An Object Detection System using Image Reconstruction with PCA

Luis Malagón-Borja and Olac Fuentes

Instituto Nacional de Astrofísica Óptica y Electrónica, Puebla, 72840 Mexico
jmb@ccc.inaoep.mx, fuentes@inaoep.mx

Abstract

We present an object detection system that is applied to detecting pedestrians in still images, without assuming any a priori knowledge about the image. The system works as follows: In a first stage a classifier examines each location in the image at different scales. Then in a second stage the system tries to eliminate false detections based on heuristics. The classifier is based on the idea that Principal Components Analysis (PCA) can compress optimally only the kind of images that were used to compute the principal components (PCs), and that any other kind of images will not be compressed well using a few components. Thus the classifier performs separately the PCA from the positive examples and from the negative examples, when it needs to classify a new pattern it projects it into both sets of PCs and compares the reconstructions. The system is able to detect frontal and rear views of pedestrians, and usually can also detect side views of pedestrians despite not being trained for this task. Comparisons with other pedestrian detection systems are presented; our system has better performance in positive detection and in false detection rate.

1. Introduction

The problem of object detection can be seen as a classification problem, where we need to distinguish between the object of interest and any other object. In this paper we focus on a single case of the object detection problem, detecting pedestrians in images.

Pedestrian detection is more difficult that detecting many other objects due to the fact that people can show widely varying appearances when the limbs are in different positions; in addition people can dress in clothes with many different colors and types. For the characteristics of the pedestrian class we need a robust method that can learn the high variability in the pedestrian class.

Many object detection systems that have been developed focus on face detection. One of the most successful is the face detection system of Rowley et al. [10], which consists of an ensemble of neural networks and a module to reduce false detections. Similar example-based face detection systems have been developed by Sung et al. [11], Osuna et al. [8], and M. H. Yang et al.[12].

Most pedestrian detection systems use motion information, stereo vision, a static camera or focus on tracking; important works include [1], [2], [3] and [13]. Papageorgiou has reported a system [6], [7], [9] to detect pedestrians in images, without restrictions in the image, and without using any other information besides the image. It uses the wavelet template to represent the image and a Support Vector Machine (SVM) to classify. The system has been improved in [4] and [5], detecting pedestrians through the detection of four components of the human body: the head, legs, left arm and right arm.

We present an object detection system to detect pedestrians in gray level images, without assuming any a priori knowledge about the image. The system works as follows: In a first stage a classifier based on Principal Components Analysis (PCA) examines and classifies each location in the image at different scales. Then, in a second stage, the system tries to eliminate false detections based on two heuristics.

The system uses PCA as a classification tool; the main idea is that PCA can compress optimally only the kind of images that were used to do the PCA, and that any other kind of image will not be compressed well in a few attributes, so we do PCA separately for positive and negative examples, when a new pattern needs to be classified we compare the reconstruction made by both sets of principal components (PCs). In order to improve the performance of the classifier we have used the edge image as additional information for it.

The reminder of the paper is organized as follows: Section 2 presents a detailed description of the system. In Section 3 the performance of our system, and a comparison with similar systems are presented. Section 4 reports conclusions and possible directions for future work.

