Improving Gradient Histogram Based Descriptors for Pedestrian Detection in Datasets with Large Variations.

Prashanth Balasubramanian
Indian Institute of Technology Madras
Chennai, India.
bprash@cse.iitm.ac.in

Sarthak Pathak
The University of Tokyo
Tokyo, Japan.
pathak@robot.t.u-tokyo.ac.jp

Anurag Mittal
Indian Institute of Technology Madras
Chennai, India.
amittal@cse.iitm.ac.in

Abstract

Gradient histogram based descriptors, that are constructed using the gradient magnitudes as votes to orientation bins, are successfully used for Pedestrian Detection. However, their performance is hampered when presented with datasets having many variations in properties such as appearance, texture, scale, background, and object pose. Such variations can be reduced by smoothing the images. But, the performance of the descriptors, and their classifiers is affected negatively by this, due to the loss of important gradients along with the noisy ones. In this work, we show that the ranks of gradient magnitudes stay resilient to such a smoothing. We show that a combination of image smoothing and the ranks of gradient magnitudes yields good detection performances, especially when the variations in a dataset are large or the number of training samples is less. Experiments on the challenging Caltech and Daimler Pedestrian datasets, and the Inria Person dataset illustrate these findings.

1. Introduction

Pedestrian Detection is a significant task in automotive applications, surveillance systems, and assistive technologies for the challenged among others. Apart from the common challenges like clothing and pose variations, there are also additional ones such as clutter, scale changes, blur, Gaussian noise, compression artefacts, illumination and background changes. Any reliable pedestrian detection system will have to address these challenges to meet the stringent demands posed by real-world scenarios.

The literature is replete with attempts to solve the problem of pedestrian detection. [16, 22] and [27] suggested the use of Haar or Haar-like wavelets to learn the patterns of pedestrians and detect them in a sliding window manner. To handle more variations, Dalal and Triggs [4] presented their landmark paper that used the Histograms of Gradient features with an SVM classifier. HOG [4] was subsequently improved by incorporating colour features [17, 28], texture reasoning [29, 24], motion cues [30], modifying and extending to multiple scales [13], improving run-time efficiency [32], and being used as a low-level component for other features in [1, 11]. Felzenszwalb et al. [10] included contrast-sensitive gradient directions and texture information, following an interesting eigenvalue analysis of HOG vectors. This revealed that the top few HOG eigenvectors exhibit a certain pattern that can be used to reduce the dimensionality of HOG. Dollar et al. [8] have provided an extensive survey of the state-of-the-art in Pedestrian Detection. They benchmark the performance of many descriptors such as HOG-LBP [29], Felzenszwalb's HOG [10], Multiple Features and Colour Self Similarity [28], and those proposed in Partial Least Squares Analysis [24], Feature Synthesis [1]. They also observe that the information contained in gradient histograms is an important component of most of the competitive descriptors. Recent work [31, 5, 23, 7, 2, 21, 3, 6] use information such as motion, spatially-pooled covariance [26], features of different colour spaces or a combination of these in addition to gradient histograms with Random Forests classifier. [2] also studies the effect of using appropriate feature pools for effective Pedestrian Detection. Works based on deep-nets have also been attempted, refer [18, 19, 20].

Dollar et al.[8] also introduce a dataset for on-board pedestrian detection whose images have many variations,

such as object pose, appearance, texture, scale, and background changes that challenge a pedestrian detector. These variations can be suppressed to an extent by smoothing the images. However, doing so inherently affects the gradient magnitudes which leads to a reduction in the detection rate. As smoothing is a non-linear change, linear normalization techniques are not sufficient to recover the important gradients. In this paper, we show that a *non-linear normalization technique* based on the ranks of gradient magnitudes within a window is able to recover these gradients to a large extent. To suppress any spurious gradients, it is combined with a simple *min-max* normalization technique. This combination gives good gains in performance especially in datasets with many variations.

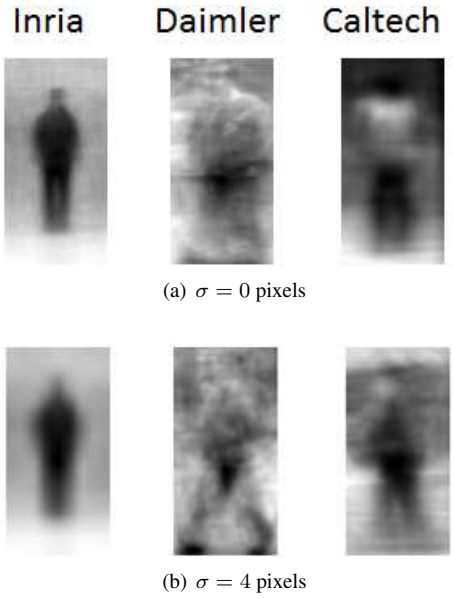
Apart from the notable attempts in [14, 29] that use LBPs or their variations, non-linear ranking techniques have seldom been used for the task of Pedestrian Detection. We propose new techniques which use the robust *ranks* of the gradient-magnitudes of the pixels within a window to form the votes to the bins of the orientation histograms. The major contribution of this work is to improve gradient histogram based descriptors on challenging real-world scenarios that contain many variations. We show that smoothing the images in the training and testing phases, followed by our normalization techniques can increase performance by good margins.

