Characterisation of the spatial sensitivity of classifiers in pedestrian detection

Daniel Quinteros[1], **Sergio A Velastin**[1,2], **Gonzalo Acuña**[1]

[1]Universidad de Santiago de Chile, Chile, {daniel.quinterosc,gonzalo.acuna}@usach.cl
[2]Universidad Carlos III de Madrid, Spain, sergio.velastin@ieee.org

Keywords: Pedestrian detection, Spatial sensitivity, NMS, HOG, Classifier.

Abstract

In this paper, a study of the spatial sensitivity in the pedestrian detection context is carried out by a comparison of two descriptor-classifier combinations, using the well-known sliding window approach and looking for a well-tuned response of the detector. By well-tuned, we mean that multiple detections are minimised so as to facilitate the usual non-maximal suppression stage. So, to guide the evaluation we introduce the concept of spatial sensitivity so that a pedestrian detection algorithm with good spatial sensitivity can reduce the number of classifications in the pedestrian neighbourhood, ideally to one. To characterise spacial sensitivity we propose and use a new metric to measure it. Finally we carry out a statistical analysis (ANOVA) to validate the results obtained from the metric usage.

1 Introduction

There are today many applications for pedestrian detection in well-established applications such as video surveillance and relatively newer ones such as self-driven vehicles. Thus, the problem of pedestrian detection is relatively well-studied in the area of computer vision, but where it is still possible to seek improvements especially for cluttered conditions where incorrect multiple detections around each pedestrians can have a significant negative effect.

In the sliding window pedestrian detection approach, the movement of the window around the pedestrian neighbourhood may result in many inaccurate classifications, so that an additional cleaning up process such as a non-maximal suppression (NMS) algorithm is needed. In this work we study how prone to these problems are different descriptor-classifier combinations and also how the problem can be reduced using a different detector confidence measure. To do this we propose a metric called Weighted Average Differences (WAD). This metric calculates the difference between an ideal model and the actual result of the classification. After that, each detector response is weighted according to how far it is from the actual position of the pedestrian. To test the validity of the metric we analyse its behaviour, measuring the results of two descriptor-classifier combinations in a pedestrian neighbourhood classification task. The data for doing this was taken from the INRIA person dataset as this dataset has been used by many related works. Finally, to validate the result we carry out an ANOVA analysis, so we can test if one descriptor-classifier is better than another in terms of spatial sensitivity.

2 Metric

Spatial sensitivity is understood as the variation of the output of a classifier when a detection window slides close to the object to be detected. Considering this, the proposed metric is based on three basic elements. First, we define an expected outcome model to measure the distance between this and the actual result. Second, the actual result of the classification in the neighbourhood, i.e. the output of each classification is collected in a matrix that represents a confidence map of the classification in a given neighbourhood. Third, a weighting function is applied. This function increases the values far away from the actual position of the pedestrian in the image i.e. a lower WAD value indicates better spatial sensitivity.

The outcome model used in this experiment was an extension of the Gaussian probability density function i.e. normal distribution, to a bi-dimensional space. This model looks like a peak in the middle between zero to one, so it is suitable to represent a well-tuned (with high spatial sensitivity) pedestrian detection.

The result of pedestrian classification is explained ahead, but it is important to realise that this result must have the same sampling density to model i.e. if the classification slide was made pixel by pixel the model has to have the same number of points as the image pixels.

The weighting function seen as a three-dimensional image, has the form of a cone. This means that the value of the function increases as we move away from the centre and this behaviour allows us to give greater importance to the farthest points. Hence a far point with a high value indicates a problem with the sensitivity of the classification.