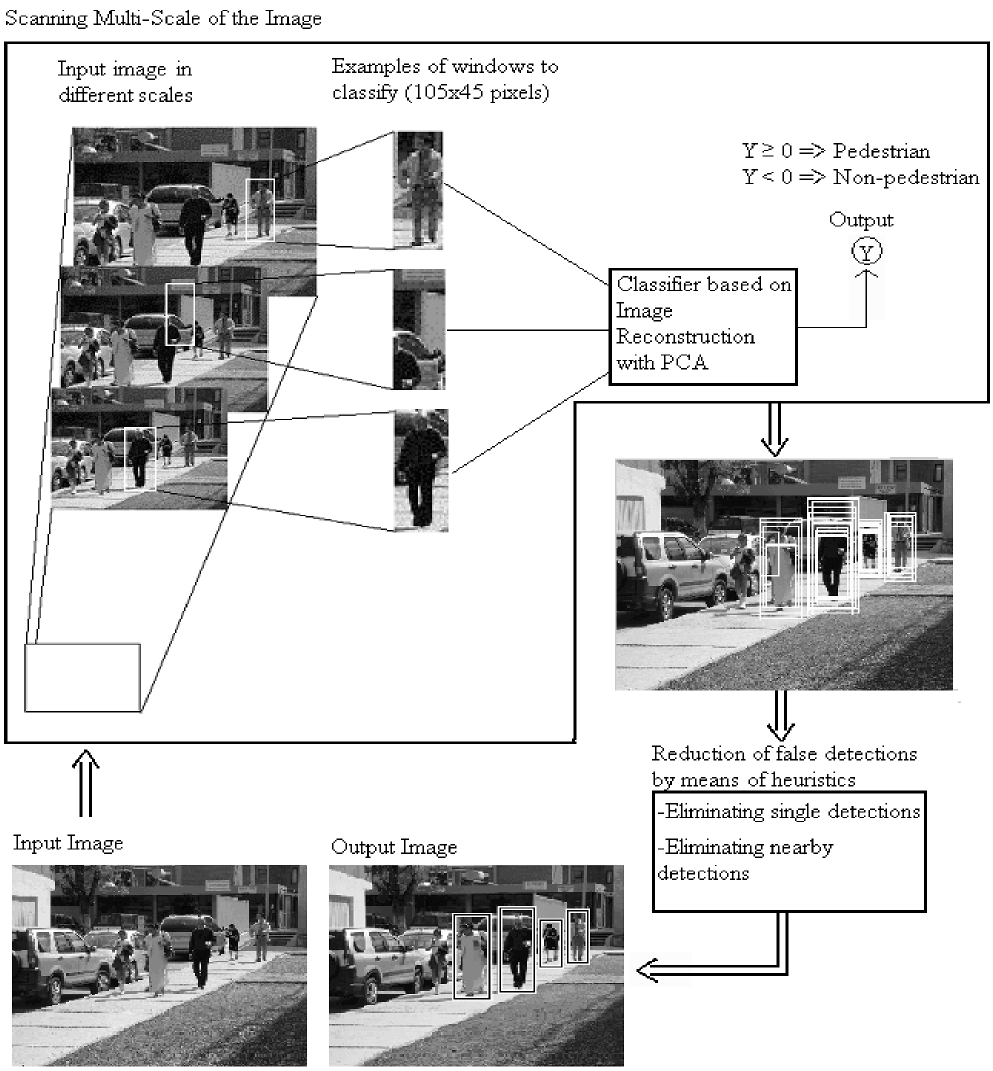


Figure 1: Architecture of the system for pedestrian detection in images

2. The Detection System

2.1. Overview of System Architecture

The system works scanning the whole image by means of a detection window of size 105x45 pixels; the window is shifting with two pixel jumps to accelerate the process without losing a lot of information from one window to another. We need a classifier that decides for each window if it contains a pedestrian or not. The construction of the classifier is the most complicated stage, we have created a classifier based on image reconstruction with PCA , this classifier uses beside the gray level image, the edge image.

The scanning of the whole image is part of an iterative process where the image is resized several times to achieve multi-scale detection. For our experiments, the image has been scaled from 0.26 up to 1.35 times its original size, with increases of approximately 17% in every cycle, thus the image is processed with 12 different scales that are the following: 0.26, 0.3, 0.35, 0.4, 0.47, 0.55, 0.64, 0.74, 0.86, 1, 1.17 and 1.35, this implies that pedestrians of size between 78x33 and 404x173 pixels will be detected by the system.

When the system has finished examining the image in all scales, a second process eliminates some detections that are believed to be false detections. The form in which this process works is eliminating the detections that do not repeat several times and eliminating the detections that overlap.

Figure 1 shows the complete process to detect pedestrians in an image, starting with the gray level image and finishing with the image with the detected pedestrians.

Figure 2: Edge Images. The edge images eliminate information about color and texture therefore they present less variation among pedestrians.

2.2. Stage 1: A Classifier based on Image Reconstruction with PCA

In this stage we present a classifier that decides if an image of size 105x45 belongs or does not belong to the pedestrian class. This classifier is based on doing image reconstruction using PCA and comparing the reconstructed with the original images. First, the reasons to work with both the gray level image and the edge image are explained, later it is explained how the reconstruction of an image is performed using PCA and finally, we present the way in which a classifier can use these reconstructions to decide if an image belongs or does not belong to the pedestrian class.

2.2.1. Edge Images

Because pedestrians appear in many colors and different textures, it is not advisable to use characteristics based on color or texture to do pedestrian detection, for this reason, we have chosen to use the edge image with the idea of obtaining the typical silhouette of a pedestrian and to eliminate a lot of useless information for the classifier.