We validate our study on the Caltech Pedestrian Detection dataset [8], the Daimler Pedestrian Dataset [9] and the Inria Person Dataset [4]. The proposed techniques have achieved an improvement over HOG by ~20% on the Caltech [8] dataset, and by ~12% on the Daimler [9] set in the acceptable operating ranges on the *FPPW* axis. We also achieve competitive results on the Inria person dataset [4]. Further, we perform experiments by reducing the sizes of the training sets in order to show the capacity of the proposed techniques to effectively represent many variations of the object (i.e. pedestrian). We also demonstrate the ability of the proposed techniques to work in conjunction with other complementary information to improve their individual performances.

The paper is organized as follows. In Section 2, we present the motivation to smooth the images and use the ranks of gradient magnitudes in building the descriptor. Section 3 describes the proposed techniques in detail, analysing their behaviour and gives the construction of the feature descriptors. Experimentation and results are presented in Section 4. Finally, in Section 5, we summarize and conclude this paper.

2. The Effect of Smoothing on Variations

In any real world pedestrian dataset, such as the Caltech Dataset [8], there exist a large number of variations in the positive images, apart from the usual pose and clothing variations, such as clutter, scale changes, blur, Gaussian noise, compression artefacts, and background and illumination changes. These can greatly affect the performance of any pedestrian detection algorithm. Figure 1(a) shows the average intensity images for three datasets - Inria [4], Daimler [9], and Caltech [8]. Essentially, these show the extent of the variations that are present in each dataset. Evidently, the image for the Inria dataset in Figure 1(a) is quite sharp and appears more human-like, implying the lesser number of variations in this dataset. Hence, the computation of gradients at the finest scale, without smoothing was suggested by Dalal and Triggs [4].



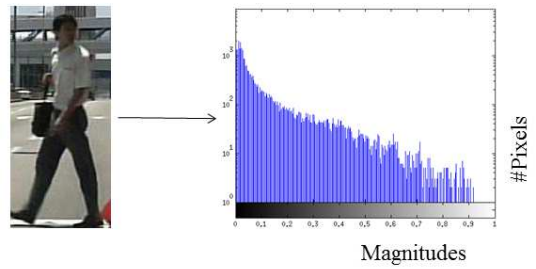
(a) $\sigma = 0$ pixels

(b) $\sigma = 4$ pixels

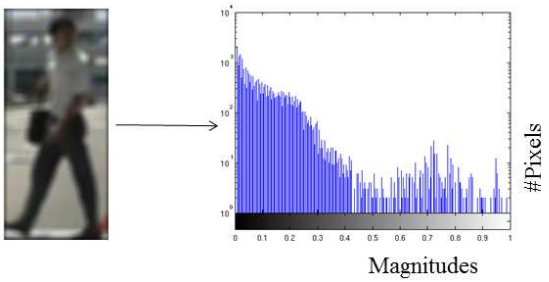
Figure 1. Figure best viewed in colour. The average intensity images for Inria, Daimler and Caltech datasets, before and after smoothing, are shown in Figures 1(a) and 1(b). These give a good measure of the variations present in each dataset. The images were smoothed by convolving with a Gaussian kernel of size 4 pixels.

In datasets with large variations, however, we observe that smoothing can decrease the impact of these variations as it reduces clutter, noise, compression artefacts, and weaker gradients. Figure 1(b) portrays the average images of the blurred counterparts shown in Figure 1(a). We notice that all the images have improved in their portrayal of the average 'pedestrian'.

However, smoothing (which can also be inherently present in images in the form of motion blur or defocussing) strongly affects gradient magnitudes in a non-linear manner. As can be seen in Figure 2, stronger gradients (which occur on the right ends of the histograms in Figure 2) are especially affected. These gradients are more important as they might represent object contours, and hence, it is imperative to retain them in order to have a better classification or de-



(a) $\sigma = 0$ pixels



(b) $\sigma = 4$ pixels

Figure 2. Figure best viewed in colour. The effect of smoothing on gradient magnitudes. The histograms of gradient magnitudes for a positive window from the Caltech dataset [8] before and after smoothing are shown in Figures 2(a) & 2(b). The stronger gradient magnitudes (on the right side of the histograms) that can represent object contours are shifted significantly under smoothing, as seen in Figure 2(b).

tection result. Furthermore, since smoothing is a non-linear process, a normalization scheme, such as L_2-hys * which is a clipping function and so is essentially (piece-wise) linear, cannot deal with such losses effectively. Hence, a non-linear normalization scheme needs to be adopted. Such a normalization can effectively recover gradients that were lost in the smoothing. This could result in better detection rates, since the smoothing gets rid of clutter and unimportant gradients.

3. Robust ranking of gradient magnitudes

The objective of the ranking procedure is to effectively recover important gradients that were lost in smoothing the image. This needs to be done keeping in mind the non-linearity imposed by the process of smoothing. The gradient magnitudes are ordered such that the ranks are designed to be robust to non-linear changes in gradient magnitudes induced by smoothing.

