Improving pedestrian detection using Support Vector Regression

Mounir ERRAMI, Mohammed RZIZA
LRIT, Unit Associated to CNRST (URAC 29)
Mohammed V University, Faculty of Sciences, P.O.B. 1014
Rabat, Morocco
mounirerrami@gmail.com, rziza@fsr.ac.ma

***Abstract*—Pedestrian detection has been always a challenging problem in computer vision. Numerous approaches based on features extraction and classification have been proposed over the years. In this paper, we present a novel pedestrian detection approach based on supervised classification. We propose here the use of basic statistical operators to adapt support vector regression (SVR) to binary classification. The classification chain adopted in this work is presented as follows : First, we use Haar wavelet decomposition and Histograms of Oriented Gradients (HOG) for features extraction. For the classification task, we use our proposed method and compare it with KNN and SVM classifiers. Experiments have been done on a public pedestrian data set. The obtained results prove the high performance of our proposed classification approach.**

***Keywords*-Features extraction; classification; SVR; wavelet transform; pedestrian images;**

I. Introduction

With the constant increase in road traffic, the risk of accidents increases too. Each year, more than 1.2 million people around the world die on the road [1]. All statistics of road safety show that each year, nearly 10 million people around the globe are involved in a road accident. Thus, Pedestrian detection is a problem of considerable social, economical and scientific interest.

However, this task proves to be very difficult due to several constraints, namely the acquisition conditions, internal and external noises, pedestrians variability in term of size, type of clothing and position in the image.

Following a comprehensive and deep inspection of previous survey works [2] developed in this context, the classification based approach seems to be the most used. It consists of two essential steps : features extraction and supervised learning (classification). Each image of the pedestrian database will be described using a features extractor (descriptor), then the vector extracted will be given to a classifier for training. The main difference between descriptors concern the nature of extracted information (texture, edge, colors ...). In this context, Papageorgiou and Poggio [3] proposed the use of support vector machines (SVM) as a classifier with an overcomplete dictionary of multiscale Haar. Following this idea, Viola and Jones [4] found a new way to calculate Haar features by using integral images, known as the Haar-like features, with an AdaBoost classifier. Recently, descriptors based on Haar-like features took researchers interest and many works have been proposed in this context [5].

The gradient-based features came also with large gains.

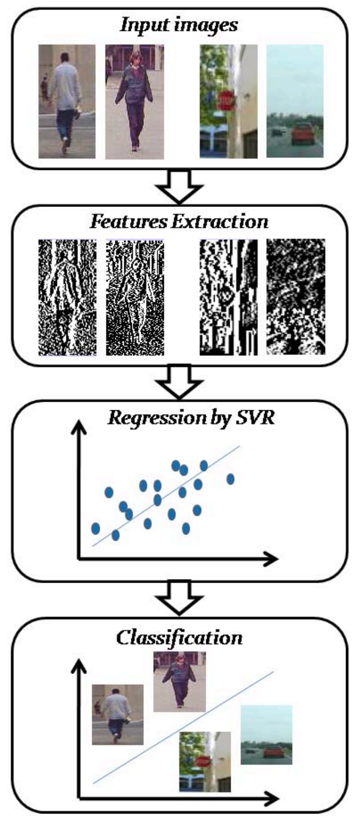

Figure 1. Overall layout of the proposed classification chain.

Dalal and Triggs [6] expose a powerful descriptor based on the calculation of histograms of oriented gradients (HOG). Giving interesting results in different applications areas in pattern recognition, the number of variants of HOG increase greatly in recent works [7] [8]. In order to outperform the HOG descriptor, several authors investigate the way of using features fusion to take in count the complementarity of information. Wojek and Schiele [9] proved how a combination of HOG, Haar-like features and shape features gave higher performances than any single descriptor.

978-1-5090-0811-7/16 $31.00 © 2016 IEEE
DOI 10.1109/CGiV.2016.38

Of concern in this paper, we propose the use of SVR for a binary classification task. The adopted classification chain is illustrated in the Figure 1 .

The remainder of this paper is structured as follows. Section 2 gives a brief preview of Haar wavelet decomposition and HOG. In section 3 we present the concept of support vector regression. Section 4 is dedicated to the experimental results. Finally, we present our conclusion and future works.

