PERSON CLASSIFICATION IN IMAGES: AN UNBIASED ANALYSIS FROM MULTIPLE POSES AND SITUATIONS

Alexandre Franco*, Raul Alves*, Luciano Oliveira*

*Intelligent Vision Rearch Lab
Federal University of Bahia
Salvador, Brazil

Emails: acsfranco@ufba.br, ralves@ufba.br, lrebouca@ufba.br

Abstract— Person classification is one of the most important study topics in the field of image pattern recognition. Over the past decades, novel methods have been evolved, and object features and classifiers created. Applications such as person detection and tracking, in intelligent transportation systems or video surveillance, benefit from person classification for real-life applications. Nevertheless, for that systems to be employed there is a need of assessing their performance to assure that will be effective in practice. From plots of classification performance to real-life applications, there seems to be a gap not yet solved, since a near perfect performance curve is not a guarantee of a flawless detection system. In this paper, we present a thorough study toward comprehending why person classifiers are so perfect in plots but not yet completely successful in practice. For that, several features (histogram of oriented gradients (HOG), pyramid HOG, local binary pattern, local phase quantization and Haar-like), two of the most applied classifiers (support vector machine and adaptive boosting) are analyzed over the 2012 person classification Pascal VOC dataset with 27647 cropped images, grouped into 8 person poses and situations. By relying on receiver operating characteristic and precision-recall tools, it was observed that person classification, in several poses and situations, demonstrated to have two different dominant performances, or even different variances among those two performance tools. One main conclusion drawn from the present study was that there is an inherent biased analysis, while assessing a novel proposed method performance. Important guesses are given in the direction of explaining why most of classification performance analyses is somewhat biased.

Keywords— computer vision and pattern recognition, ROC curve, precision-recall curve, person classification performance.

Resumo— Classificação de pessoas é um dos tópicos de estudo mais importante na área de reconhecimento de padrões de imagem. Nas últimas décadas, novos métodos têm evoluído e descritores de objetos e classificadores criados. Aplicações, tais como detecção e rastreamento de pessoas, nos sistemas de transporte inteligente ou de vigilância de vídeo são beneficiados pela classificação de pessoas em aplicações da vida real. No entanto, para que os sistemas sejam utilizados há uma necessidade de avaliar o seu desempenho para assegurar que será eficaz na prática. Dos gráficos de desempenho da classificação para as aplicações da vida real, parece haver uma lacuna ainda não resolvida, uma vez que a curva de desempenho quase perfeita não é uma garantia de um sistema de detecção sem falhas. Neste artigo, apresentamos um estudo aprofundado para compreender por que os classificadores de pessoas são tão perfeitos em gráficos, mas ainda não completamente bem sucedido na prática. Para isso, diversos descritores (histograma de gradientes orientados (HOG), pirâmide HOG, padrão binário local(LBP), quantização de fase local(LPQ) e Haar-like), dois dos classificadores mais aplicados (máquinas de vetores de suporte e *adaptive boosting*) são analisados sobre o dataset de classificação de pessoas Pascal VOC 2012, com 27.647 imagens cortadas, agrupadas em pessoas com 8 poses e situações. Utilizando as ferramentas de avaliação de desempenho *receiver operating characteristic* e *precision-recall*, observou-se que a classificação de pessoas, em diversas poses e situações, demonstrou ter dois diferentes domínio, ou mesmo diferentes variações de desempenho entre essas duas ferramentas. Uma das principais conclusões extraídas do presente estudo foi que existe uma análise tendenciosa inerente, ao avaliar o desempenho de um novo método proposto. Suposições importantes são dadas no sentido de explicar por que a maioria das análises de desempenho de classificação é um tanto tendenciosa.

Palavras-chave— visão computacional e reconhecimento de padrões, curva ROC, curva precision-recall, desempenho de classificação de pessoas.