To this end, the space of gradient magnitudes in the window is partitioned into R ranges such that each range encompasses nearly an equal number of gradient magnitude values. The limits of each range are decided in the following manner. Let $m(\mathbf{X})$ denote the gradient magnitude of $\mathbf{X}$. Let $S = \{m_i\}_{i=0}^{n-1}$ be the set of distinct, *sorted* gradient magnitudes of the window, where i is the index of m_i when the gradient magnitudes are sorted. Partition S into R ranges (an integer parameter) using $\{m_{\gamma_0}, ..., \ m_{\gamma_{R-1}}, \ m_{\gamma_R} = m_{n-1}\}$ such that

$$S = \bigcup_{i=0}^{R-2} \left[m_{\gamma_i}, \ m_{\gamma_{i+1}}\right) \bigcup \left[m_{\gamma_{R-1}}, m_{\gamma_R}\right], \ \ m_{\gamma_i} \leq m_{\gamma_{i+1}}$$

We define index γ_i as:

$$\gamma_i = i \left\lfloor \frac{n}{R} \right\rfloor, \ \ 0 \leq i < R < n, \ \ i \in \mathbb{Z}_{\geq 0} \qquad (1)$$

The rank, $r_{nl}(\mathbf{X})$, of pixel $\mathbf{X}$ is defined as

$$r_{nl}(\mathbf{X}) = \begin{cases} i+1, & m_{\gamma_i} \leq m(\mathbf{X}) < m_{\gamma_{i+1}}, \\ R, & m_{\gamma_{R-1}} \leq m(\mathbf{X}) \leq m_{\gamma_R}, \\ & 0 \leq i \leq R-2, \ \ i \in \mathbb{Z}_{\geq 0} \end{cases}$$

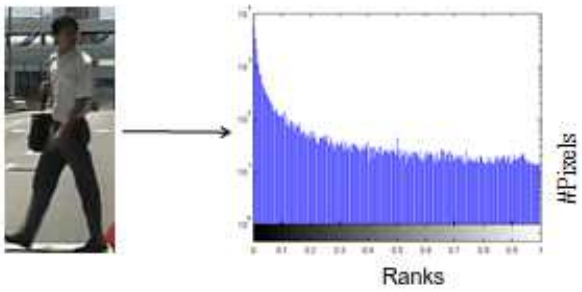
For example, if $R = 10$, and the number of distinct gradient magnitudes in the window are 100, then the pixels of the window having the first 10 gradient Magnitude values are assigned rank 1, the next 10 are assigned rank 2, and so on. This type of a division ensures that under non-linear changes induced by smoothing, the rank of a particular pixel remains the same, because the ranges move accordingly. In our experiments, we have used $R = 256$.† Although Ordinal and Spatial Intensity Distribution (*OSID*) by Tang et al., [25] use the technique in Equation 1 for pixel intensities, their method has been primarily designed for feature point matching and making it robust against monotonic changes in illumination, while the proposed technique operates in the gradient space and aims at making an object descriptor robust for classification or detection.

In case of non-linear changes brought about by smoothing, the ranks are affected little, see Figure 3. The figure shows the distribution of ranked-magnitudes of a positive window from the Caltech dataset [8] before and after smoothing the window with a Gaussian kernel of size 4 pixels. We notice that the distributions of ranks of the original and smoothed images are similar, highlighting the robustness of the usage of ranks over that of gradient magnitudes.

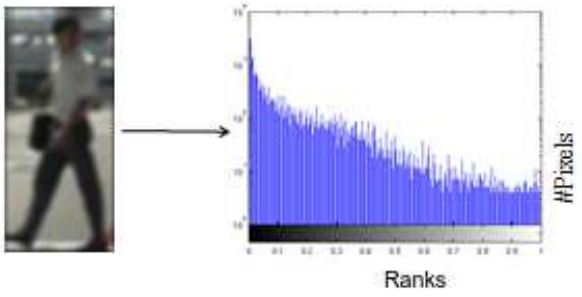
As the aforesaid scheme attempts to maintain an almost equal number of unique gradients per range, some

*L_2-hys proposed by Lowe [12] curbs down the effect of large gradients by L_2 normalizing the orientation histogram followed by clipping each value to be no larger than 0.2.

†If $R = 256$, the procedure is similar to histogram equalization of gradient magnitudes, while setting $R = 1$ discards the magnitudes completely. The latter is comparable to LBP [15] in the way pixel information is used.



(a) $\sigma = 0$ pixels



(b) $\sigma = 4$ pixels

Figure 3. Figure best viewed in colour. The effect of non-linearly ranking the gradient magnitudes is demonstrated here. The distribution of the ranks, $r_{_{nl}}$ (for $R = 256$), of the pixels of the pedestrian window, before and after smoothing, are shown in Figures 3(a) and 3(b). The distribution changes little under smoothing.

of the low-valued gradients can become spurious, and so one could employ a linear ranking procedure $r_{_{minmax}}$ using min-max normalization to curb these.

Simply put, S is partitioned using $\{m_0,\ m_1,\ ...,\ m_R\}$ into R ranges such that they are all equi-sized. Here, m_0 & m_R are the minimum and the maximum gradient magnitudes.

The rank, $r_{_{minmax}}(\mathbf{X})$, of pixel $\mathbf{X}$, using this normalization, is defined as

$$r_{_{minmax}}(\mathbf{X}) = \begin{cases} 1, & m_0 \leq m(\mathbf{X}) \leq m_1 \\ i+1, & m_i < m(\mathbf{X}) \leq m_{i+1}, \\ & 1 \leq i < R,\ \ i \in \mathbb{Z}_{\geq 0} \end{cases}$$

Please note that the ranking scheme $r_{_{minmax}}$, although helpful in suppressing spurious gradients, is a linear method and so suffers from the demerits explained in Section 2. It can, however be combined with the ranking scheme $r_{_{nl}}$ as they yield complementary benefits. The experimental Section 4 shows the efficacy of such a combination in terms of classification performance.

