High Occupancy Vehicle Detection

Alberto J. Pérez-Jiménez, Jose Luis Guardiola, and Juan Carlos Pérez-Cortés

Instituto Tecnológico de Informática*
Universidad Politécnica de Valencia
Camino de Vera s/n, 46022 Valencia, Spain
aperez@disca.upv.es, joguagar@iti.upv.es, jcperez@disca.upv.es

Abstract. High occupancy vehicles lanes (HOV) are highway lanes usually reserved to vehicles carrying at least two persons. They are designed to help move more people though congested areas. In this context, automatic passenger counting systems could be useful to grant access to or to control vehicles in those lanes.

In this work, we propose a real-time passenger detection system based on the analysis of visual images. Each person is detected by mixing the information from different types of classifiers in order to make the detection process faster and more robust.

1 Introduction

Detecting the number of passengers in vehicles can be useful to grant access to or to control vehicles in high occupancy lanes (HOV). Those lanes are usually reserved to vehicles carrying at least two users and they are designed to help move more people though congested areas. Variable highway toll fees depending on passenger number and traffic conditions (HOT) is another possible application.

The control of vehicles in those lanes is usually performed by humans, although some attempts to automate or ease this task have been performed. In [1] a video-based system operated by humans is proposed to assist in this task. A first attempt to automate the problem is proposed in [2]. In that work, authors propose to detect passenger faces by analyzing differences in visible and near infrared images by means of computers. Based on that technology, some commercial systems have appeared recently [3].

In order to work correctly, those visual systems require the occupants to explicitly make themselves visible. For this reason in [4] the use of internal passenger sensors already present in the vehicles (safety belt detection, airbags activation, etc.), is proposed along with transponder technology to count the number of passengers. The major drawback of this method is that it implies vehicle modification.

We propose a passenger detection system based on the analysis of visual images. The passengers are detected by combining the results of different types of classifiers, searching for different features that characterize people (faces, safety

* Work partially supported by the Spanish *Ministerio de Educación y Ciencia* under project DPI2006-15542-C04 and Consolider CSD2007-018.

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 782–789, 2008.

belts, etc.). A cascade of boosted classifiers [5,6] is implemented for fast feature detection. To minimize false detections, more powerful but slower classifiers, as k nearest neighbor (k-NN) [7], are used to filter the cascade classifier results.

This work is organized as follows: section 2 describes the proposed system, section 3 describes the experiments performed to test the system. section 4 shows the results obtained and, finally, a 5 section is included.

2 System Description

2.1 Faces Detector

In a first attempt to count passengers, a face detector classifier was used. This classifier analyzes subwindows of the image at different positions and sizes (see figure 2). The classifier used is in fact a cascade of boosted classifiers based on *haar-like* features [5,6]. As can be seen in figure 1, this classifier is formed by a set of classifiers arranged in a cascaded structure. Each classifier refuses the input data or allows them to pass to the next stage. Input windows not refused in any stage are detected as positives.

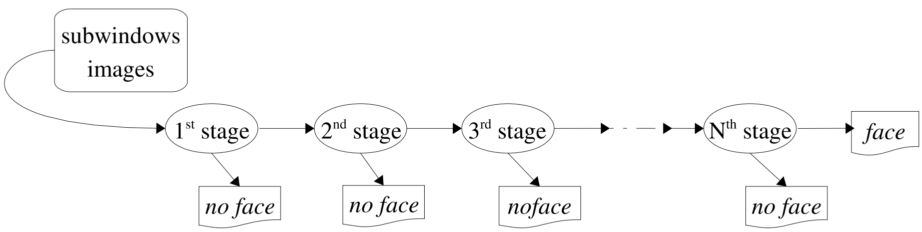

Fig. 1. Cascade classifier structure

At each stage, the classifier is designed to allow a maximum false positive rate (F_p), while having a minimum success rate (T_p). Doing so, the first stages generate very simple classifiers that refuse very quickly those subwindows of the image that are very different from the objects to detect (faces in our case), while more complex and costly last stages deal with the more ambiguous subwindows. This strategy allows to analyze images very fast but, due to the complexity of the captured images (bad contrast, highlights in the windshield, wildly changing illumination, etc., see figure 3), a too high false positive (false detection) rate is obtained.