II. Features extraction

Features extraction techniques consist of builiding from an initial set of data a derived values (features) which are intented to be informative and lead to a better performance in the learning and generalization steps. In what follows, two features extraction methods are briefly introduced.

A. Histogram of Oriented Gradients (HOG)

Histogram of oriented gradients is now one of the most wild powerful descriptor. The idea behind, is that object shape can be easily and well characterized by the distribution of local intensity gradients. First works on the orientation histograms were done by Freeman et al [10] for hand gesture recognition. Then come Dalal and Triggs [6] and present a deep study on HOG for human detection.

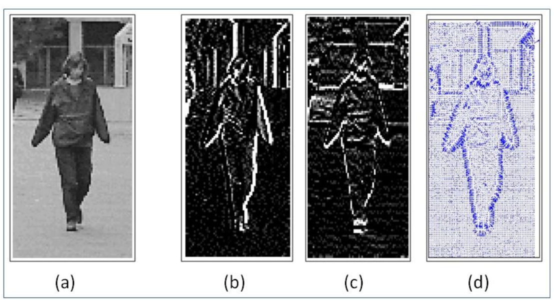

Figure 2. Formation of the HOG descriptor; a) Original image, b) Horizontal gradient computation c) Vertical gradient computation d) Gradients orientations.

In practice, the HOG feature extraction method is done following three essential steps. First, we apply two derivations masks at each point (pixel) of the image :

$$M_1 = [-1, 0, 1] \qquad\qquad M_2 = [-1, 0, 1]^T. \tag{1}$$

then we compute the orientation :

$$\theta = arctan(\frac{G_y}{G_x}) \qquad with \;\; \theta \in [0, \pi]. \tag{2}$$

Finally the image is divided into several cells. For each one, a weighted histogram of oriented gradients is constructed by counting the occurrences of gradient in a bar corresponding to a specific range of orientation.

B. Haar wavelet decomposition

A discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. In general, wavelets are grouped into families, including an initial wavelet ψ called "mother wavelet" and all its images obtained by affine transformations in $\mathbb{R}^n$. it is defined as follows :

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}). \tag{3}$$

With t the variable to study and $\psi_{s,\tau}(t)$ the wavelet dilatation s and translation τ deriving from the mother wavelet.

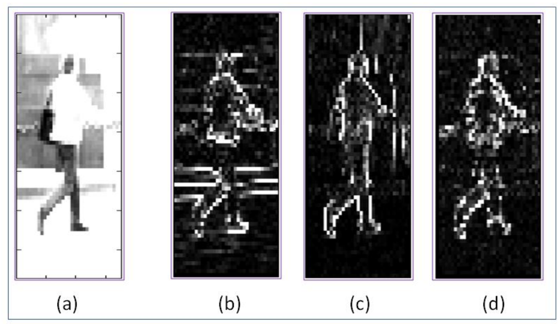

Figure 3. Example of Haar wavelet transform for a pedestrian grayscale image; a) Original image, b) Horizontal coefficients c) Vertical coefficients d) Diagonal coefficients.

Wavelet transform is represented by two low-pass L and high-pass H with several wavelet basis functions. in the case of an image, the wavelet transformation is a multi-scale, a multi-orientation (horizontal, vertical and diagonal) decomposition giving the image signal on different frequency channels (LL), (HL), (LH), (HH) with :

- LL : An approximation of the image which can be decomposed again.
- HL : Horizontal orientation detail coefficients.
- LH : Vertical orientation detail coefficients.
- HH : Diagonal orientation detail coefficients.

Haar wavelet has been introduced by the mathematician Alfred Haar [11]. Nowadays, it is one of the most popular wavelet family used for pattern recognition. it is defined as follows :

$$\psi(x) = \begin{cases} 1 & \text{if} \quad 0 \leq x < \frac{1}{2} \\ -1 & \text{if} \quad \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Using the following Haar scale function :

$$\phi(t) = \begin{cases} 1 & \text{if} \quad 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

the three sub-images obtained from the decomposition (Figure 3) are transformed into vectors, sampled and concatenated to form a single feature vector that it will be given as an input to the classifier.