We use orientation histograms to build the feature vectors of images. The construction of the orientation histograms closely follows that of HOG[4, 10]. It is to be noted that the orientation bin to which a pixel contributes its vote is calculated based on its gradient direction. The proposed descriptors are obtained by using the *ranks* (of gradient magnitudes), instead of the raw gradient magnitudes, to form the *votes* to the bins of the orientation histograms. Specifically, the *vote* that a pixel contributes is its *rank* rather than its gradient magnitude. By individually using either $r_{_{nl}}$ or $r_{_{minmax}}$ on Dalal's HOG [4], and its derivative *felz* by Felzenszwalb et al.[10], we obtain 4 variants: *r-nl_HOG*, *r-minmax_HOG*, *r-nl_felz* and *r-minmax_felz*. A fifth variant of Dalal's HOG, *r-minmax-nl_HOG* is obtained by concatenating the descriptors obtained by the two ranking schemes and aims to tap the advantages of both.

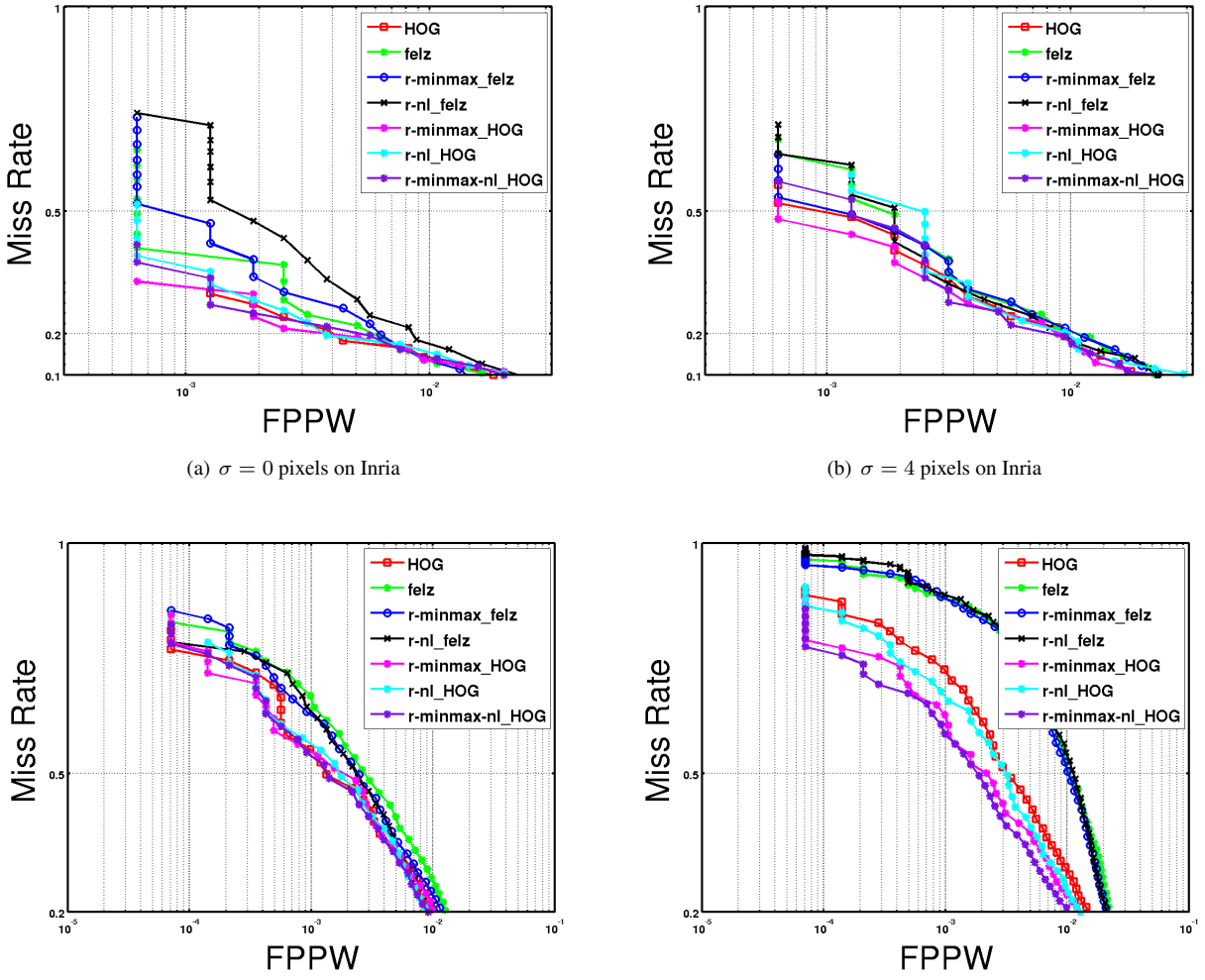
4. Experimentation and Results

Descriptors based on gradient histograms and their many variants have been successfully used for the task of Pedestrian Detection. Dalal's [4] & Felzenszwalb's HOGs [10], HOG-LBP [29], Multiple Features and Colour Self Similarity [28], PLS [24], Colour Segmentation Features [17], and Integral & Aggregated Channel Features [7, 5, 31], Roerei & Spatially-pooled detectors [21, 2] either add information to HOG-like gradient histograms or use such gradient histograms in an appropriate form. In this paper, we compare the performances of the proposed techniques with the baseline performances of Dalal's [4] and Felzenszwalb's HOGs [10]. It would be illustrative to study the effect of the proposed methods on the other variants in a future work.