To understand how the metric works it is easier to think of a one dimension problem. If the sliding classification window slides only horizontally we can have an outcome like that shown in Figure 1. We calculate the absolute value of the difference between the actual classification outcome and the ideal outcome model. This value is weighted using the weighting function and then all results are summed up. The actual pro-

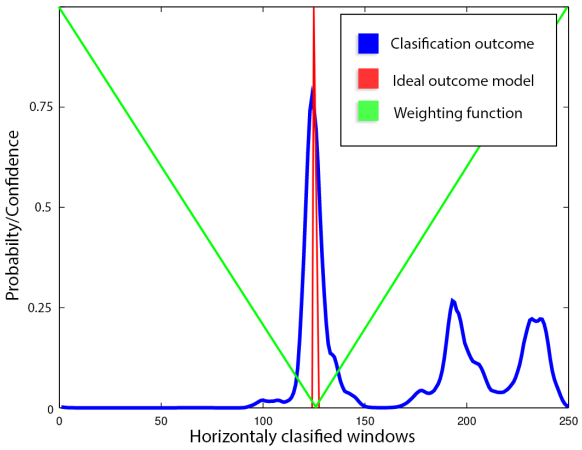

Figure 1. One dimensional weighted average differences graph to illustrate how the metric works (the metric works in the 2-dimensional image classification space).

cess is the same explained before but in three dimensions, so the only real difference is computation time.

3 Dataset

The dataset used in this work is the INRIA person dataset with some modifications. The original dataset contains three groups of files, two of images for training and test (positive and negatives) and one group of plain text files with annotations of the pedestrians position in the images. The training positive set, originally with 64x128px, was scaled to four different sizes. The negative training dataset was selected randomly from the negative images and scaled to the same four sizes of the positive set. The same treatment was applied to the test set. With these modified images we can study the behaviour of the pedestrian detector for different sizes. Studied sizes have the same scale ratio of 1:2 based on the original size 64x128px.

For the classification task we use the images in the test part of the dataset, but the area of analysis was limited to a close neighbourhood of the pedestrian. We use the annotations provided with the dataset to locate the pedestrian and then select the area of analysis. This area contains the pedestrian in the middle and a margin with the same size of the pedestrian around him. This way we can see the behaviour of the detector in a very near neighbourhood. In general, according to the results, this area is still rather wide and maybe for other analysis we can use only 50% or 30% of the pedestrian size.

4 Pedestrian detector

A typical pedestrian detection algorithm has three main elements: descriptor, classifier and NMS. In this work we select two combinations of descriptor and classifier: HOG plus SVM and HOG plus AdaBoost. We suppress the NMS to study the outcome directly form the output of classifier. In this way we can be sure to analyse the behaviour of the classifier and not the result of an NMS algorithm. This also opens the possibility of finding an alternative method to NMS.

The reason for selecting this descriptor-classifier combinations can be found in Dalal[2] work, in fact much of this work is based on Dalal[2, 3] work on pedestrian detection problem, as is much of the literature on this subject.

The model training step was made using the same parameters used by Dalal for the SVM based detector [2]. In the case of the AdaBoost, the training parameters were obtained from the literature [12, 6, 5]. Thus, we have used the same descriptors, dataset, classifiers and parameters as reported by reference works so as to allow direct comparisons.

5 Evaluation Process

In the evaluation process we use the information contained in the dataset annotations as ground truth to select a neighbourhood area. This area is exactly nine times the pedestrian size with the pedestrian in the middle. The neighbourhood area is not always free of other pedestrians occluded by or side by side with the pedestrian in the centre.

Each pedestrian detector performed the same task. This consists on classifying each neighbourhood area obtained from the dataset. At this point a normalisation process is needed because the outcome of an SVM has a distinct scale than the outcome of AdaBoost. To normalise the outcome of the classifiers into a confidence map, the Platt Sigmoid [9] was used. This technique consists in training two constants of a sigmoid function and use them to map the actual outcome to a normalised one into a probability/confidence scale. We note here that many of the works on pedestrian detection that use SVM use the output of the SVM directly rather than converting it into a probability. We found that this generates more false detections and hence complicates the final non-maximal suppression stage. Finally, each confidence map was measured using WAD.

The implementation of the Platt's Algorithm to train the aforementioned constants is known to have some numerical difficulties described in Lin [7] work. To avoid these difficulties we use the algorithm proposed by Lin instead of the original one by Platt.

From the evaluation process we obtain two kinds of results: confidence maps and WADs measures. Confidence maps are useful for a better understanding of the problem. Its difficult to see only the numbers and understand whats happen on each classification, so this maps are useful tools to provide us with a qualitative analysis. WADs measures allow easier calculations to compare results. This results are the opposite of the confidence maps and provide us a mathematical and comparative analysis. Both kind of results are complementary and help us in different ways to problem better understanding.