The edge images were computed using x and y Sobel filters, this edge image serves as complementary information to the gray level image and it allows the classifiers to obtain more data to decide if an image is a pedestrian or not.

In Figure 2 we can see examples of the corresponding edge images of some pedestrian gray level images. In these images we can observe that although the gray level images are very different in color and background, the edge images present fewer changes from one image to another. This is the reason why the edge images are very important to help in the classifiers' task.

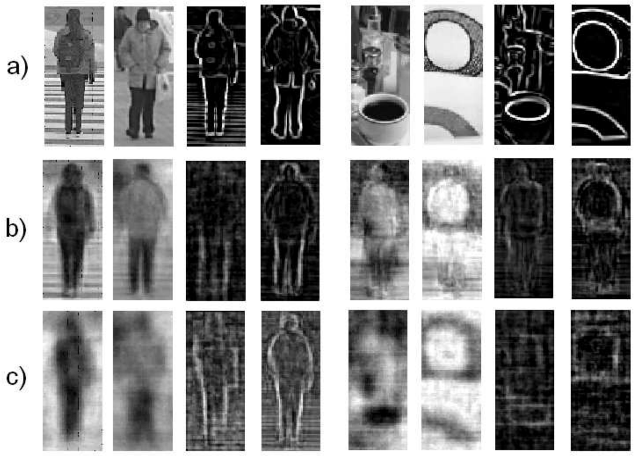


Figure 3: Image Reconstruction with different sets of PCs. In row a) the original images, in row b) the images reconstructed using 100 PCs obtained from pedestrian images and in row c) the images reconstructed using 100 PCs obtained from non-pedestrian images. We can see that for both the gray level images and the edge image, the pedestrian images are better reconstructed with the PCs obtained from pedestrian images (row b) than with the PCs obtained from non-pedestrian images (row c). This does not happen with the non-pedestrian images, which are better reconstructed with the PCs obtained from non-pedestrian images (row c).

2.2.2. Image Reconstruction with PCA

When we speak of reconstructing an image with PCA, what we understand is to project the image into the PCs, and from this projection, we try to recover the original image.

Let A be the matrix with n PCs and X the mean object, obtained when we do PCA with a set of M input images, and suppose that we want to reconstruct an image of size 105x45, which is described by the column vector I of size 4725x1, what we do first is to obtain the difference with the mean object.

$$\Phi = I - X$$

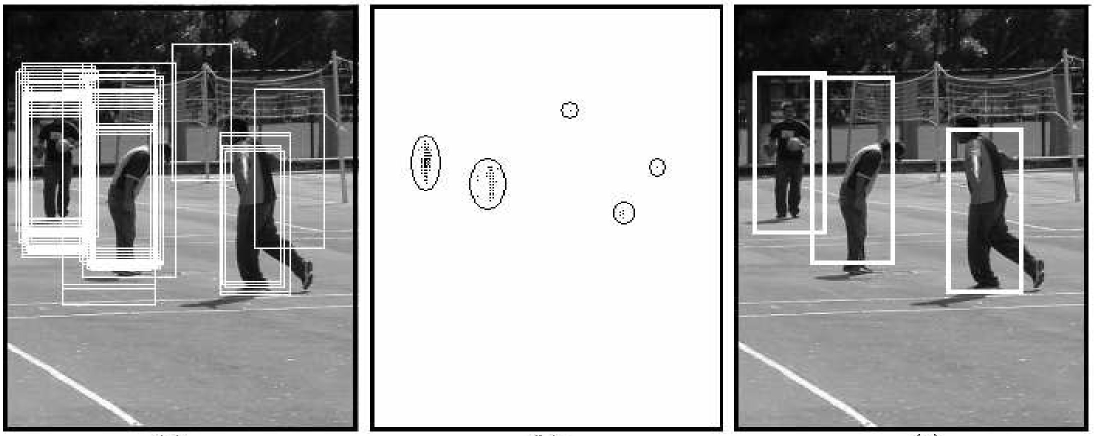
Next we obtain the projection

$$P = A^T \cdot \Phi$$

Finally to obtain the reconstruction I' of the image starting from projection P, we do the following operations

$$I' = A \cdot P + X$$

The more PCs we use to obtain the projection, the less loss of information we will have, therefore the reconstruction of the image will be more accurate.