1 Introduction

Several studies have been done in the field of image pattern recognition, having person classification as one of the most important topic in this field. Person classification is a key component for person detection (so-called pedestrian detection in intelligent transportation systems - ITS) (Mogelmose et al., 2012; Broggi et al., 2009) and tracking, in ITS or video surveillance systems (Zhang et al., 2007; Bischof, 2008). In this sense, by having a high performance classification system, those aforementioned systems can be more suitable to be employed in practice. To improve the classification performance, proposed methods have been conceived by a variety of features and classifiers, in the last few decades. In order to broadly assess the performance of the proposed methods, several image datasets have been gathered, and competitions created to challenge the methods. However, evaluation results usually privilege some characteristics or properties of the datasets, which make the result transfer, most of the time, unreliable toward application of person classification in real-life systems.

The performance of a classification in a pattern recognition problem depends on the choice of appropriate classifiers and suitable features to appropriately represent objects. The goal of this representation is to encompass distinctiveness, uniqueness or rarity of each object class. To reach that goal, much has been attempting toward conceiving feature spaces which can aid the classifiers to separate objects from non-objects, in a binary fashion. In this sense, many researches can be found in the literature: Papageorgiou and Poggio (2000) proposed the Haar-like features, which are a type of square wavelets that capture changes of contrast in objects of the image on overlapped image regions. The original proposed used the support vector machine (SVM) to classify this feature vector. After,Viola and Jones (2001) proposed an adaboost classifier to improve the performance of face recognition. Dalal and Triggs (2005) proposed a different feature space based on a dense representation defined by histograms of oriented gradient (HOG). An object feature based on image pyramid representation and HOG descriptor, called pyramid HOG (PHOG) was proposed by Bosch et al. (2007) to encode the local shape of the object by capturing its spacial distribution of edges and representing as a vector descriptor. In the context of texture descriptors, the local binary pattern (LBP) was suggested by Ojala et al. (1994) and the local phase quantization (LPQ) was proposed by Ojansivu and Heikkilä (2008).

Evaluation of the proposed methods and systems usually involves assessing system performance by receiver operating characteristic (ROC), precision-recall (PR) or detection error trade-off curves, over public datasets, in order to pinpointing system performance comparatively. Nevertheless, Torralba and Efros (2011) have showed, by cross-comparison of training and testing datasets, that this evaluating methodology can lead to a biased look at the classification performance, since it can fail to infer the classifier generalization throughout many different datasets. Although some works as in Drummond and Holte (2004) and Davis and Goadrich (2006) already pointed to this bias in the way of evaluating the classification performance, proposing other tools for that job, none of them has explored this issue throughout an extensive investigation of the results of a particular classification problem as was done here.

In order to investigate those issues, the contribution of this paper is to present a thorough performance analysis of person classification, in different poses and situations, investigating the bias and the possible reason for that. To accomplish that, two of the most used classifiers (support vector machine – SVM, and adaptive boosting – adaboost) and some state-of-the-art object representations (HOG, PHOG, LBP, LPQ and Haar-like) are used in order to build several possible classification systems. Those systems are analysed by two of the most commonly used tools of performance evaluation – the receiver operating characteristic (ROC) and the precision-recall (PR) curves. The 2012 person classification Pascal Visual Object Classes (Pascal VOC) dataset (Everingham et al., 2010), comprised of 27647 cropped images, grouped into 8 person poses and situations, was used to allow a thorough analysis of person classification. That chosen dataset is a challenging one, being widely used in international classification competitions. We could noticed that person classification, in several poses and situations, demonstrated to have two different dominant performances, or even different variances among the ROC and PR tools in some aspects that can be explained by the characteristics of those performance tools. Moreover, there is a strong relation between the two performance tools that makes a single analysis in one of them to be unreliable in certain way.

The remainder of this text is organized as follow: In section 2, ROC and PR curves, their relations, as well as their biases are addressed. Section 3 shows an experimental performance analysis of image person classifications, discussing the results. Finally, Section 4 draws conclusions.

