

HOL, GDCT AND LDCT FOR PEDESTRIAN DETECTION

Sanaa Tayb¹, Youssef Azdoud¹, Aouatif Amine¹, Bouchra Nassih¹, Hanaa Hachimi¹ and Nabil Hmina¹

¹LGS, National Schools of Applied Sciences,
Ibn Tofail University, Kenitra, Morocco

sanaa.tayb@gmail.com, azdoud.youssef@univ-ibntofail.ac.ma,
amine_aouatif@univ-ibntofail.ac.ma, bouchra.nassih@gmail.com,
hanaa.hachimi@univ-ibntofail.ac.ma, hmina@univ-ibntofail.ac.ma

ABSTRACT

In this paper, we present and analyze different approaches implemented here to resolve pedestrian detection problem. Histograms of Oriented Laplacian (HOL) is a descriptor of characteristic, it aims to highlight objects in digital images, Discrete Cosine Transform DCT with its two version global (GDCT) and local (LDCT), it changes image's pixel into frequencies coefficients and then we use them as a characteristics in the process. We implemented independently these methods and tried to combine it and used there outputs in a classifier, the new generated classifier has proved it efficiency in certain cases. The performance of those methods and their combination is tested on most popular Dataset in pedestrian detection, which are INRIA and Daimler.

KEYWORDS

Pedestrian detection, HOL, DCT local, DCT Global, Classification, SVM, Inria Person, Daimler.

1. INTRODUCTION

In the last years, pedestrian detection and tracking systems have been an important major and a subject of investigation in the vision system studies. Various approaches and methods have been proposed to fulfil the task of detection. It is an application that intend to determine the pedestrian position in an image or a scenery, its importance is noticeable in so many areas of our everyday life, such as traffic safety, video surveillance, car safety, or automatic driver-assistance systems in vehicles.

A robust pedestrian detection system have be able to detect people in different circumstances: different shapes and scales, huge amount of variation in the poses, and some cases of partially occluded pedestrian or whose parts blend with complex background sometimes. Therefore, we proposed a new method and tested it on two different databases with completely different set of contents that considered all the above circumstances.

This paper impart an efficient approach to deal with the pedestrian detection problem, for this purpose we propose a method based on three major tasks: features extraction using HOL, GDCT and LDCT, dimensionality reduction of feature vectors and finally the classification employing the SVM classifier tested out with many kernel functions.

The primary idea behind this work is to prove the effectiveness of the combination of two approaches on the performance of the pedestrian detection rate.

This paper is organized as follows: we will start by displaying the pre-processing techniques in section 3 and 4, where HOL and its important properties is presented, next both Global and Local DCT are introduced. The experimental results, discussion and the main experiments were conducted on INRIA person and Daimler sets. Finally yet importantly, a comparison between different approaches results is given in section 6.

2. RELATED WORK

Pedestrian detection remains a two-class classification problem that aims to classify an input image as a positive if a pedestrian occurs in the image or a negative if no pedestrian is found in. Previous approaches have been proposed in literature based on so many features extraction methods, we note VJ [14] as short for the Viola and Jones one of the first descriptor proposed and used to detect objects efficiently in real-time images, it was also trained using the AdaBoost classifier in [11] and merged both motion and appearance information to spot a walking person, Hog descriptor displayed by Dalal and Triggs [4] it is a feature descriptor used in computer vision and image processing in the purpose of detecting objects, Hol [2] and so many others that can be uncouncted in [3]. In [15] David et al. represent a new approach using Haar wavelets and edge orientation histograms fed into an Adaboost classifier to generate a good pedestrian detection compared to the approach results displayed in [4], in which Dalal and Triggs used support vector machine with histogram of oriented gradient. In the context of detecting object in the paper [16] an integral image was adopted to allow the Adaboost classifier to compute quickly the features judged critical. Then their generated classifier is combined in cascade way in order to discard quickly some background regions while concentrating computations on interest object regions. In [17] Gavrilu and Munder presents a multi-cue vision system which use cascade modes that analyse visual criteria to narrow down the search space to perform detection, similarly [18] exhibit a framework that counts four detectors trained individually to get four components of the human body, the results are then combined and classified either a “person” or a “nonperson.”, besides the work in [19] is an evaluation of the performance of descriptors which compute local interest regions and it proposes an extension of SIFT descriptor according to their results. The most popular approach for improving detection quality is to increase/diversify the features computed over the input image, by having richer and higher dimensional representations, the classification task becomes somehow easier, enabling improved results. To that matter a large set of feature types have been explored: edge information [5] [6] [7] [8], color information [6] [12], texture information [13], local shape information co-variance features [9], among others. More and more diverse features have been shown to systematically improve performance.