For all the experiments, linear kernel SVM was chosen as the classifier, owing to its efficiency in runtime, and the images were resized to 64 x 128 before computing the descriptors. Cells were defined as 8 x 8 pixels and blocks as 2 x 2 cells with a stride factor of 1 cell in each direction. The performance comparisons are brought out by the *Detection Error Trade-off* (DET) curves i.e. Miss Rate vs. False Positives Per Window (FPPW) on a log-linear scale Since they characterize Miss Rates, lower their values better is the performance.

We conducted our experiments on three datasets: the Inria Person Dataset [4], the Caltech Pedestrian Detection Benchmark [8] and the Daimler Pedestrian Dataset [9]. As *FPPW* metric is used, for evaluating on the Caltech dataset [8], we extracted all pedestrians of heights ≥ 20 pixels from the Caltech-USA test set (*set06 - set10*) and classified them.

We show the performance plots resulting from three different experiments: 1) images smoothed with kernels of differing radii to minimize the variations of the dataset and study their impact on the performance of the descriptors. 2) training sets of differing sizes to analyse how well each descriptor handles the variations of the dataset. 3) showing how the proposed technique can be used in combination with other descriptors having complementing information to characterize an object better.



(a) $\sigma = 0$ pixels on Inria

(b) $\sigma = 4$ pixels on Inria

(c) $\sigma = 0$ pixels on Daimler

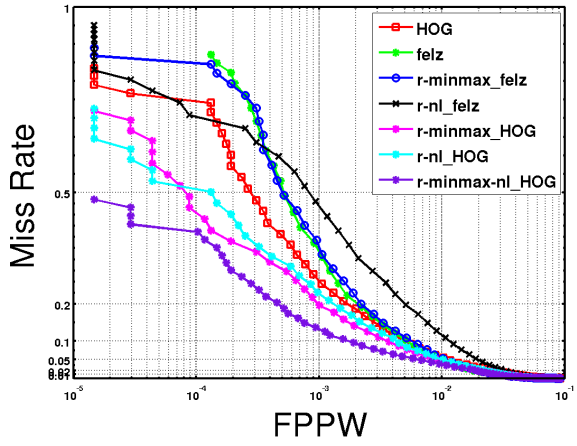
(d) $\sigma = 4$ pixels on Daimler

Figure 4. Figure best viewed in colour. Performance on the Inria [4](row-1) and the Daimler [9](row-2) datasets for Gaussian smoothing kernels of different radii.

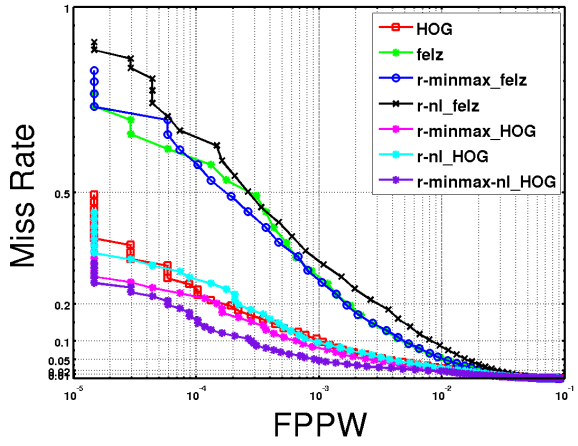
4.1. Impact of Smoothing

Smoothing affects the gradients of the image non-linearly, and two competing factors occur - one is the reduction of variations and the other is the reduction of gradient magnitudes that are useful for detection. Both these changes counter each other and can impact the performance depending on which one gains dominance over the other. To understand the effect, we convolved both the training and the test images with Gaussian smoothing kernels of differing radii and studied the performance of various descriptors. The plots in Figures 4 and 5 illustrate these.

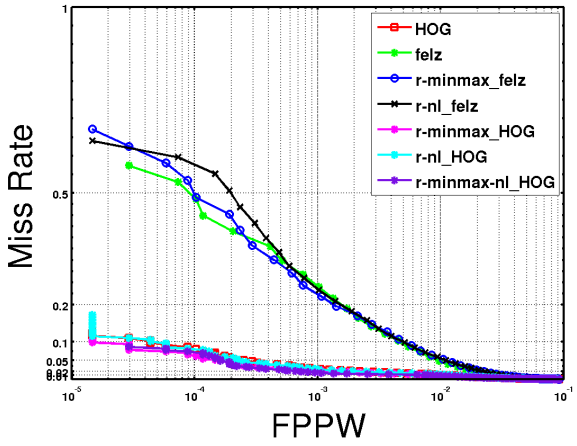
Row-1 of Figure 4 shows the effect of smoothing on the images of the Inria dataset[4]. As Inria images are focused, centred, and have little variations, the gradients which are well-defined when $\sigma = 0$ get affected, on smoothing, beyond the point of recovery, and so, lead to a loss of discriminative information. Hence, we see a dip in the performance of all the descriptors moving from $\sigma = 0$ to $\sigma = 4$. Row-2 of Figure 4 shows the behaviour of the descriptors on the Daimler dataset [9]. The proposed HOG-variants yield performance comparable with HOG when $\sigma = 0$ and improve by ~12% when $\sigma = 4$ and *FPPW* $\in \left[10^{-5}, 10^{-4}\right]$. We notice that the performances of all the descriptors go down on moving from $\sigma = 0$ to $\sigma = 4$ pixels. The proposed descriptors suffer by ~4% while the others suffer by more than 12%. The reason for the small dip in performance may be due to the fact that Daimler images, though more challenging than Inria, still contain significant useful information at the native scale which is lost in smoothing and is hard to recover.



