Visual Monitoring of Railroad Grade Crossing

Yaser Sheikh, Yun Zhai, Khurram Shafique, and Mubarak Shah

University of Central Florida,
Orlando FL-32816, USA.

ABSTRACT

There are approximately 261,000 rail crossings in the United States according to the studies by the National Highway Traffic Safety Administration (NHTSA) and Federal Railroad Administration (FRA). From 1993 to 1998, there were over 25,000 highway-rail crossing incidents involving motor vehicles - averaging 4,167 incidents a year. In this paper, we present a real-time computer vision system for the monitoring of the movement of pedestrians, bikers, animals and vehicles at railroad intersections. The video is processed for the detection of uncharacteristic events, triggering an immediate warning system. In order to recognize the events, the system first performs robust object detection and tracking. Next, a classification algorithm is used to determine whether the detected object is a pedestrian, biker, group or a vehicle, allowing inferences on whether the behavior of the object is characteristic or not. Due to the ubiquity of low cost, low power, and high quality video cameras, increased computing power and memory capacity, the proposed approach provides a cost effective and scalable solution to this important problem. Furthermore, the system has the potential to significantly decrease the number of accidents and therefore the resulting deaths and injuries that occur at railroad crossings. We have field tested our system at two sites, a rail-highway grade crossing, and a trestle located in Central Florida, and we present results on six hours of collected data.

Keywords: Railroad Crossing, Surveillance, Tracking, Object Detection

1. INTRODUCTION

According to the U.S. Department of Transportation, a motorist is 40 times more likely to be killed if involved in a vehicle-train crash than in any other type of highway collision. There are approximately 261,000 highway-rail and pedestrian crossings in the United States according to the studies by the National Highway Traffic Safety Administration (NHTSA) and Federal Railroad Administration (FRA). From 1993 to 1998, there were over 25,000 highway-rail crossing incidents involving motor vehicles-averaging 4,167 incidents a year. In the US alone, a train collides with a vehicle or person once every 115 minutes, and in an average year, more people die at highway-rail crossings than in commercial airline crashes. According to the FRA's Railroad Safety Statistics Annual Report (1998), there were 75 highway-rail grade crossing incidents in Florida in 1998, resulting in 7 fatalities. Thus, there is significant motivation for the exploration of the use of innovative technologies to monitor railroad grade crossings. A number of cities worldwide now employ camera-based public monitoring systems. Cities in the U.S. include: Baltimore, Newark, Tampa, Virginia Beach, Memphis, Tacoma, Hollywood, Anchorage, San Diego, and Fort Lauderdale. These systems are first generation, and consist of a set of video cameras feeding into a central station that must be staffed by an attentive human monitor. Their effectiveness and response is largely determined not by the technological capabilities or placement of the cameras but by the vigilance of this person. Recently some commercial video monitoring systems have started to appear in the market. However, these systems are limited in their capabilities, need a large number of training samples, work only in predefined scenarios, do not have learning capability, and are expensive. Another problem with these systems is that they mostly employ only static information in the images, and make use of shape properties of objects of interest. However, these shape-based approaches encounter difficulties in scenes containing clutters, during occlusion, shadows etc. Also, these systems do not have the capability to analyze the movements and activities of pedestrians, bikers, and animals, which may be very useful for railroad grade crossings.

Further author information: (Send correspondence to Mubarak Shah)
Mubarak Shah: E-mail: shah@cs.ucf.edu, Telephone: 1 407 823 5077

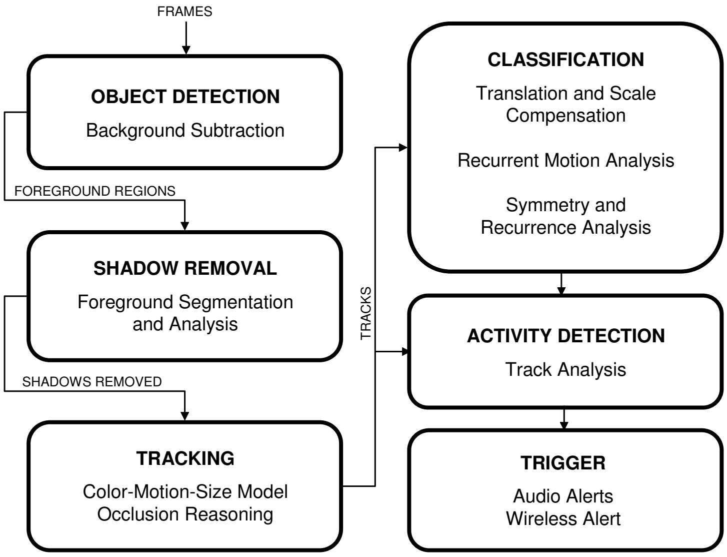


Figure 1. Components of the visual surveillance system.