3. HOL

3.1. Generalities

The Histogram of oriented Laplacien proposed in [2] is an image descriptor similar to Hog descriptor. HOG is a set of characteristic descriptors, which serves to detect objects in digital images. It was introduced by Dalal and Triggs [4] and has contributed in so many works in literature in the field of features extraction. The Hog descriptor is based on the principle that the shape and the appearance of an object is described by the gradient intensity distribution and orientation of edges. In fact, the implementation of these descriptors starts by cutting the image in adjacent small areas called cells. For each cell, the histogram of the gradient orientations is commuted for each pixel. Hence, combining all those calculated histograms gives the HOG descriptor. To polish the given results, the histograms are normalized by calculating a measure of the intensity of a set of cells called blocks. This measure is used to normalize all the cells in a block. The process of building a HOG descriptor for a given image is described as follows:

- First; the input image is divided into small spatial regions corresponding to the cells.
- Then, the input image is divided into small special regions corresponding to the cells.
- For each cell, accumulate a local histogram of gradient directions or edge orientation in every pixels in the cell
- The normalization is applied on local histogram to reduce the effect of illumination variance.

This can be done by accumulating a measure of these local histograms over regions called blocks. A block is a set of cells and then use the results to normalize all cells in the block.

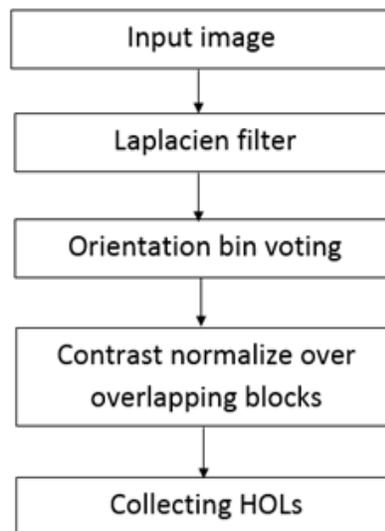


Figure 1. Operating principle of HOL

The HOL descriptor (see Figure 1), is based on the processing of the HOG descriptor where the gradient filter was replaced by the Laplacien one.

4. THE DISCRETE COSINE TRANSFORM

GDCT (Global Discrete cosine transform) and LDCT (Local discrete cosine transform) for pedestrian detection were two transforms introduced and used for pedestrian detection. Unlike other transforms, the DCT attempts to represent each image by a set of features called DCT coefficients where most of the visually significant information about the image is concentrated in just a few coefficients of the DCT, most of the time we combine the coefficient extraction to coefficient reductions to reduce the computation time.

4.1. Generalities

The Discrete Cosine Transform helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier Transform since it transforms a signal or image from the spatial domain to the frequency representation.

As known in an image, most of the energy is concentrated in the lower frequencies, so if we transform an image into its frequency components and neglect the higher frequency coefficients, we can reduce the amount of data needed to describe the image without sacrificing too much image quality. Ahmed, Natarajan, and Rao (1974) [1] first introduced the DCT in the early seventies.

The general equation for a 2D (N x M image) DCT is defined by the following equation:

$$F(u, v) = \left(\frac{2}{N}\right)^{1/2} \left(\frac{2}{M}\right)^{1/2} \sum_{j=0}^{M-1} \sum_{i=0}^{N-1} \Lambda(i, j) \cos \left[\frac{\pi u}{2N} (2i + 1) \right] \cos \left[\frac{\pi v}{2M} (2j + 1) \right] f(i, j) \quad (1)$$

Where

$$\Lambda(i) \begin{cases} 1/\sqrt{2} & \text{for } u = 0 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

$$\Lambda(j) \begin{cases} 1/\sqrt{2} & \text{for } v = 0 \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