(a) (b) (c)

Figure 4: Process to eliminate single false detections. In Figure (a) we can see the original detections found by the classifier. In Figure (b) each detection is grouped with those detections whose centroid is in the same neighborhood. Finally in Figure (c) we have the grouped detections composed by two or more original detections.

2.2.3. Classification using Reconstruction

By definition, PCA looks for the set of PCs that better describes the distribution of the data that are being analyzed. Therefore, these PCs are going to preserve better the information of the images from which PCA was performed, or of those that are similar, thus, if we have a set of PCs that was obtained from a set of pedestrian images only, these must reconstruct better the images of other pedestrians than any other type of images, and viceversa, if we have a set of PCs obtained from images of anything except pedestrians, the reconstruction of the pedestrian images will not be as good. We can observe this fact in Figure 3, both for gray level images and for edge images.

From this fact we can create a classifier based on image reconstruction with PCA, which decides if an image belongs or does not belong to the pedestrian class. The algorithm to do this classification is the following:

Before doing any classification:

1. Perform PCA on the set of pedestrian gray level images to obtain PCs_1.
2. Perform PCA on the set of pedestrian edge images to obtain PCs_2.
3. Perform PCA on the set of non-pedestrian gray level images to obtain PCs_3.
4. Perform PCA on the set of non-pedestrian edge images to obtain PCs_4.

   When we want to classify a new gray level image:

5. Obtain the edge image from the gray level image.
6. Do four reconstructions, two from the gray level image and two from the edge image, using in every reconstruction one of the four sets of PCs.
7. Compare the reconstructions with the original images, pixel to pixel, in order to obtain four differences d_1, d_2, d_3 and d_4.
8. Add the two differences obtained in the reconstructions with the sets of PCs obtained from pedestrian images d_1 and d_2, to obtain the difference with respect to the pedestrian class D_p.
9. Add the two differences obtained in the reconstructions with the sets of PCs obtained from non-pedestrian images d_3 and d_4, to obtain the difference with respect to the non-pedestrian class D_n.
10. Define the classification value $Y = D_n - D_p$.
11. Classify the image according to the following criterion

$$\begin{cases} \text{Pedestrian} & Y \geq 0 \\ \text{Non-pedestrian} & Y < 0 \end{cases}$$

2.3. Stage 2: Reduction of False Detections by means of Heuristics

The output after classifying all the windows of the image in multiple scales still contains an important number of false detections, in this stage we present two heuristics that allow to diminish the false detections by means of two processes, namely, eliminating single detections and eliminating nearby detections.

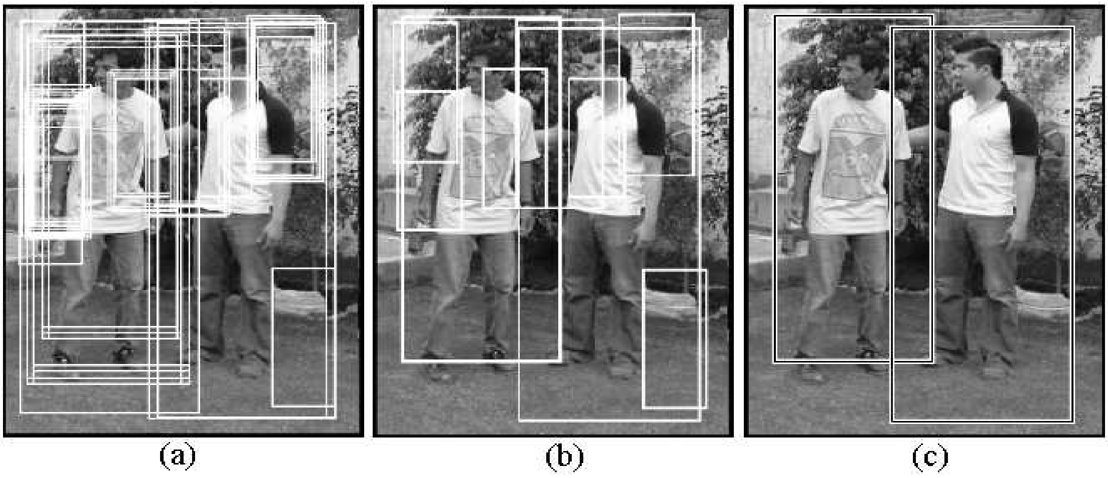


Figure 5: Process to eliminate nearby false detections. Figure a) shows the detections found by the classifier. Figure b) shows the grouped detections. In Figure (c) the grouped detections with the greatest $Preference$ have been preserved and the nearby grouped detections have been eliminated.