To reduce the false positive rate, a k nearest neighbor classifier (k-NN) [7] is trained to filter the results of the cascade classifier. This type of classifier is simple to train and can be very powerful, although its computational cost is high, which restricts its use in this case to the analysis of a moderate number of selected hypothesis rather than the whole image.

The system works as follows: the subwindows detected as faces by the cascade classifier are then fed into the k-NN classifier to be classified again as *face* or

Fig. 2. Subwindows of different sizes are analyzed at different positions all over the input image

Fig. 3. Low contrast image with windshield highlights obtained by the acquisition system

false . To do so subwindows are first resized to 20×20 pixels to obtain a 400-dimensional vector representation. In order to simplify the k-NN task, and speed up the classification process, the dimensionality is reduced by means of principal component analysis (PCA) [7]. A fast implementation of the k-NN classifier based on *kdtrees* [8] has been implemented.

Finally, a join process is performed to merge face detections in the same area, since one face can be detected by the cascade classifier at several scales at slightly different positions.

2.2 Safety Belt Detectors

A significant number of occluded faces occurs in the captured images due to windshield highlights, car sunshades, etc. In order to be able to correctly account for the

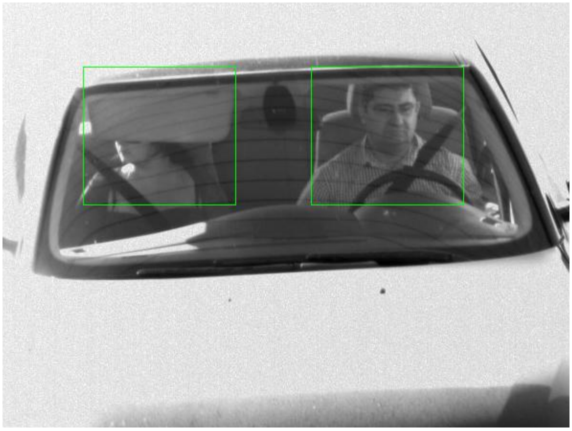

Fig. 4. Right passenger face is hidden behind the car sunshade but the safety belt is perfectly visible. AOIs are drawn in the image.

real number of passengers, other features can be used. After studying the images, safety belts seem to be a good additional feature to look for (see figure 4).

Due to the simple geometry of the safety belts, the cascade classifier seems specially suitable for detecting them. Two new cascade classifiers were created for left and right safety belts, this simplified the training process and allowed us to filter out false detections that are at the wrong side of the car.

After safety belt detection, a join process is performed as before. In this case, an additional filter is used, taking into account the number of detections in a group, allowing us to easily filter out false detections.

2.3 Areas of Interest

The areas of interest (AOI) are rectangular areas defined to bound the position of each passenger in the image (see figure 4). The AOIs are defined relatively to the windshield location and allow us to filter out false detections in wrong positions, simplifying the merge of the information from faces and safety belt detectors.

The filtering process is done by excluding detected faces that are not completely inside the AOIs. In the case of safety belt detections, a certain amount of overlap is allowed. A passenger detection will be signaled in an AOI if either a face or a safety belt in the right position is detected.

Because the AOIs are defined relative to the windshield, they only have to be redefined when the camera view is changed.

2.4 Context Information

Finally, any information about the capture context that could simplify the problem could be added to the system. For instance, we can consider that the driver will always be in the car, this simplifies the problem since it is only necessary to identify the existence of the right passenger (or the left one some countries).

2.5 Final System

A graphical summary of the whole system can be seen in figure 5.