And

f (i,j) is the 2D input sequence

The basic operation of the DCT is as follows:

- 1) The input image is N by M;
- 2) f(i,j) is the intensity of the pixel in row i and column j;
- 3) F(u,v) is the DCT coefficient in row k1 and column k2 of the DCT matrix;
- 4) For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT;
- 5) The output array of DCT coefficients contains integers; these can range from -1024 to 1023;
- 6) It is computationally easier to implement and more efficient to regard the DCT as a set of basic functions which given a known input array size (8 x 8) can be precomputed and stored. This involves simply computing values for a convolution mask (8 x8 window)

that get applied (sum m values \times pixel the window overlap with image apply window across all rows/columns of image). The values are simply calculated from the DCT formula.

In Natural images, the majority of the DCT energy is concentrated on low frequencies, so each image retain a set of low frequency features and high frequency features (edges) which are rarely encountered.

4.2. Feature reduction

The advantages of DCT are based on the fact that most images have sparse edges. Therefore, most blocks contain primarily low frequencies, and can be represented by a small number of coefficients without significant precision loss. In general, the first DCT coefficients contain most of the information about the global statistics of the processed block, which is necessary to achieve high classification accuracy. Based on that, the idea of reducing the dimension of the DCT coefficient vector length of GDCT and LDCT is eventual.

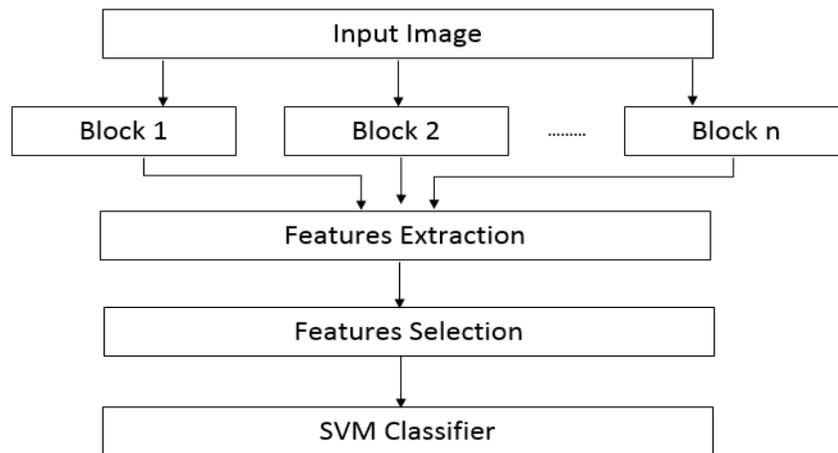


Figure 2. The principle of applying GDCT

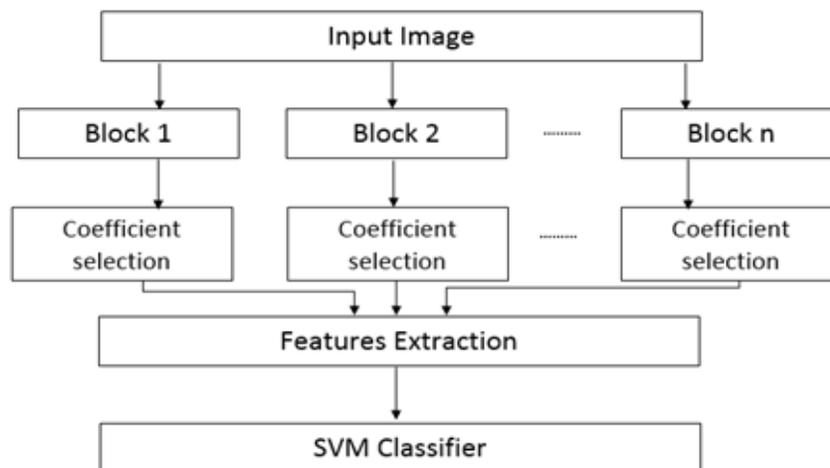


Figure 3. The principle of applying LDCT

5. SUPPORT VECTOR MACHINE

The theory of SVMs is from statistics, which is proposed by Vapnik [10]. The basic principle of SVM is finding the optimal linear hyper plane in the feature space that maximally separates the two target classes. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different classes memberships. The SVM is a state-of-the-art classification method and it is widely used in supervised classification in machine learning applications. Apart from its simplicity and accuracy, the SVM classifier is popular because of the ease with which it can deal with high-dimensional data. It performs binary classification by defining a hyper-plane that classifies the input data into two classes (Pedestrian) and (Non pedestrian). SVM uses kernel functions to classify data in a high dimension space, where the separation is performed much easily. The separation is given by the equation:

$$F(x) = \sum_{i=1}^n (\alpha_i y_i K(x_i, x)) + b \quad (4)$$

6. EXPERIMENTAL RESULTS AND DISCUSSION

Much of the progress in the past few years has been driven by the availability of challenging public datasets. In this paper, the proposed approach has been mainly applied on INRIA person along with the Daimler datasets, and simulated on the platform of Matlab.

6.1. Databases

Multiple pedestrian datasets have been collected over the years and used for testing different approaches in the field of pedestrian detection or tracking.

The INRIA person dataset is very popular in the Pedestrian Detection community, both for training detecting and reporting results. It helped drive recent advances in pedestrian detection and remains one of the most widely used databases despite its limitations. This dataset was collected as part of research work on detection of upright people in images and video.

In addition to the Inria person dataset, our approach has been tested on the Daimler pedestrian.



Figure 4. Samples of images from the INRIA person Database

The Daimler pedestrian dataset contains a collection of pedestrian and non-pedestrian images. It is made available for download over the internet for benchmark purposes, in order to advance research on pedestrian classification.



Figure 5. A sample of DC database

6.2. Proposed Method

In this paper, we propose a new approach for the pedestrian detection; it is based on the concatenations of the HOL characteristic vector respectively with the GDCT and the LDCT coefficients.

In this paper, all the images of the Inria person dataset were re-sized to square 290x290 pixels for the GDCT application, and to 288x288 pixels so they can be an integral number of 8x8 pixels per block for the LDCT application, as a result we end up with 1296 blocks of 8x8 pixels. Based on the principle of the features reduction represented in the Figure 2 and Figure 3, we reduced the number of coefficients for both GDCT and LDCT, the obtained vectors are then concatenated to the HOL feature vector and then fed into a Bi-class SVM to classify the input data as a pedestrian ID or not. LIBSVM (a library for Support Vector Machines) and different kernel functions were tested here to improve detection, by reducing the miss rate, and increasing the accuracy rate Table 1.

Table 1. Test on the Inria person.

	Kernel	Coefficients	Accuracy (%)	Miss Rate (%)
HOL	RBF	-	77.42	18
GDCT	Polynomial	84100 coef / Image	76.91	23
	Polynomial	40000 coef / Image	76.91	23
LDCT	Polynomial	64 coef / Block	76.91	23
	Polynomial	16 coef / Block	76.91	23
GDCT+HOL	Polynomial	-	76.91	23
	Linear	-	78.60	21
LDCT+HOL	Polynomial	-	32.08	86
	Linear	-	32.56	83.03

Many combination of SVM parameter were tested, the best performances were obtained using the Polynomial kernel function, and we reduced the number of the coefficients per images since we obtain better performance. The outcome of the concatenation of the GDCT and the HOL descriptor is the increase of the detection rate by 2 % and the decrease of the miss rate compared to the application of HOL or GDCT as a feature descriptor on the Inria person dataset each one alone .

As for the application GDCT of the LDCT on the Daimler, we had to rescale the images respectively to 39x39 pixels which give us a total of 1296 coefficients per image, and to 32x32 pixels for the LDCT that provide us with 16 blocks of 8x8 pixels per image, and each block enumerate 64 coefficients, the deferent results are summed up on Table 2

Table 2. Test on the DC database

	Kernel	Coefficients	Accuracy (%)	Miss Rate (%)
HOL	RBF	-	53.01	23
GDCT	Polynomial	1296 coef / Image	65.5	18
LDCT	Polynomial	64 coef / Block	67.27	19.7
GDCT+HOL	Polynomial	-	65.48	18
	Linear	-	65.12	39.80
LDCT+HOL	Polynomial	-	32.36	64.7
	Linear	-	48.59	52.37

6.3. Comparison

This section provides a comparative study of our approaches and multiple well-known techniques for pedestrian detection applied on Inria person database and the DC pedestrian data set.