In this work, we present a 'smart' real-time computer vision system, using only COTS hardware, for monitoring motions of pedestrians, bikers, animals, vehicles, etc. The computer vision system monitors a railroad grade crossing constantly and detects any moving object, tracks the object and analyzes its shape and motion. Based on this analysis, at the onset of predefined events, an active device is triggered to warn pedestrians. With the availability of low cost, low power, and high quality video cameras, increased computing power and memory capacity, and recent success in developing computer vision algorithms the proposed approach provides a cost effective, deployable solution to this important problem. The system has impact beyond monitoring railroad crossing, the research results can be used for homeland security, counter drug interdiction, border patrol, etc. The presented system is scalable to the use of multiple cameras (with or without overlapping fields of views) in order to monitor larger areas.[1, 2]

2. RELATED WORK

A large number of monitoring systems have been proposed in recent years. For example, Pfinder[3] uses a single gaussian background model to locate interesting objects in the scene. The full body of a person is tracked based on the assumption that only one person is present in the scene. Stauffer and Grimson[4] proposed an adaptive multi-modal background subtraction method that can deal with slow changes in illumination, repeated motion from background clutter and long term scene changes as opposed to unimodal background model of Pfinder. The detected objects were tracked using a multiple hypothesis tracker and common patterns of activities were statistically learned over time. Ricquebourg and Bouthemy[5] proposed tracking people by exploiting spatiotemporal slices. Their detection scheme involved the combined use of intensity, temporal differences between three successive images and comparison of the current image to a background reference image which is reconstructed and updated online. Recovered spatiotemporal signatures of each object were used to perform classification between persons and vehicles. In W4,[6] dynamic appearance models were used to track people. Single person and groups were distinguished using projection histograms. Shadows in the scene cause many problems in computer vision based monitoring systems. To deal with shadows, Pfinder[3] performs the background subtraction in normalized color space, whereas, Horprasert et el.[7] proposed a color model that separates the brightness from the chromaticity component in RGB space. However, these normalized color spaces only deal with light shadows and cannot handle the illumination variation in strong shadows.

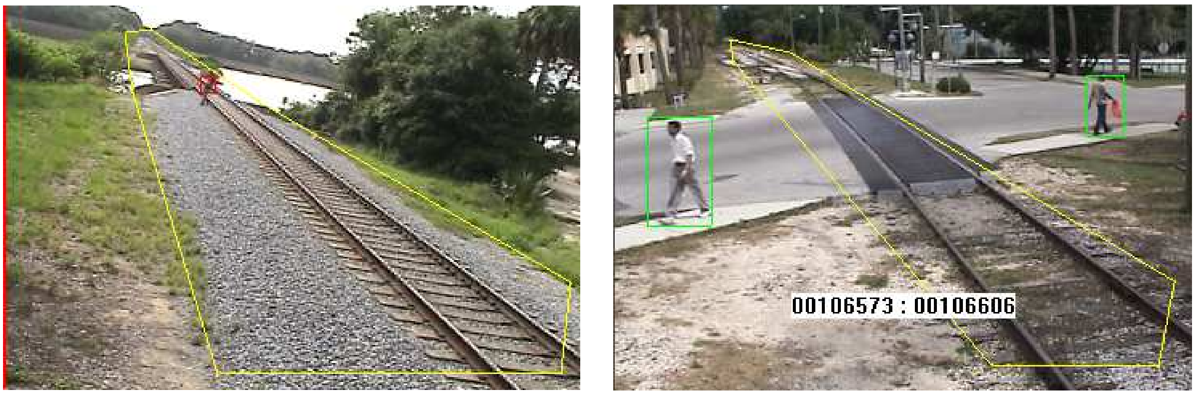


Figure 2. Detecting entry into the danger zone. A danger zone is defined through a GUI and objects are warned as they enter the danger zone. The bounding box of individuals turn red as they enter the polygon defining the danger zone.