2 Evaluating classification performance

There are several methods of measuring classification performance. Graphical methods are usually more useful, showing at a glance the classification performance pattern. There are two graphical methods widely used for this purpose, which are based on the evaluation of the trade-off between two parameters of performance rate. One of them is the ROC curve and the other one is the PR curve. There are important assumptions to be made about these curves as well their advantages, limitations and biases. The area under curve (AUC) is one of the most single metric used in order to evaluate these methods. The remainder of this section shows an overview of these methods, their characteristics and the relation between them.

2.1 ROC curve

In a binary classification problem there are two class labels. There are four categories of measures over the class labels, as follow: True positives (TP) are the examples of correct true labels classification, true negatives (TN) are the examples of correct negative labels classification, false positives (FP) are related to the negative examples incorrectly classified as positive, and, finally, false negatives (FN) are related to the positive examples incorrectly classified as negative.

ROC curve is a graph plot that represents the classification performance in a binary fashion. The main idea of this curve is to illustrate the behavior of the true positive rate (TPR) when in-

Table 1: Performance considering each feature classified by an SVM with respect to the ROC (a) and PR curves (b).

(a) AUC results of SVM classification over ROC.

Classification performance ranking				
	Object person situations			
Ranking	**NO**	**O**	**NT**	**T**
1 st	0.806(H)	0.812(L)	0.891(H)	0.799(H)
2 nd	0.788(L)	0.806(H)	0.867(L)	0.779(L)
3 rd	0.760(P)	0.748(B)	0.844(P)	0.756(B)
4 th	0.733(B)	0.742(P)	0.772(B)	0.736(P)
5 th	0.539(A)	0.515(A)	0.506(A)	0.575(A)
	Object person poses			
Ranking	**FR**	**RE**	**LE**	**RI**
1 st	0.799(H)	0.835(H)	0.848(H)	0.855(H)
2 nd	0.796(L)	0.813(P)	0.763(P)	0.771(P)
3 rd	0.754(B)	0.757(L)	0.685(B)	0.727(B)
4 th	0.741(P)	0.712(B)	0.630(L)	0.714(L)
5 th	0.518(A)	0.636(A)	0.544(A)	0.492(A)

AUC (Feature)

(b) AUC results of SVM classification over PR.

Classification performance ranking				
	Object person situations			
Ranking	**NO**	**O**	**NT**	**T**
1 st	0.603(H)	0.573(H)	0.731(H)	0.615(H)
2 nd	0.494(P)	0.517(L)	0.556(P)	0.550(B)
3 rd	0.476(L)	0.459(B)	0.517(L)	0.550(L)
4 th	0.436(B)	0.421(P)	0.426(B)	0.482(P)
5 th	0.244(A)	0.225(A)	0.177(A)	0.303(A)
	Object person poses			
Ranking	**FR**	**RE**	**LE**	**RI**
1 st	0.627(H)	0.476(H)	0.419(H)	0.443(H)
2 nd	0.559(L)	0.323(P)	0.226(P)	0.244(P)
3 rd	0.535(B)	0.170(B)	0.143(B)	0.187(B)
4 th	0.492(P)	0.144(L)	0.112(L)	0.173(L)
5 th	0.279(A)	0.084(A)	0.086(A)	0.071(A)

AUC (Feature)

Table 2: Performance considering each feature classified by an adaboost with respect to ROC (a) and PR (b) curves.

(a) AUC results of adaboost classification over ROC.

Classification performance ranking				
	Object person situations			
Ranking	**NO**	**O**	**NT**	**T**
1 st	0.776(L)	0.787(H)	0.867(H)	0.778(L)
2 nd	0.762(P)	0.754(P)	0.849(P)	0.772(H)
3 rd	0.735(H)	0.699(A)	0.814(L)	0.741(P)
4 th	0.687(A)	0.681(B)	0.738(A)	0.716(A)
5 th	0.680(B)	0.630(L)	0.587(B)	0.690(B)
	Object person poses			
Ranking	**FR**	**RE**	**LE**	**RI**
1 st	0.779(H)	0.824(H)	0.807(H)	0.827(H)
2 nd	0.741(P)	0.818(P)	0.796(L)	0.811(L)
3 rd	0.685(B)	0.815(L)	0.782(P)	0.803(P)
4 th	0.663(A)	0.737(A)	0.690(A)	0.668(A)
5 th	0.552(L)	0.646(B)	0.630(B)	0.662(B)

AUC (Feature)

(b) AUC results of adaboost classification over PR.