(a) $\sigma = 0$ pixels



(b) $\sigma = 2$ pixels



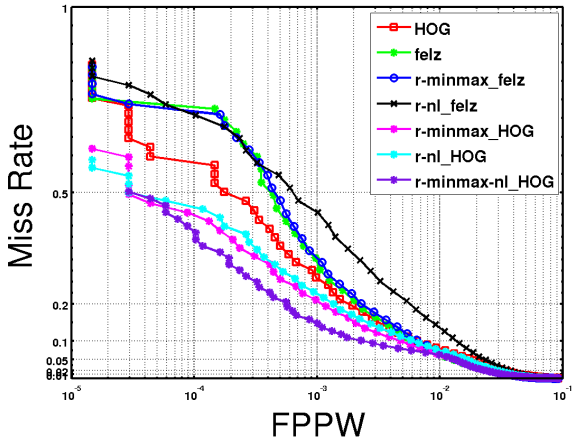
(c) $\sigma = 4$ pixels

Figure 5. Figure best viewed in colour. Performance on the CALTECH [8] dataset for Gaussian smoothing kernels of different radii.

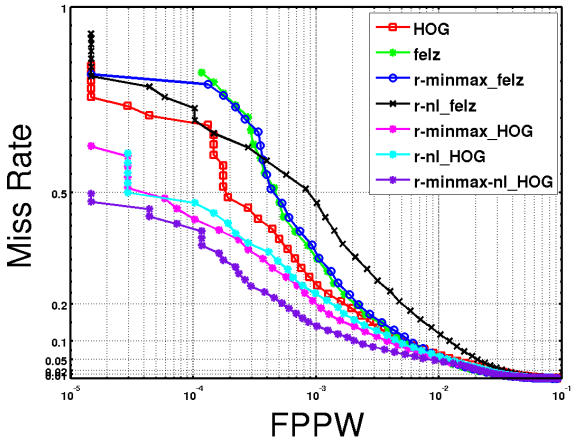
Figure 5 shows the performances of the descriptors on the Caltech dataset [8]. Smoothing results in an increase in the performances of our techniques – the combined descriptor improves its miss-rate from ~48% at $\sigma = 0$ to ~25% at $\sigma = 2$ at 10^{-5} *FPPW* – which are consistently better than HOG [4] with a gain of $\geq 12\%$ when *FPPW* $\in \left[10^{-5}, 10^{-4}\right]$. In the same *FPPW* range, when $\sigma = 0$, the proposed combined technique gives a gain of ~25%. The reason for the large gain despite no smoothing is due to the many number of non-linear variations in the Caltech dataset [8] which are better represented by the ranks that are used to build the descriptor. Also when $\sigma = 4$, the gain of the proposed descriptors over HOG is nullified. This may be because the images have been overly smoothed, and so have lost almost all of the valuable gradients without a possible recovery and thus saturate the performances of the descriptors.

4.2. Changing Training Set Size

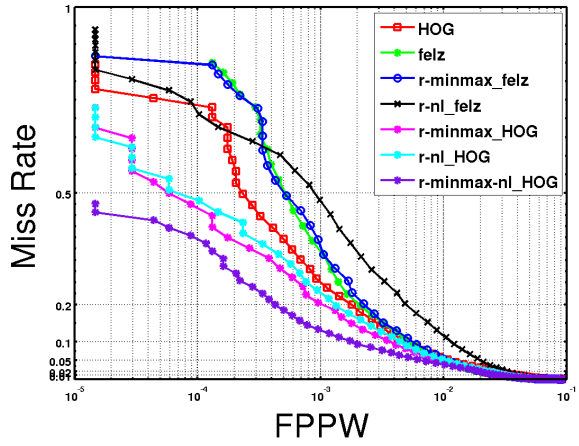
The necessity of a large training set is to capture as many generic variations of the object as possible, although, a robust descriptor should be able to sufficiently represent the object in its generality even with a smaller number of samples. In order to study this property of the proposed technique, we measured the detection performances as a function of the sizes of the training set from the three datasets. Changing the sizes of the training sets can be seen as a way to increase the number of variations in the test set as well. Subsets of sizes ranging from 10% - 90% were randomly sampled from the training sets of the datasets, the classifiers were trained on them and then tested on the corresponding entire test sets.