3. VISUAL MONITORING

The visual monitoring system used in the surveillance of the railroad grade crossing consisted of six main modules: (1) Object Detection, (2) Shadow Removal, (3) Tracking, (4) Classification, (5) Activity Detection and (6) Triggering Decision. Object detection is performed using a cascade coupling of color based background subtraction and gradient based confirmation as detailed in.[8]

3.1. Object Detection and Shadow Removal

Pixel-wise statistical models of color are first used to detect candidate foreground pixels. Candidate regions are then tested for gradient-based foreground pixels at their boundaries. The pixel-level models of both color and gradient are then updated based on this classification process. This approach provides robust detection in the presence of common phenomenon, such as quick illumination changes and cloud shadows, repositioning of static objects, and the initialization of the background model with moving objects present in the scene. Such phenomenon are not handled by even the most current background modelling schemes. In order to deal with the shadow problem, we use a shadow model, using the observation that the color of the shadow at a point is largely independent of the object generating the shadow. The estimate of the shadow is improved with time, by noting that the pixel intensity is a mixture of 3 distributions: the background Gaussian; the shadow Gaussian, which has a lesser mean than the background Gaussian; and the object Gaussian, which has low weight and high variance. For every pixel, another class is generated to represent shadow: thus a pixel may be assigned to the background class, the shadow class, or any of the foreground classes.

3.2. Tracking and Classification

Since object detection is being performed per frame, tracking is essentially the task of establishing correspondence between detected objects across frames. Each object is modelled by color and spatial pdfs, a Gaussian distribution to represent spatial position (variance equal to the sample variance of the object silhouette), and the color is represented by a normalized histogram. To establish correspondence, each pixel detected in a subsequent frame votes for membership to a single model, based on its color and position, and a region is corresponded to the object to which more than k_p votes go. If k_p pixels vote for more than two objects, it is concluded that occlusion is occurring. Using a feature vector called the 'Recurrent Motion Image',[9] the periodicity of behavior of objects is observed, provided a means to distinguish between cars (with no periodicity) and humans (periodic motion of arms and legs during walking).

3.3. Monitoring Railroad Grade Crossing

The domain of monitoring railroad grade crossing has specific surveillance requirements. For a situation of 'interest', a specific combination of events need to occur, namely the train should be approaching while a pedestrian, vehicle or animal enters within a certain boundary of the railroad. In our system, since a stationary camera is placed for surveillance, an area can be demarcated at the time the system is setup using the GUI. Figure 2 show two scenes with the yellow boundary displaying the so-called 'danger-zone'. In the figure, the bounding boxes of objects within the danger zone are red, while the bounding boxes of objects outside the danger zone are green. The system receives two inputs, one from the traffic signal (triggered when a train approaches) along with visual input on the position of pedestrians and vehicles with respect to the danger zone. At the correct combination of events, a rule-based algorithm detects activities based on the object classification, track patterns, and the object model. Speed, direction and orientation of silhouettes, with respect to each other, provide a language in which to express events of interest. A warning setup is attached to the system, to trigger an alarm at the onset of an undesirable event. An audio alert is generated, and since the system is online, an email or message can be sent to an authorized individual.

4. EVALUATION

To test the performance of different features of the system (detection, tracking and classification), we obtained six hours of videos at two different locations in Central Florida, with a history of trespasser violations. The data set is composed of 5 videos taken from different views and in different conditions (time of day, lighting, wind, camera focus, traffic density etc.). The duration and number of objects (persons/vehicles) in each data set are shown in Table 1.

Table 1. Data Sets Used for System Evaluation.

	Duration (minutes)	Number of Interesting Objects
Data Set 1	65	85
Data Set 2	72	171
Data Set 3	80	215
Data Set 4	70	160
Data Set 5	65	94

We ran the system on all these videos and assigned the following attributes to each moving object in the field of view (The same system parameters were used for *all* tested data.):

1. Whether the object was correctly detected or not.
2. Whether the tracking was performed correctly or not. Note that we only say that tracking is correct if the tracking was completely correct over the time the object was present in the field of view.
3. Whether the classification of object (Person, Vehicle or Group) was correct or not and what was the incorrect label that was assigned to it.