III. Support Vector Regression (SVR)

Support vector machines are a binary classification method introduced by Vapnik [12]. They are based on tl concept of decision planes (hyperplane) that define decisic boundaries. An hyperplane separates between a set of objec having different class memberships. The training instanc that are closest to this hyperplane are called support vecto In other terms, it is a supervised classification technique th predicts the class of new samples on the basis of a trainir samples already classified. The classification function given as follows :

$$f_{SVM}(x) = \sum_{x_i \in VS} \alpha_i y_i K(x, x_i) + b, \tag{6}$$

where (x_i, y_i) the training set, x the new sample to classify and K is the kernel function defined as follows depending on the nature of the data (linear/non linear case) :

Table I
Kernel functions

Kernel	Function	Parameters
Linear	$K(x,y) = x.y$	$-$
Polynomial	$K(x,y) = (x.y+1)^d, d \in \mathbb{R}$	d
RBF	$K(x,y) = e^{-\gamma \|\|x-y\|\|^2}, \gamma > 0$	γ
Sigmoidal	$K(x,y) = tanh(\gamma x.y + r)$	γ, r

SVMs can also be applied to regression problems by the introduction of an alternative loss function [13]. Generally regression aims to find a function f that fits the relationship between the input x and the output $y = f(x)$. Various loss functions exist [13]. Two loss functions are commonly used in SVR : the "ϵ-SVR" and "ν-SVR". The formulation of linear and non linear regression problem remain the same. As opposed to the SVM, the value of SVR epsilon(ϵ) and nu (ν) parameters controls the sensitivity of the results to the outputs and leads to a better generalization even with large number of training samples.

However, regression techniques are applicable for prediction type of problems as opposed to classification, they gives as an output numeric or continous values. So, the idea here is to adapt regression to classification by changing the continuous ouput values of SVR to categorical output (classes). Statistically, when the studied data present a good distribution (high similarity intra-class, high dissimilarity inter-class), basic statistical operators such as the mean, median, variance or standard deviation can be used to find a midpoint separating data space into two classes.

IV. Experimental results

This section gives comparative experimental results of the techniques described in sections 2 and 3. All experiments have been done on a benchmark database proposed by [14]. The database is divided in three sets, each one contains about 9800 images scaled to common size 18×36 see Figure 4.

Figure 4. Pedestrian (upper row) and non pedestrian (lower row) samples from the benchmark data set.

- Experimental setting :

First, we train our classifiers using 1000, 2000 and 4000 samples. based on classification rate, we decide to keep 4000 samples for training and test. Classification algorithms are evaluated quantitatively by computing three measures :
First the normalized confusion matrix composed by the numbers CMi_1i_2 of elements of the class i_1 which are classified in class i_2. It is defined as follows :

$$NCM_{i_1,i_2} = \frac{CM_{i_1,i_2}}{\sum_{i_k=1}^{N_c} CM_{i_1,i_k}} = \frac{CM_{i_1,i_2}}{N_{i_1}} \tag{7}$$

where N_c is the number of classes and N_{i_1} is the number of elements of class i_1. An average of the classification rate CR_m can be calculated as follows :

$$CR_m = \frac{\sum_{i_k=1}^{N_c} CM_{ii}}{N} \tag{8}$$

where N represents the total number of data used by the classification algorithm. The classification error is also calculated using the Balanced Error Rate (BER) defined as follows :

$$BER = \frac{(\frac{FN}{TP+FN} + \frac{FP}{FP+TN})}{N_c} \tag{9}$$

where TP and FP represent respectively the true positive and false positive, TN and FN the true negative and false negative.
We configure our classifiers as follows :

- we decide emprically the number of neighbors $k = 4$ and $Euclidean$ distance for the KNN classifier.
- For both SVM and SVR, and based on a comparison between kernels using the receiver operating characteristic

curve (ROC) (Figure 5), we choose the linear kernel, considering the fact that it gives good trade off between precision and runtime.