2.3.1. Eliminating single detections

As we can see in Figure 4(a), most of the pedestrians are detected at multiple nearby positions and scales, while false detections usually appear at a single position. This observation allows us to eliminate some false detections, eliminating detections that appear only once.

Each detection found can be grouped with those detections whose centroid is inside the same neighborhood, obtaining a new set of detections which we will call grouped detections, composed by one or more of the original detections. Once we have the set of grouped detections we will ignore the original detections, and we will eliminate the grouped detections composed by only one original detections. We can see this process in Figure 4.

2.3.2. Eliminating nearby detections

If a window is identified correctly as a pedestrian, then it is very likely that there are no pedestrians either above or below it, and if there are pedestrians beside it, they cannot be too overlapped. This heuristic allows us to eliminate nearby detections. With this purpose, we define a region around a detection which we are going to use to eliminate any detection whose centroid is inside this region. The size of the region was defined empirically as 1.4 times the detection height upwards and downwards from the centroid and between 0.5 and 0.75 times the detection width towards each side of the centroid.

Until now we know that when two detections are in the same region we must eliminate one, but how are we going to decide which one we are going to maintain? The most reasonable way to choose one would be to maintain the grouped detection composed by more original detections, nevertheless, we observe that usually the biggest detections were the correct ones, due to the fact that frequently arms, legs and head are confused with pedestrians, so when we need to decide among a set of detections that are in the same region, we must consider the number of times that they have been detected originally as well as the size of the detections.

To achieve this, the detections that compose a grouped detection are weighted by their height, then the grouped detection with the greatest $Preference$, according to the next formula is chosen.

$$Preference = Detections \cdot Weight(height)$$

where $Detections$ is the number of original detections that compose the grouped detection that we are evaluating and $Weight$ is a function that determines the value that each detection has, according to the $height$ of the grouped detection, and it is given by the formula:

$$Weight(height) = (height - 50)^2$$

There are very few cases where this heuristic does not work, and thus it allows to eliminate many false detections when the classifier confuses the arms, the legs, or some other object with a pedestrian. In Figure 5 it is possible to see an example where some false detections are eliminated applying this heuristic to the output of the classifier.

3. Experimental Results

As we explained in the previous section, we need a set of pedestrian images and a set of non-pedestrian images to obtain the four sets of PCs from which we are going to perform the four reconstructions.

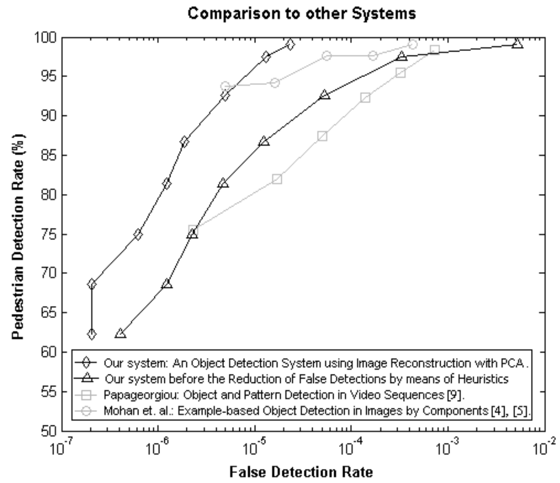


Figure 6: ROC curves comparing the performance of our system versus the best reported in the literature. The detection rate is plotted against the false detection rate measured on logarithmic scale.

The pedestrian images were obtained from the MIT pedestrian database, which contains scenes with pedestrians in frontal or rear views. We convert these color images to gray level images and we crop part of the background to diminish the variation that exists among pedestrian images, finally obtaining a set with 500 gray level images. The set of negative images had 2315 images that were obtained randomly from a set of 90 images of scenery that did not contain any pedestrian.

In the experiments we used 200 PCs in each set, that contain between 75% and 85% of the variance, to do the reconstructions. It was observed that this number of PCs allowed a good classification.

The system was tested with a database containing 204 pedestrian images in frontal or rear view to determine the pedestrian detection rate; these images were not used before. The false detection rate was obtained by running the system over a database with 17 images that did not contain any pedestrian; by running the system over these 17 images, 4,850,103 windows were classified.