6 Results

Confidence maps can be represented graphically (Figure 2) for each neighbourhood classification. Also the average confidence map (over the whole set of images) for each detector-scale combination can be computed. These maps give visual

information about the spatial sensitivity and the expected values of WAD. To obtain a high quality confidence map, as it mentioned above, we use dense (pixel to pixel) sampling in the classification task. The map contains the result values of each classification for a given neighbourhood.

The spatial sensitivity metric was developed to be applied in pedestrian detection problems, more specifically in the neighbourhood of one pedestrian detection i.e. given a pedestrian in an image, we can use WAD to measure which detector improves the accuracy of the pedestrian detection for this specific pedestrian. Then we need an ANOVA analysis to generalise the results of the WAD measure.

ANOVA analysis allows testing if the average of the values obtained from the neighbourhood classification process may have statistically significant differences. The samples obtained from the classification task give us eight different groups according to the scale and person detector used. The graph resulting of the ANOVA (Figure 3) show the average WAD value for the eight groups. Four of the values are for the SVM based combinations and the other four are for the AdaBoost based combinations.

From these results we can see that the average WAD measure of the SVM based detector have a clear tendency to grow up when the window size increases. The AdaBoost based detector has resulted in greater uncertainty.

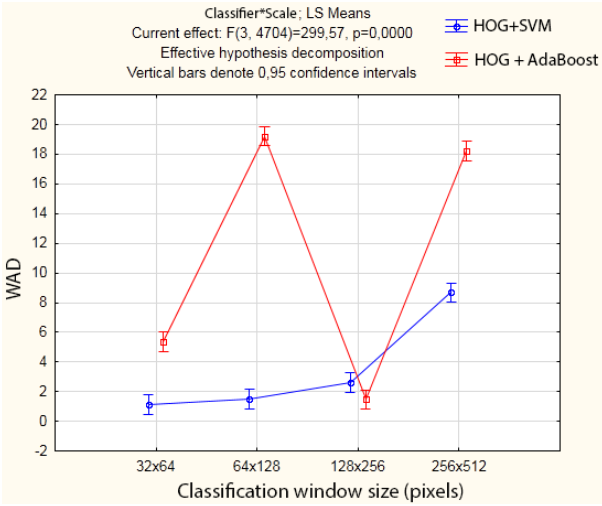

Figure 3. Results of ANOVA. The results of the analysis of variance indicates statistically significant differences between mean WAD values.

7 Discussion

The design of the metric was created thinking on pedestrian detection, but the structure with some adjustments allows it to be used in other object detection problems. In fact it is important to do that in order to generalise the metric and validate its usage.

The results obtained for the SVM based combination is consistent in scale growing, but AdaBoost shows some irregularity. The irregularity may be mostly due to the classifier training process, but the selected values are consistent with what the literature reports [2, 12, 6, 5]. To study this in more detail it may be necessary to implement our own hard training algorithm to obtain optimal values of each classifier for each instance of the assessment task.

Descriptor elements in the detector selection have not been deeply analysed, as we have restricted the experiments to the well-known HOG descriptor so that readers can make comparisons with the standard literature. This point may introduce some bias in the results of this study, but this has been done for two main reasons. First, the study of new image descriptors is somewhat outside the scope of this study and, as indicated earlier, HOG has proven popular in the pedestrian detection field [12, 10, 11, 4]. Secondly, this work is a first approach to full spatial sensitivity characterisation, so it is important to keep things simple to see clearly the effect of each element in the spatial sensitivity.

If we can prove that the metric actually works as designed on a more general group of descriptors and classifiers, ANOVA might not be necessary every time. Then it is important to study the behaviour of the WAD metric in a more general study as mentioned above.

Platt's method to obtain probabilities from classification is not a new method and is not the most effective method today [1]. However, Platt's method (as modified by Lin [7]) is robust enough to be applied to another classifier algorithms and not only on SVMs [8].