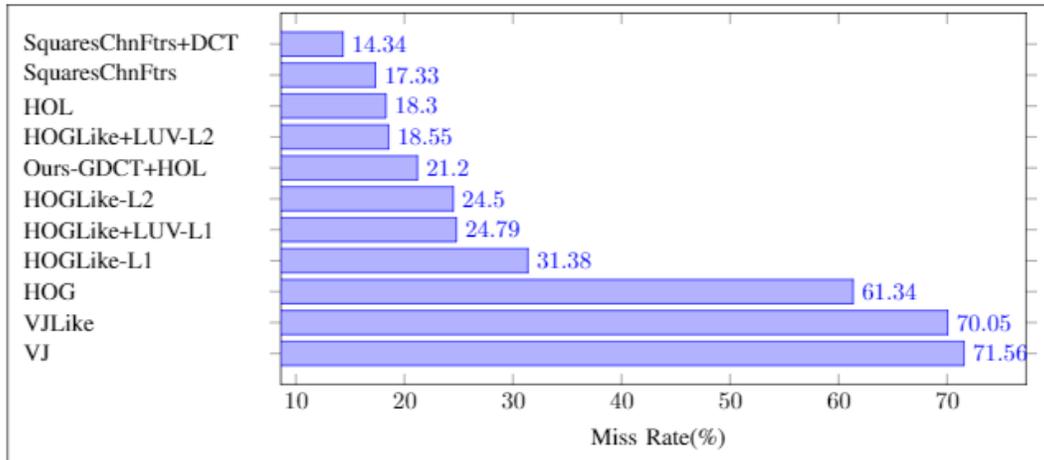


Figure 6. Effect of features on detection performance Inria Dataset

Classification performance of our proposed method is evaluated using the bar plot represented on the Figure 6, which plot the miss rate.

The miss rate is a measure that calculate the proportion of false negative rate miss detection of a pedestrian in an image, it is considered in so many research works as a procedure that shows the effectiveness of different features in the detection applied on the Inria person database.

As we have shown experimentally, our approach brings several advances when compared to commonly used and existing procedures; the reduction of the number of coefficients calculated by the mean of GDCT and LDCT, since they don't capture discriminative information, combining it to the HOL descriptor reduces the Miss rate, and places our approach in a decent position compared to the HOG [4] combined to the SVM classifier and VJ [11], the most used descriptors in the field of pedestrian detection.

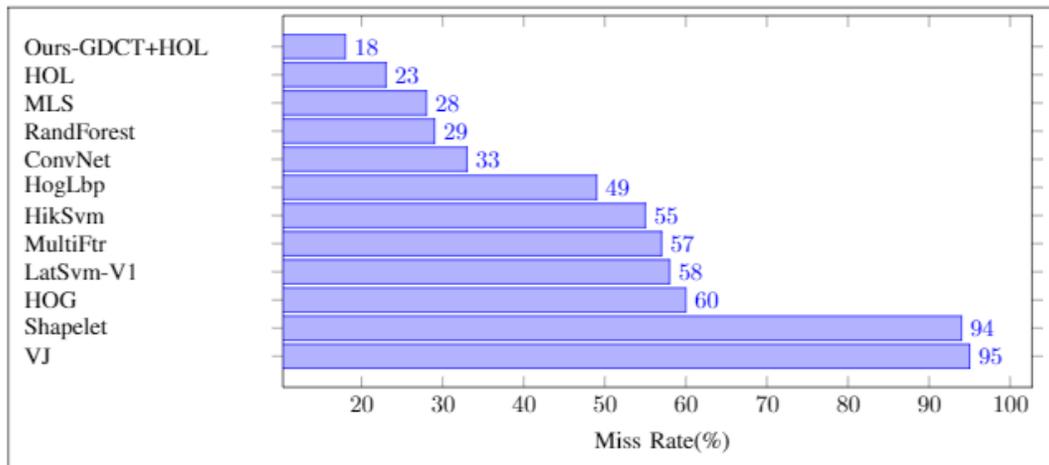


Figure 7. Effect of features on detection performance Daimler Dataset

The Daimler dataset is grayscale unlike the other datasets. As such, only a subset of the detection algorithms were applicable (those that did not rely on color information). The Figure 7 shows that the proposed feature selection proposed in this article outperform different methods proposed out there and applied on the Daimler database.