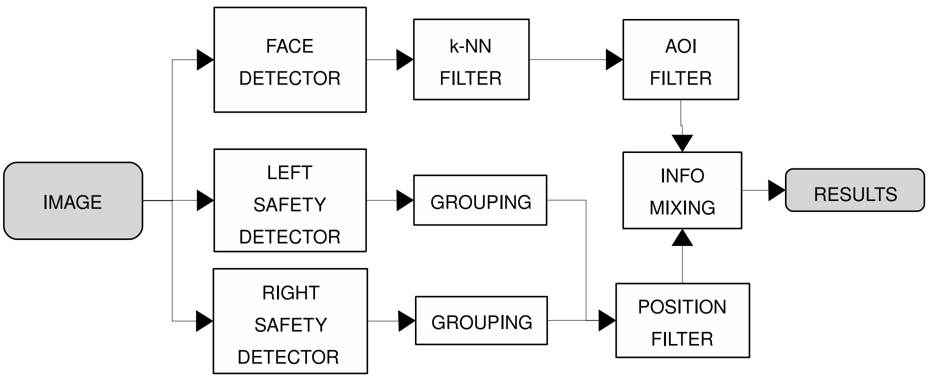

Fig. 5. Architecture of the system for passenger detection

3 Methods and Experiments

To develop and test the system, a set of 3357 images with a 1280×1024 pixel resolution was employed. The images were taken from a highway toll station using the same capture system throughout several days. This set was split randomly into a training set of 2347 images and a test set of 1210 images. Every image in the training and in test sets was manually labeled to mark the location of the faces and the windshield.

The cascade classifier for face detection was trained using the whole training set. The T_p and the F_P rates for every stage were chosen to be 0.999 and 0.5 respectively, obtaining a final classifier with 19 stages.

The k-NN classifier was trained using the subwindows detected as faces by the cascade classifier. Those subwindows were labeled as *face* or *false* according to the ground truth. In order to obtain a larger number of samples, the last 4 stages of the cascade classifier were not used to obtain the training corpus for the k-NN classifier. As stated in the previous section, each subwindows was rescaled to 20×20 pixels and the dimensionality reduced by means of PCA from 400 to 50. A total of 505000 training vectors were obtained. A value of 13 for the k parameter was estimated to give the best success rate.

For the safety belts detectors, a set of 200 images randomly selected from the training set were used. The images were labeled manually to indicate the safety belt positions. Two cascade classifiers were trained using the values of 0.99 and

0.5 for T_p and F_p respectively. After the training process, a number of 9 and 7 stages were obtained for the right and left safety belt detectors respectively.

After experimentation a group size parameter of 4 is selected for safety belts joining filtering. That is, after safety belts detection, groups are formed and those not having at least 4 members are eliminated. This helps to minimize the number of false detections.

The test set (1280 images) has been used to evaluate the performance of the developed system. In the next section, the obtained results are shown.

4 Results

4.1 Detection Results

Table 1 shows the results obtained over the whole test in different stages of the system development. The *ok* label refers to the percentage of images where the number of passengers was correctly detected, while the label *false detection* refers to the percentage of images where a number of passengers larger than the real one were detected.

Table 1. Passenger detection results obtained over the test set in the different stages of the system

	Faces	**Faces + k-NN**	**Faces + k-NN + safety belts**
Ok	79.35 %	77.66 %	84.42 %
False detection	12.48 %	1.61 %	1.92 %

We can see how the cascade classifier for faces presents a high false detection rate, this rate is significantly reduced by the k-NN classifier, losing only a small percentage of correctly identified images. After adding the safety belt detectors to the system, the *ok* success rate increased significantly without an important increase of the *false detection* percentage.

In many areas, where most vehicles have the wheel at the same side, we can improve the success rate considering that “context information”. In table 2 the improvement obtained using this information is shown.