Classification performance ranking				
	Object person situations			
Ranking	**NO**	**O**	**NT**	**T**
1 st	0.475(P)	0.510(L)	0.631(H)	0.542(L)
2 nd	0.459(L)	0.505(H)	0.568(P)	0.528(H)
3 rd	0.427(H)	0.447(P)	0.446(L)	0.491(P)
4 th	0.367(A)	0.364(B)	0.369(A)	0.439(A)
5 th	0.345(B)	0.363(A)	0.363(B)	0.426(B)
	Object person poses			
Ranking	**FR**	**RE**	**LE**	**RI**
1 st	0.555(H)	0.315(H)	0.326(H)	0.352(H)
2 nd	0.495(P)	0.282(P)	0.255(P)	0.290(P)
3 rd	0.417(B)	0.211(A)	0.232(L)	0.275(L)
4 th	0.377(A)	0.190(L)	0.158(A)	0.138(A)
5 th	0.245(L)	0.092(B)	0.113(B)	0.138(B)

AUC (Feature)

creasing the number of false positive rate (FPR). Figure 1 shows examples of ROC curves, in the first columns. The curve closest to the upper left corner is the one of the best performance. The computation of TPR and FPR are given by

$$TPR = \frac{TP}{TP + FN}\,. \tag{1}$$

$$FPR = \frac{FP}{FP + TN}\,. \tag{2}$$

2.2 PR curve

The PR curve represents the trade-off between TPR, called recall, and the hit rate of those objects classified as positive, called precision. The precision is computed according to

$$PRECISION = \frac{TP}{TP + FP}\,. \tag{3}$$

In the second column of Fig. 1, there are examples of PR curves. The best curve is the one closest to the upper right corner in the plot.

2.3 AUC value

Looking at a ROC and a PR, the classification performance of a method is directly proportional to the area under the curve (AUC). The AUC is a single value and the closer to one means a near perfect classification performance.

2.4 Relation between ROC and PR curves

Since the ROC curve is a tool commonly used to present the performance of a binary classification, there is a problem when dealing with datasets that contains a very small number of positive examples. As a result, the ROC tends to be an optimistic curve. Another issue is that sometimes two classification algorithms with quite distinct performance in PR curves have ROC curves very close to each other. The reason for these is that the ROC does not consider the true negatives examples. Therefore, when a classifier returns a high

Table 3: Feature with same performance ranking in both ROC and PR curves, considering SVM. The character "-" means an uncorrelated rank in that situation.

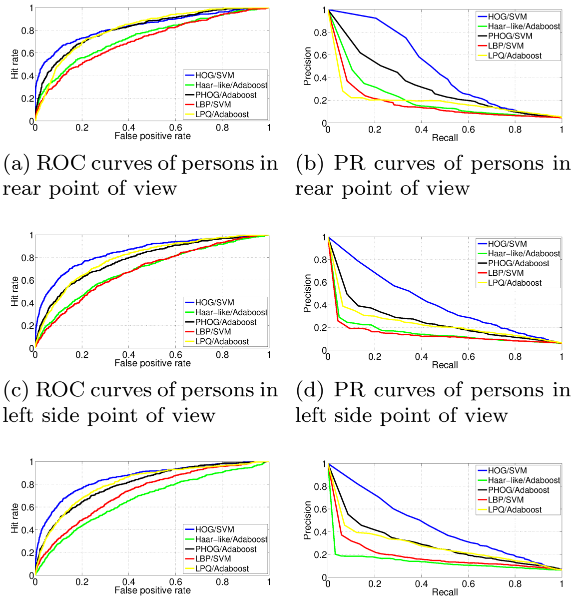
Classification performance ranking					
		Object person situations			
Ranking		NO	O	NT	T
1 st	Features	HOG	-	HOG	HOG
2 nd		-	-	-	-
3 rd		-	LBP	-	-
4 th		LBP	PHOG	LBP	PHOG
5 th		Haar-like	Haar-like	Haar-like	Haar-like
		Object person poses			
Ranking		FR	RE	LE	RI
1 st	Features	HOG	HOG	HOG	HOG
2 nd		LPQ	PHOG	PHOG	PHOG
3 rd		LBP	-	LBP	LBP
4 th		PHOG	-	LPQ	LPQ
5 th		Haar-like	Haar-like	Haar-like	Haar-like