7. CONCLUSION

In this study, we putted forward an extended approach to the proposed method GDCT and LDCT applied for the pedestrian detection by combining these methods to HOL approach.

Before the combination of the two separate approaches, we dug the possibility of features reduction in both cases of GDCT and LDCT, reducing the number of studied coefficients and keeping only the more significant ones, then concatenating them to the HOL coefficient.

We showed experimentally that our proposed method for detecting pedestrian improves over some recent techniques proposed in the article [2] as well as the earlier method of [4], from which HOL seeks out its idea.

This article is an in depth comparison of the proposed method and the state of the art methods for the pedestrian detection, applied and verified on the Inria Person and Daimler databases; regarding the classification methods, we used the SVM classifier and tested various kernel functions and kept the most performing ones.

REFERENCES

- [1] N. Ahmed, T. Natarajan and K. Rao, (1974). 'Discrete Cosine Transform', IEEE Transactions on Computers, vol. -23, no. 1, pp. 90-93.
- [2] R. Benenson, M. Omran, J. Hosang and B. Schiele, (2015). 'Ten Years of Pedestrian Detection, What Have We Learned?', Computer Vision - ECCV 2014 Workshops, pp. 613-627.
- [3] A. Costea and S. Nedeveschi.,(2014). 'Word Channel Based Multiscale Pedestrian Detection without Image Resizing and Using Only One Classifier', 2014 IEEE Conference on Computer Vision and Pattern Recognition.

- [4] N. Dalal and B. Triggs,(2005). 'Histograms of Oriented Gradients for Human Detection',IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).
- [5] P.Dollár, Z.Tu, P.Perona and S.Belongie, (2009). 'Integral Channel Features', 2009, In BMVC, volume 2,page 5.
- [6] J. Jaewhan Lim, P.Dollar and C.Lawrence Zitnick III, (2013). 'Learned mid-level representation for contour and object detection', US Patent Application pp.13/794,857.
- [7] P. Luo, Y. Tian, X. Wang and X. Tang, (2014). 'Switchable Deep Network for Pedestrian Detection', IEEE Conference on Computer Vision and Pattern Recognition.
- [8] Y.Azdoud ,A. Amine , N.Alioua and M.Rziza, (2015). 'Precollision detection system for pedestrian safety based on hol'. IEEE/ACS 12th International Conference on Computer Systems and Applications (AICCSA).
- [9] S. Paisitkriangkrai, C. Shen and A. Hengel, (2013). 'Efficient Pedestrian Detection by Directly Optimizing the Partial Area under the ROC Curve', IEEE International Conference on Computer Vision.
- [10] V. Vapnik, (1998). Statistical learning theory. New York: Wiley.
- [11] P. Viola, M. Jones and D. Snow, (2005). 'Detecting Pedestrians Using Patterns of Motion and Appearance', Int J Comput Vision, vol. 63, no. 2, pp. 153-161.
- [12] S. Walk, N. Majer, K. Schindler and B. Schiele, (2010). 'New features and insights for pedestrian detection', IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- [13] X. Wang, T. Han and S. Yan, (2009). 'An HOG-LBP human detector with partial occlusion handling', IEEE 12th International Conference on Computer Vision.
- [14] P. Viola and M. Jones, (2004). 'Robust Real-Time Face Detection', International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154.
- [15] D. Gerónimo, A. López, D. Ponsa and A. Sappa, (2007). 'Haar Wavelets and Edge Orientation Histograms for On-Board Pedestrian Detection', Pattern Recognition and Image Analysis, pp. 418-425.
- [16] P. Viola and M. Jones, 'Rapid object detection using a boosted cascade of simple features', (2001). Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR.
- [17] Gavrila and S. Munder, 'Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle', (2006). International Journal of Computer Vision, vol. 73, no. 1, pp. 41-59.
- [18] A. Mohan, C. Papageorgiou and T. Poggio, 'Example-based object detection in images by components', (2001), IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 4, pp. 349-361.
- [19] K. Mikolajczyk and C. Schmid, 'A performance evaluation of local descriptors', (2005). IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630.