Finally, the detection results using only the safety belt detectors are shown in table 3. As we can see, the success rate is only moderate, but it is important

Table 2. Passenger detection results using context information (driver always present) or not

	No context	**Driver present**
Ok	84.42 %	89.60 %
False detection	1.92 %	1.92 %

Table 3. Passenger detection results based only in the safety belts detector

	Safety belts
Ok	66.27 %
False detection	0.51 %

to recall that safety belts can not be seen in all of test-set images. On the other hand, the *false detection* percentage is really small, and therefore this detector will not introduce a significant number of *false detections* in the final result (as can be seen in table 1).

4.2 Computational Cost

The system takes approximately 5 seconds to analyze a whole image of 1280×1024 pixels resolution in a computer with an Intel Core 2 Duo 2.0Ghz. If the windshield is pre-located (something easily done in most cases), an important reduction of the computational cost is obtained, taking about 0.5 seconds the image process with windshields of size around 600×200 pixels.

A multithreaded implementation was performed to take advantage of new multicore processors. The cascade classifier and the k-NN applications were parallelized using the support provided by the *openCV* libraries [9] and the *Open MultiProcessing* directives provided in the 4.2 GCC compiler.

With the parallelized code the computational cost can be reduced to 60 % in double core systems. Quadruple core systems only attain a reduction of 50 % due probably to the dual port memory bottleneck.

5 Conclusions and Future Works

The system described is able to detect the frontal passengers with a high success rate (89.60 %), while maintaining the false detection percentage relatively low (1.92 %). The system could be used to control vehicles in the HOV with a low probability of allowing unauthorized vehicles (i.e., with only one person) to stay in those lanes.

The use of different cascade classifiers to detect different passenger features has proved to be effective in our problem, maintaining a reasonable computational cost. Finally, combining cascade classifiers with k-NN classifiers can lead to fast and precise classification.

An important percentage of images from the set used show a very low contrast, windshield highlights and other acquisition problems. Improving the quality of images will lead to significantly better results.

A confidence rate measure of the hypothesis obtained for every image could be useful for practical purposes. Having this measure, images over a certain confidence threshold can be processed automatically, while images with a confidence value under the threshold will be supervised by human operators. This way, even if the succes rate is not close to 100%, the system can be used to effectively reduce the installation costs.

In our work the windshield location was indicated manually during the labelling process, a robust automatic system to detect the windshield based on the Hough transform and corner classification is already being tested.

Finally because our method only uses images in the visual range, the required image adquisition equipment presents a lower cost compared to IR-based methods.

References

1. Billheimer, J.W., Kaylor, K., Shade, C.: Use of videotape in hov lane surveillance and enforcemente. Technical Report 6429, California Department of Transportation (1990)
2. Pavlidis, I., Symosek, P., Fritz, B., Bazakos, M., Papanikolopoulos, N.: Automatic detection of vehicle occupants: the imaging problem and its solution. Machine Vision and Applications 11(6), 313–320 (2000)
3. Video-Occupancy, C.: Cyclops, occupancy counter system (2005), `http://www.vehicleoccupancy.com/products.htm`
4. Schijns, S., Matthews, P.: A breakthrough in automated vehicle occupancy monitoring sustems for hov/hot facilities. In: 12th HOV Systems Conference (2005)
5. Viola, P., Jones, M.: An extended set of haar-like features for rapid object detection. In: IEEE ICIP, vol. 1, pp. 900–903 (2002)
6. Lienhart, R., Maydt, J.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
7. Fukunaga, K.: Statistical Pattern Recognition, 2nd edn. Academic Press, London (1990)
8. Friedman, J., Bentley, J., Finkel, R.: An algorithm for finding best matches in logarithmic expected time. In: ACM Transactions on Mathemathical Software, vol. 3, pp. 209–226 (1977)
9. Davies, B., Bornet, O., Bradki, G.: altera: Open Computer Vision Library. Version 1.0 (2006), `http://sourceforge.net/projects/opencvlibrary/`