Table 4: Feature with same performance ranking in both ROC and PR curves, considering adaboost. The character "-" means an uncorrelated rank in that situation.

Classification performance ranking					
		Object person situations			
Ranking		NO	O	NT	T
1 st	Features	-	-	HOG	LPQ
2 nd		-	-	PHOG	HOG
3 rd		HOG	-	LPQ	PHOG
4 th		Haar-like	LBP	Haar-like	Haar-like
5 th		LBP	-	LBP	LBP
		Object person poses			
Ranking		FR	RE	LE	RI
1 st	Features	HOG	HOG	HOG	HOG
2 nd		PHOG	PHOG	-	-
3 rd		LBP	-	-	-
4 th		Haar-like	-	Haar-like	Haar-like
5 th		LPQ	LBP	LBP	LBP

number of true positives, even that the number of true negatives approaches zero, the ROC misleadingly points to a high performance result of the classifier. The lack of true negative rate in the ROC curve also explains the low distance between the two performance curves, even if the difference of performance between them is high. This happens when the difference between true positive of the two curves is low and the difference between true negative is high, for instance.

Davis and Goadrich (2006) show that there is a deep correlation between PR and ROC evaluation tools, proving that one curve is dominant if, and only if, it is dominant in both PR and ROC curves. This leads one to observe that conclusions should not be drawn simply by analysing the performance on the point of view of only one performance analysis tool.



(a) ROC curves of persons in rear point of view (b) PR curves of persons in rear point of view

(c) ROC curves of persons in left side point of view (d) PR curves of persons in left side point of view

(e) ROC curves of persons in right side point of view (f) PR curves of persons in right side point of view

Figure 1: Some evaluations of person classification with ROC and PR that show the ROC biases.

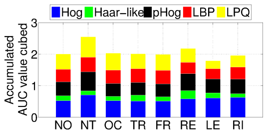
3 Experimental analysis

Toward the observation of classification performance in a more accurate way, we accomplished a thorough performance analysis by considering 5 feature extractors (HOG, PHOG, LBP, LPQ and Haar-like), 2 classifiers (SVM and adaboost), 4 situations (occluded, non-occluded, truncated and non-truncated) and 4 poses (frontal, rear, left and right). All that in the perspective of the two aforementioned analysed tools (ROC and PR) and their behaviors. The classifiers and features were implemented in Matlab and the tests were performed on a computer with 8Gb of RAM and 2.4 GHz processor. Some hints of which person classification system provides the best real performance in all poses and situations are also given.

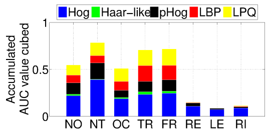
3.1 *Methodology*

First the AUC values were computed for all combinations of feature/classifier/performance tool ($5 \times 2 \times 2$), and the results were ranked according to the best AUCs (see Tables 1 and 2). Fig. 2 graphically summarizes the information contained in Tables 1 and 2. Bars in Fig. 2 depicts the values of AUC in PR and ROC curves. Next, a correlation analysis among AUCs of SVM and adaboost were made (see Tables 3 and 4).