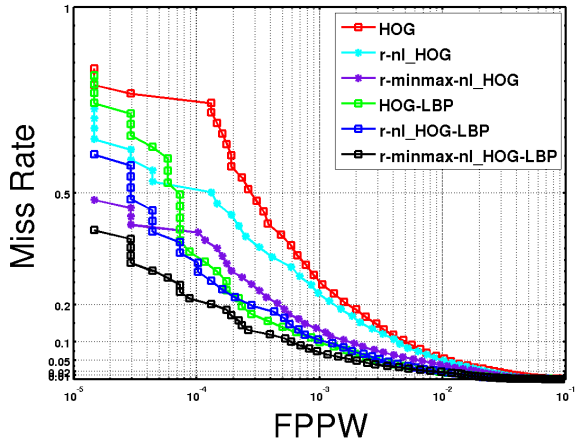
(a) Training set size = 30%



(b) Training set size = 60%



(c) Training set size = 90%



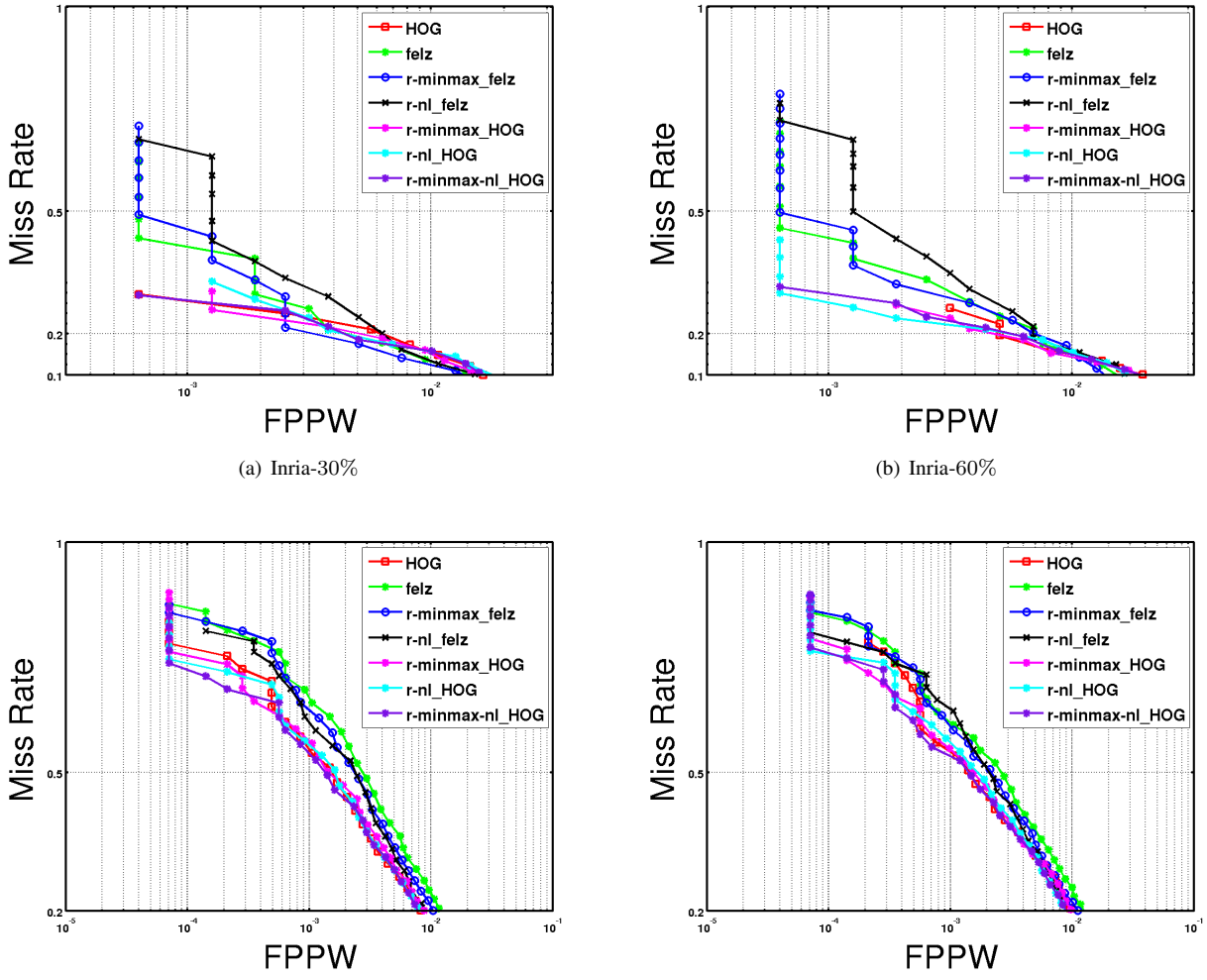
(d) Concatenation with the LBP descriptor

Figure 6. Figure best viewed in colour. Figures 6(a),6(b) and 6(c) portray the study of the impact of training set sizes on the detection performance on the Caltech [8] dataset. Figure 6(d) shows the performance of the LBP variants on the Caltech dataset [8].

Figure 6(a), 6(b) and 6(c) show the detection performances on the Caltech-USA test set when the classifiers were trained on 3 subsets of the Caltech training set [8]. We observe that the proposed techniques improve over HOG [4] at all points on the *FPPW* axis for all the 3 subsets. When *FPPW* $\in \left[10^{-5}, 10^{-4}\right]$ the gain is nearly 20% for the combined descriptor showing its ability to represent an object(i.e. Pedestrian) better. Figure 7 shows the performances of the descriptors on the entire Inria and Daimler test sets, for 2 training subsets. Inria being a dataset with not many variations, the performances are nearly the same, as seen in Figures 7(a) and 7(b), though when *FPPW* $\in \left[10^{-5}, 10^{-4}\right]$, HOG and its proposed variants are better than Felzenszwalb's HOG and its variants. On the Daimler dataset also, in Figures 7(c) and 7(d), the proposed variants of HOG show improvements. Although not as large as in Figures 6(a), 6(b) and 6(c) (due to the nature of Daimler set [9]), the gains in the performances are ~5% for *FPPW* $\in \left[10^{-5}, 10^{-4}\right]$, showing the ability of the proposed descriptors to represent an object when the training sets are small in their sizes.