In general, the performance of any object detection system shows a tradeoff between the positive detection rate and the false detection rate. We ran the system over the test images at several different thresholds. The results were plotted as a Receiver Operating Characteristic (ROC) curve, given in Figure 6; we also show results before the reduction of false detections by means of heuristics. The curve indicates that the system achieves a detection rate of 99.02% with one false positive every 43,304 windows examined, or if we want a more conservative system, it can achieve a detection rate of 81.37% with one false detection every 808,351 windows examined. Also we can see that our system has better performance that the best reported in the literature of pedestrian detection in systems than do not assume any a priori scene structure or use any motion information.

Figure 7 shows the result of applying the system to sample images in cluttered scenes under different lighting conditions.

4. Conclusions and Future Work

In this paper we have presented an object detection system for static images, without assuming any a priori knowledge, applied to the specific problem of locating pedestrians in cluttered gray level images.

Our system is able to detect frontal and rear views of pedestrians, and usually it can also detect side views of pedestrians despite not being trained for this task.

The success of PCA for pedestrian detection comes from its capability to capture the most of the information that allows to distinguish between a pedestrian image and any other image in the huge universe of non-pedestrian images.

The current system does not work as well for side views of pedestrians as for pedestrians in frontal or rear views. To solve this, we can add side views of pedestrians to the training set, or we can create an additional part of the system that could be specialized for these views.

Another way of improving the system's performance is to obtain more positive and negative examples for training. We only use 500 positive examples and 2315 negative examples, while other works in object detection use around 2000 positive examples and 10000 negative examples.

A promising direction for future work is to apply the method presented in this paper in a component-based approach. This approach has shown better performance in pedestrian detection (see [4] and [5]) than a similar full-body pedestrian detector [6], [7], [9].

The framework described here is applicable to other domains besides pedestrians; it may be generalized to the detection of several different types of objects, such as faces, vehicles and others.

We have presented a classification method based on PCA applied to the pedestrian detection problem. The edge image helps to improve the performance of the classifier, however it is not necessary if we were working in a domain different from images. It would be interesting to investigate if the approach described in this paper could be extended to other classification problems besides object detection.

References

[1] Fablet, R., Black, M.J. "Automatic detection and tracking of human motion with a view-based representation." In *European Conference on Computer Vision.* pp. 476-491 2002.

Figure 7: These images demonstrate the capability of the system for detecting people in still images with cluttered backgrounds.

[2] I. Haritaoglu, D. Harwood, and L. Davis, "Who, When, Where, What: A Real Time System for Detecting and Tracking People," *Procedings Third Face and Gesture Recognition Conference,* pp. 222-227, 1998.

[3] Jabri, Duric, Wechsler, and Rosenfeld. "Detection and location of people in video images using adaptive fusion of color and edge information." *International Conference on Pattern Recognition,* vol.4 pp. 46274631, 2000.

[4] A. Mohan. "Object detection in images by components." *A.I. Memo 1664, Center for Biological and Computational Learning, M.I.T., Cambridge, MA,* 1999.

[5] A. Mohan, C. Papageorgiou and T. Poggio, "Example-based object detection in images by components,"*IEEE transactions on Pattern Analysis and Machine Intelligence,* vol.23, no. 4, Apr., 2001.

[6] M. Oren, C. Papageorgiou, P. Shinha, E. Osuna, and T. Poggio, "A Trainable System for People Detection."*In Procedings of Image Understanding Workshop,* pp.207–214, 1997.

[7] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian Detection Using Wavelet Templates,"*Procedings IEEE Conference Computer Vision and Pattern Recognition,* pp. 193-199, June 1997.

[8] E. Osuna, R. Freund, and F. Girosi, 'Training Support Vector Machines: An Application to Face Detection,' *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 130 - 136, 1997.

[9] C. Papageorgiou. "Object and Pattern Detection in Video Sequences." *Masters thesis, MIT,* 1997.

[10] H. Rowley, S. Baluja, T. Kanade, "Neural Network-based face detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(1), pp. 23-38, 1998.

[11] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection" *IEEE Transactions Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 39-51, Jan. 1998.

[12] M.-H. Yang; D. Roth, and N. Ahuja, "A SNoW-based face detector", *In Advances in Neural Information Processing Systems 12,* pp. 855-861, 2000.

[13] L. Zhao and C. Thorpe, "Stereo and Neural Network-Based Pedestrian Detection." *IEEE Transactions on Intelligent Transportation Systems,* Sept. 2000.