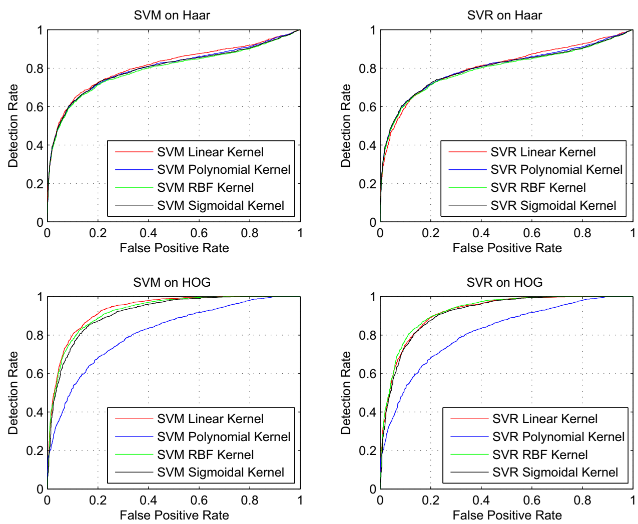

Figure 5. SVM and SVR classification performances on HOG and Haar with different kernels.

The obtained results of normalized confusion matrices (NCM), classification rates CR_m and the BER for SVM, KNN and SVR, with both Haar and HOG descriptors are presented in Table 1. It proves that the proposed SVR approach for classification gives results as good if not better than SVM with HOG descriptor, but outperform it significantly with Haar wavelet.

Conclusion

In this paper, we presented a new classification approach based on support vector regression (SVR). The main topic of this study is to adapt SVR to classification using basic statistical operators. Our SVR classification approach allows more flexibility in the choice of penalties and loss functions. In addition, it gives good generalization performance even with large numbers of training samples. Our experiments have been done in the context of pedestrian detection. We have used a benchmark data set and two powerful descriptors (HOG and Haar). Results showed that the proposed approach achieves better accuracy in comparison to the well-known SVM and KNN classifier. In future work we plan to make a modification of SVR loss functions to further increase the detection accuracy.

Table II
Classification accuracy achieved using the proposed classification approach.

Classifier/Descriptor	Results on DaimlerChrysler database [14]		
	NCM(%)	CR(%)	BER
SVM/HOG	$\begin{pmatrix} 85.15 & 14.85 \\ 15.85 & 84.15 \end{pmatrix}$	84.65	0.15
KNN/HOG	$\begin{pmatrix} 76.90 & 23.10 \\ 9.80 & 90.20 \end{pmatrix}$	83.55	0.16
SVR/HOG	$\begin{pmatrix} 81.75 & 18.25 \\ 11.85 & 88.15 \end{pmatrix}$	**85.00**	**0.14**
SVM/Haar	$\begin{pmatrix} 66.25 & 33.75 \\ 20.30 & 79.70 \end{pmatrix}$	72.97	0.26
KNN/Haar	$\begin{pmatrix} 89.75 & 10.25 \\ 73.85 & 26.15 \end{pmatrix}$	59.00	0.36
SVR/Haar	$\begin{pmatrix} 82.30 & 17.70 \\ 30.15 & 69.85 \end{pmatrix}$	**76.07**	**0.24**

References

[1] M. Peden, *Rapport mondial sur la prévention des traumatismes dus aux accidents de la circulation*. World Health Organization, 2004.

[2] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, 2012.

[3] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[4] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[5] K.-Y. Park and S.-Y. Hwang, "An improved haar-like feature for efficient object detection," *Pattern Recognition Letters*, vol. 42, pp. 148–153, 2014.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[7] L. Weixing, S. Haijun, P. Feng, G. Qi, and Q. Bin, "A fast pedestrian detection via modified hog feature," in *Control Conference (CCC), 2015 34th Chinese*. IEEE, 2015, pp. 3870–3873.

[8] A. Mahmoud, A. El-Barkouky, J. Graham, and A. Farag, "Pedestrian detection using mixed partial derivative based histogram of oriented gradients," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2334–2337.

[9] C. Wojek and B. Schiele, "A performance evaluation of single and multi-feature people detection," in *Pattern Recognition*. Springer, 2008, pp. 82–91.

[10] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *International workshop on automatic face and gesture recognition*, vol. 12, 1995, pp. 296–301.

[11] R. S. Stanković and B. J. Falkowski, "The haar wavelet transform: its status and achievements," *Computers & Electrical Engineering*, vol. 29, no. 1, pp. 25–44, 2003.

[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag New York, USA, 1995.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[14] S. Munder and D. M. Gavrila, "An experimental study on pedestrian classification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 11, pp. 1863–1868, 2006.