916 (2010)
2. Boix, X., Gonfaus, J., van de Weijer, J., Bagdanov, A., Serrat, J., Gonzàlez, J.: Harmony potentials for joint classification and segmentation. In: IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, CA, USA (2010)
3. Carreira, J., Sminchisescu, C.: Constrained parametric min-cuts for automatic object segmentation (2010)
4. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence24(5), 603–619 (2002)
5. Dalal, N.: Finding people in images and videos. PhD Thesis, Institut National Polytechnique de Grenoble / INRIA Rhône-Alpes (2006)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conf. on Computer Vision and Pattern Recognition. San Diego, CA, USA (2005)
7. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. on Pattern Analysis and Machine IntelligenceIn press (2011)
8. Enzweiler, M., Gavrila, D.: Monocular pedestrian detection: survey and experiments. IEEE Trans. on Pattern Analysis and Machine Intelligence31(12), 2179–2195 (2009)
9. Enzweiler, M., Gavrila, D.: A multi-level mixture-of-experts framework for pedestrian classification. IEEE Trans. on Image ProcessingIn press (2011)
10. Everingham, M., van Gool, L., Williams, C., Winn, J., Zisserman, A.: The PASCAL visual object classes (VOC) challenge. Int. Journal on Computer Vision88(2), 303–338 (2010)
11. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: IEEE Conf. on Computer Vision and Pattern Recognition. Anchorage, AK, USA (2008)

12. Gandhi, T., Trivedi, M.: Pedestrian protection systems: issues, survey, and challenges. IEEE Trans. on Intelligent Transportation Systems8(3), 413–430 (2007)
13. Gavrila, D.: A bayesian, exemplar-based approach to hierarchical shape matching. IEEE Trans. on Pattern Analysis and Machine Intelligence29(8), 1408–1421 (2007)
14. Gerónimo, D., López, A., Sappa, A., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. on Pattern Analysis and Machine Intelligence32(7), 1239–1258 (2010)
15. Gould, S., Gao, T., Koller, D.: Region-based segmentation and object detection. In: Advances in Neural Information Processing Systems. Vancouver, BC, Canada (2009)
16. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Trans. on Pattern Analysis and Machine Intelligence19(6), 1657–1663 (2010)
17. Jones, M., Snow, D.: Pedestrian detection using boosted features over many frames. In: IEEE Conf. on Computer Vision and Pattern Recognition. Anchorage, AK, USA (2008)
18. Kumar, M., Torr, P., Zisserman, A.: Objcut: Efficient segmentation using top-down and bottom-up cues. IEEE Trans. on Pattern Analysis and Machine Intelligence32(3), 530–545 (2010)
19. Ladicky, L., Sturgess, P., Alahari, K., Russell, C., P.H.S. Torr: What, where & how many? combining object detectors and crfs. In: European Conf. on Computer Vision. Crete, Greece (2010)
20. Laptev, I.: Improving object detection with boosted histograms. Image and Vision Computing27(5), 535–544 (2009)
21. Liao, C.T., Lai, S.H., Wang, W.H.: A hierarchical image kernel with application to pedestrian identification for video surveillance. In: IEEE Int. Conf. on Image Processing. Cairo, Egypt (2009)
22. Lin, Z., Davis, L.: Shape-based human detection and segmentation via hierarchical part-template matching. IEEE Trans. on Pattern Analysis and Machine Intelligence32(4), 604–618 (2010)
23. Ott, P., Everingham, M.: Implicit color segmentation features for pedestrain and object detection. In: Int. Conf. on Computer Vision. Kyoto, Japan (2009)
24. Rao, M., Vázquez, D., López, A.: Color contribution to part-based person detection in different types of scenarios. In: International Conference on Computer Analysis of Images and Patterns. Seville, Spain (201)
25. van de Sande, K., Uijlings, J., Gevers, T., Smeulders, A.: Segmentation as selective search for object recognition. In: Int. Conf. on Computer Vision. Barcelona, Spain (2011)
26. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conf. on Computer Vision and Pattern Recognition. pp. 511–518. Kauai, HI, USA (2001)
27. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. Int. Journal on Computer Vision63(2), 153–161 (2005)
28. Wang, X., Han, T., Yan, S.: An HOG-LBP human detector with partial occlusion handling. In: Int. Conf. on Computer Vision. Kyoto, Japan (2009)
29. Watanabe, T., Ito, S., Yokoi, K.: Co-occurrence histograms of oriented gradients for pedestrian detection. In: PSIVT. Tokyo, Japan (2009)
30. Wyszecki, G., Stiles, W.: Color science: concepts and methods, quantitative data and formulae. Wiley Series in Pure and Applied Optics
31. Zhang, J., Huang, K., Yu, Y., Tan;, T.: Boosted local structured hog-lbp for object localization. In: IEEE Conf. on Computer Vision and Pattern Recognition. Providence, RI, USA (2011)