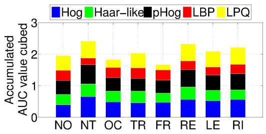
Some ROCs and PRs curves were plotted in the Fig. 1. Theses plots shows the ROC biases in some persons poses. Finally, the worst and the best pair of feature/classifier regardless of situation and pose are showed in Fig. 3.



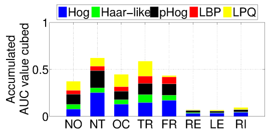
(a) Accumulated AUC values of the ROC curves considering SVM classifier



(b) Accumulated AUC values of the PR curves considering SVM classifier



(c) Accumulated AUC values of the ROC curves considering adaboost classifier



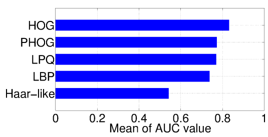
(d) Accumulated AUC values of the PR curves considering adaboost classifier

Figure 2: Accumulated AUC values of ROC and PR curves considering adaboost and SVM classifiers. Each bar patch is associated to a feature. Its length is proportional to the AUC value, which is raised by a power of three, which was done in order to increase the differences between two AUC values, graphically improving the visualization of the values.

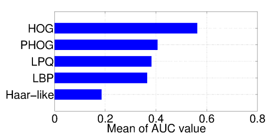
3.2 *Discussion of the results*

By investigating the figures and the values of the tables presented in the previous section, one can be taken some discussions and conclusions of the results as will be seen in the remind of this section.

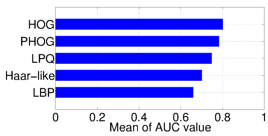
The best and the worst feature in each analysis tool Considering the values summarized in Tables 1 and 2, and Fig. 2, it is noticeable that the HOG feature presented the best performance in almost all instances, whereas Haar-like feature had the worst performance, being classified by an SVM. Yet, the LPB feature presented the worst performance in almost all situations and poses with Adaboost classifier.



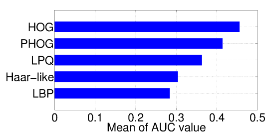
(a) Mean of AUC values of the ROC curves considering SVM classifier.



(b) Mean of AUC values of the PR curves considering SVM classifier.



(c) Mean of AUC values of the ROC curves considering adaboost classifier.



(d) Mean of AUC values of the PR curves considering adaboost classifier.

Figure 3: Mean of AUC values of the ROC and the PR curves of all person situations and poses, considering all pairs of feature/classifier.

In Fig. 2, all bars in the column of the non-truncated (NT) situation show that all features together in that bar present the best performance either in ROC or PR curve.Conversely, all bars in the column of the left pose (LE) show that all those features together present the worst performance.

According to Fig. 3 and analyzing the performance of person classification regardless the poses and situations, HOG classified by an SVM or Adaboost has presented the best performance in all plots, while the Haar-like feature, classified by an SVM, and LBP, classified by an Adaboost, are both the worst in classification performance.

Performance ranking As previously presented in Davis and Goadrich (2006), one curve dominates in ROC space if and only if it dominates in PR space. This means that the performance raking of a particular classifier/feature pair is only reliable when this performance ranking is the same in both ROC and PR space. Tables 3 and 4 summarizes the results where the pairs of feature/classifier are in the same performance ranking, in both ROC and PR, according to Tables 1 and 2. The position in the Tables 3 and 4 filled with the character "-" means that, in that particular ranking, the result of the ROC curve does not match with the PR curve. These results select reliably the most and the least appropriated pair in each particular situation/pose.

Addressing the ROC biases In Fig. 1, differences in performance between ROC and PR, as well optimistic results and low variance between ROC curves, can be seen in some situations. For example, in Fig. 1c, the ROC curve of HOG presents a better performance than the ROC curve of the same feature in Fig. 1a. Conversely, the response of PR curve in Fig. 1d and 1b presents the contrary. The same observation can be seen in Figs 1e and 1a, and Figs 1f and 1b. The ROC curves in Fig. 1a are close to each other, whilst, in Fig. 1b, the PR curves are far apart, for the same situations. These differences, previously presented, are related to person classifications in rear, right and left poses. These poses are those that have the smallest number of positive classes, which are approximately 500 positive examples for each pose, against 8777 negative examples.