4.3. Concatenation with Other Descriptors

The use of information in addition to gradient histograms can improve detection rates, refer [8]. The additional information is usually complementary to that provided by the gradient histograms. We study one particular variant - HOG-LBP [29] - to portray the feasibility of the proposed technique as a concatenating component to other complementary information. Figure 6(d) shows the performance



(a) Inria-30%

(b) Inria-60%

(c) Daimler-30%

(d) Daimler-60%

Figure 7. Figure best viewed in colour. (a) Study of the impact of training set sizes(30% & 60%) on the detection performance on the Inria [4] (Fig.7(a) & 7(b)) and the Daimler [9] (Fig.7(c) & 7(d)) datasets.

of the proposed technique in conjunction with the LBP [15] operator on the Caltech dataset [8]. When used with LBP [15], the proposed techniques improve over themselves(viz. the non-LBP versions), HOG, and HOG-LBP [29], especially when $FPPW \in \left[10^{-5}, 10^{-4}\right]$, showing their ability to perform well when concatenated with other information.

5. Conclusions

We have proposed a technique that improves gradient histogram-based features for pedestrian detection when the dataset has a large number of variations. Smoothing the images prior to gradient-computation helps in reducing these variations. To recover the object gradients lost when smoothing, a rank-based, non-linear technique is used. A simple min-max technique is also used to curb spurious gradients. Good gains in performance are seen both in the CALTECH and the DAIMLER Pedestrian sets for different smoothing amounts. It is also shown that concatenating the proposed descriptors with others containing complementing information works well in practice. It would be illustrative to study the effect of the normalizations of the proposed techniques on other variants of gradient histograms. Also, it would be of interest to figure out an online procedure to determine the optimal amount of blur as well as to see the effect of this kind of normalization on flow features which will be affected even more by variations in datasets.

References

[1] A. Bar-Hillel, D. Levi, E. Krupka, and C. Goldberg. Part-based feature synthesis for human detection. In *The Proceed-*

ings of the 11th European Conference on Computer Vision, volume 6314 of *Lecture Notes in Computer Science*, pages 127–142. Springer Berlin Heidelberg, 2010.

[2] R. Benenson, M. Mathias, T. Tuytelaars, and L. Van Gool. Seeking the strongest rigid detector. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3666–3673, June 2013.

[3] R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? In *The Proceedings of the Workshops of the 13th European Conference on Computer Vision*, volume 8926 of *Lecture Notes in Computer Science*, pages 613–627. Springer International Publishing, 2015.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[5] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

[6] P. Dollár, R. Appel, and W. Kienzle. Crosstalk cascades for frame-rate pedestrian detection. In *The Proceedings of the 12th European Conference on Computer Vision*, volume 7573 of *Lecture Notes in Computer Science*.

[7] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral channel features. In *The Proceedings of the British Machine Vision Conference*, 2009.

[8] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, 2011.

[9] M. Enzweiler and D. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, Dec 2009.

[10] P. F. 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*, 32(9):1627–1645, 2010.

[11] Z. Lin and L. Davis. A pose-invariant descriptor for human detection and segmentation. In *The Proceedings of the 10th European Conference on Computer Vision*, volume 5305 of *Lecture Notes in Computer Science*, pages 423–436. Springer Berlin Heidelberg, 2008.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[13] S. Maji, A. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[14] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou. Discriminative local binary patterns for human detection in personal album. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[15] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, Jul 2002.

[16] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–199, 1997.

[17] P. Ott and M. Everingham. Implicit color segmentation features for pedestrian and object detection. In *The Proceedings of the IEEE International Conference on Computer Vision*, pages 723–730, 2009.

[18] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. In *The Proceedings of the IEEE International Conference on Computer Vision*, pages 2056–2063, Dec 2013.

[19] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. In *The Proceedings of the IEEE International Conference on Computer Vision*, pages 2056–2063, Dec 2013.

[20] W. Ouyang, X. Zeng, and X. Wang. Modeling mutual visibility relationship in pedestrian detection. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, June 2013.

[21] S. Paisitkriangkrai, C. Shen, and A. van den Hengel. Strengthening the effectiveness of pedestrian detection with spatially pooled features. In *The Proceedings of the 13th European Conference on Computer Vision*, volume 8692 of *Lecture Notes in Computer Science*, pages 546–561. Springer International Publishing, 2014.

[22] C. Papageorgiou and T. Poggio. Trainable pedestrian detection. In *The Proceedings of the International Conference of Image Processing*, volume 4, pages 35–39, 1999.

[23] D. Park, C. Zitnick, D. Ramanan, and P. Dollar. Exploring weak stabilization for motion feature extraction. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2882–2889, June 2013.

[24] W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis. Human detection using partial least squares analysis. In *The Proceedings of the IEEE International Conference on Computer Vision*, pages 24–31, Sept 2009.

[25] F. Tang, S. H. Lim, N. Chang, and H. Tao. A novel feature descriptor invariant to complex brightness changes. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2631–2638, 2009.

[26] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, Oct 2008.

[27] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *The Proceedings of the IEEE International Conference on Computer Vision*, pages 734–741 vol.2, Oct 2003.

[28] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1030–1037, 2010.

[29] X. Wang, T. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. In *The Proceedings of the*

IEEE International Conference on Computer Vision, pages 32–39, 2009.

[30] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 794–801, 2009.

[31] J. H. H. Woonhyun Nam, Piotr Dollár. Local decorrelation for improved pedestrian detection. In *Advances in Neural Information Processing Systems 27*, 2014.

[32] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1491–1498, 2006.