8 Conclusions

One of the main contributions of this work is the proposed WAD metric, as it is very useful to analyse spatial sensitivity which in our opinion is an understudied problem in the pedestrian detection problem. The analysis of variance works as validation for the WAD metric in inferential statistical terms. So WAD metric measures spatial sensitivity and simplifies the problem of comparing confidence maps. Spatial sensitivity is a concept important to study because it allows observing from a new perspective some elements of accuracy in the problem of pedestrian detection. Refining these elements should result in better tuned pedestrian detection algorithms so that theses algorithms can have better performance in problems related to the neighbourhood such as counting and tracking in the presence of occlusion.

Another contribution is the comparison, in terms of spatial sensitivity, between the analysed pedestrian detectors. The results obtained from the ANOVA analysis indicate a clear advantage of the SVM based detector over that based on AdaBoost. This result is the same for three of four of the analysed cases, so in general we can say that SVM is better in terms of spatial sensitivity. As future work it is important to analyse other classification algorithms to investigate the generality of the WAD metric.

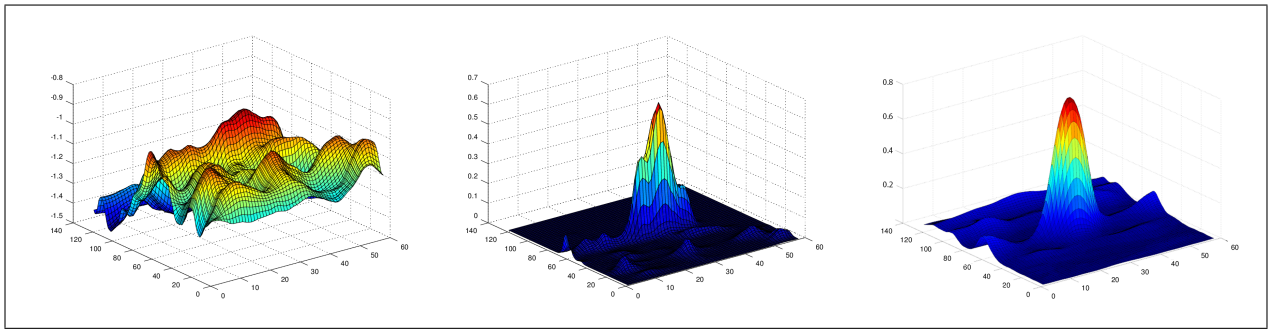

Figure 2. Confidence maps for three different steps of the process. Left: raw classification (not a true confidence map, but important to see the contrast between the raw output of a classifier such as SVM and a normalised probability output). Middle: normalised classification i.e. an actual confidence map. Right: average sample classification. With these confidence maps we can study the results of the classification in a comparative way.

9 Acknowledgement

The work described here was carried out as part of the OBSERVE project funded by the Fondecyt Regular programme of Conicyt (Chilean Research Council for Science and Technology) under grant no. 1140209. Sergio A Velastin has received funding from the Universidad Carlos III de Madrid, the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 600371, el Ministerio de Economa y Competitividad (COFUND2013-51509) el Ministerio de Educacin, cultura y Deporte (CEI-15-17) and Banco Santander.

References

[1] Zezhi Chen. *Detection, tracking and classification of vehicles in urban environments*. PhD thesis, Kingston University, 2012.

[2] Navneet Dalal. *Finding people in images and videos*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2006.

[3] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[4] Oscar Déniz, Gloria Bueno, Jesús Salido, and Fernando De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, 2011.

[5] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407, 2000.

[6] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[7] Hsuan-Tien Lin, Chih-Jen Lin, and Ruby C Weng. A note on platts probabilistic outputs for support vector machines. *Machine learning*, 68(3):267–276, 2007.

[8] Alexandru Niculescu-Mizil and Rich Caruana. Obtaining calibrated probabilities from boosting. In *UAI*, page 413, 2005.

[9] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[10] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *Intelligent Vehicles Symposium, 2006 IEEE*, pages 206–212, 2006.

[11] Xiaodong Yang, Chenyang Zhang, and YingLi Tian. Recognizing actions using depth motion maps-based histograms of oriented gradients. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1057–1060. ACM, 2012.

[12] Qiang Zhu, M-C Yeh, Kwang-Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1491–1498. IEEE, 2006.