Looking at Fig. 2, it can be seen, in the PR space, the high variance between AUC values of each person situation or pose with all features together (noted by the difference among the height of the bars). In addition, the optimistic value in the poses rear, left and right is noticeable by the high height of the respective bars columns in the ROC space.

The same ROC biases previously mentioned can be seen in the Figure 3 that shows the performance of each classifier/feature regardless the person situation and poses. For instance, the variance of the AUC value is higher in PR curves (noted by the difference among the width of the bars) than in ROC curves. Yet, the bars also show an optimist values in the ROC curves of Haar-like feature classified by an SVM.

Discussion of the results Solely considering a pair of feature/classifier, it is insufficient to guarantee a reliable classification in all poses and situations. Even considering that HOG presented, most of the time, the best performance, the classification results were yet bellow of the average in order to apply this type of feature (either with SVM or Adaboost), in a real-life application.

4 Conclusion

An extensive analysis of person classification problem was done in this work, with the aim of showing that a single plot of a unique performance tool can lead to a biased results. This is usually due to the characteristics of the datasets. The results of the ROC, in some cases, was different than those presented in the PR, which showed, in practice, the two problems of ROC curves - optimism in results and low variance. These biases were most evident on rear, left and right poses of the object where the number of positive images examples was much smaller than the number of negative examples. In that case, the PR curves presented more realistic results. The current analysis also gave some hints about which features and classifiers are more or less appropriate to the problem of person classification, showing directions on the way to tackle the problem in the future.

Acknowledgment

This work was supported by Fundação de Amparo a Pesquisa do Estado da Bahia (FAPESB), under grant 6858/2011.

References

Bischof, H. (2008). Robust person detection for surveillance using online learning, *Image Analysis for Multimedia Interactive Services, International Workshop on*, pp. 1–1.

Bosch, A., Zisserman, A. and Munoz, X. (2007). Representing shape with a spatial pyramid kernel, *Proceedings of the international conference on Image and video retrieval*, ACM, New York, pp. 401–408.

Broggi, A., Cerri, P., Ghidoni, S., Grisleri, P. and Jung, H. G. (2009). A new approach to urban pedestrian detection for automatic braking, *Intelligent Transportation Systems, IEEE Transactions on* **10**(4): 594–605.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, IEEE Conference on*, Vol. 1, pp. 886–893.

Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves, *Proceedings of the international conference on Machine learning*, ACM, New York, pp. 233–240.

Drummond, C. and Holte, R. C. (2004). What roc curves can't do (and cost curves can, *In ROCAI-2004*.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. (2010). The pascal visual object classes (voc) challenge, *International Journal of Computer Vision* **88**(2): 303–338.

Mogelmose, A., Prioletti, A., Trivedi, M., Broggi, A. and Moeslund, T. (2012). Two-stage part-based pedestrian detection, *Intelligent Transportation Systems, International IEEE Conference on*, pp. 73–77.

Ojala, T., Pietikainen, M. and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions, *Pattern Recognition, Proceedings of the International Conference on*, Vol. 1, pp. 582–585.

Ojansivu, V. and Heikkilä, J. (2008). Blur insensitive texture classification using local phase quantization, *in* A. Elmoataz, O. Lezoray, F. Nouboud and D. Mammass (eds), *Image and Signal Processing*, Vol. 5099 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 236–243.

Papageorgiou, C. and Poggio, T. (2000). A trainable system for object detection, *Int. J. Comput. Vision* **38**(1): 15–33.

Torralba, A. and Efros, A. (2011). Unbiased look at dataset bias, *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1521–1528.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *Computer Vision and Pattern Recognition, Proceedings of the Conference on*, Vol. 1, pp. 511–518.

Zhang, L., Li, S., Yuan, X. and Xiang, S. (2007). Real-time object classification in video surveillance based on appearance learning, *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1–8.