Accuracies of Detection, Tracking and Classification were then computed as follows:

$$\text{Detection Accuracy} = \left(\frac{\text{Number of Correct Detection}}{\text{Total Number of Object}}\right) \tag{1}$$

$$\text{Tracking Accuracy} = \left(\frac{\text{Number of Completely Correct Tracks}}{\text{Number of Correct Detections}}\right) \tag{2}$$

$$\text{Classification Accuracy} = \left(1 - \frac{\text{Number of Correct Classifications}}{\text{Number of Correct Detections}}\right) \tag{3}$$

While the measure *Detection Accuracy* (Recall) indicates the portion of the relevant objects, which were detected, another measure *Detection Precision* indicates the portion of the relevant objects under the detected objects, and is defined as follows:

$$\text{Detection Precision} = \left(\frac{\text{Number of Correct Detections}}{\text{Total Number of Detections}}\right) \tag{4}$$

The performance of the three tasks, detection, tracking and classification is tabulated in Table 2, while the values of accuracy and precision measures are shown in Table 3. The performance of detection module is graphically depicted in figures 3 and 4. It can be seen that the detection module achieves both recall and precision rates of more than 0.95 in all of the data sets. Similar graphs are provided for Tracking performance (See Figure 5) and Classification performance (See Figure 6). Once again, the tracking accuracy is better than 95% in most of the cases, whereas, the system achieves the classification accuracy of better than 90% for three of the data sets. Maintaining classification accuracy consistently over 90% is part of our future endeavors.

Table 2. System Performance.

	No. of Objects	Correct Detections	False Positives	Correct Tracking	Correct Classification
Data Set 1	85	85	2	83	78
Data Set 2	171	167	5	167	162
Data Set 3	215	206	10	191	187
Data Set 4	160	156	4	156	120
Data Set 5	94	92	2	92	75

Table 3. Detection Accuracy.

	Detection Accuracy	Detection Precision	Tracking Accuracy wrt Detection	Classification Accuracy wrt Detection
Data Set 1	1.00	0.98	0.98	0.94
Data Set 2	0.98	0.97	1.00	0.97
Data Set 3	0.96	0.95	0.93	0.98
Data Set 4	0.975	0.975	1.00	0.77
Data Set 5	0.98	0.98	1.00	0.82

5. CONCLUSION

In this paper, we outlined a general visual monitoring system, and show it's application for the visual surveillance of railroad grade crossing. We collected 6 hours of video data around rail crossings by COTS camcorders. The use of advanced video surveillance technology in monitoring rail road crossing has potential to decrease the number of accidents and resulting deaths and injuries that occur due to pedestrian crossings. With the continued improvement in computer vision algorithms, and due to availability of low cost, low power, high quality video cameras, and increased computing power the proposed approach provides a cost effective solution for monitoring railroad crossings. The monitoring system can be operational 24 hours day, 7 days a week without any interruption. In addition, the technology can also be used for monitoring job site security, preventing vandalism e.g. stealing of scarp aluminum at maintenance yard, constructions sites around major bridges, etc.

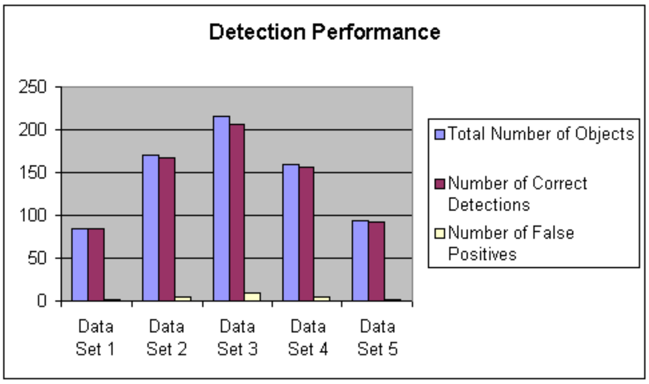
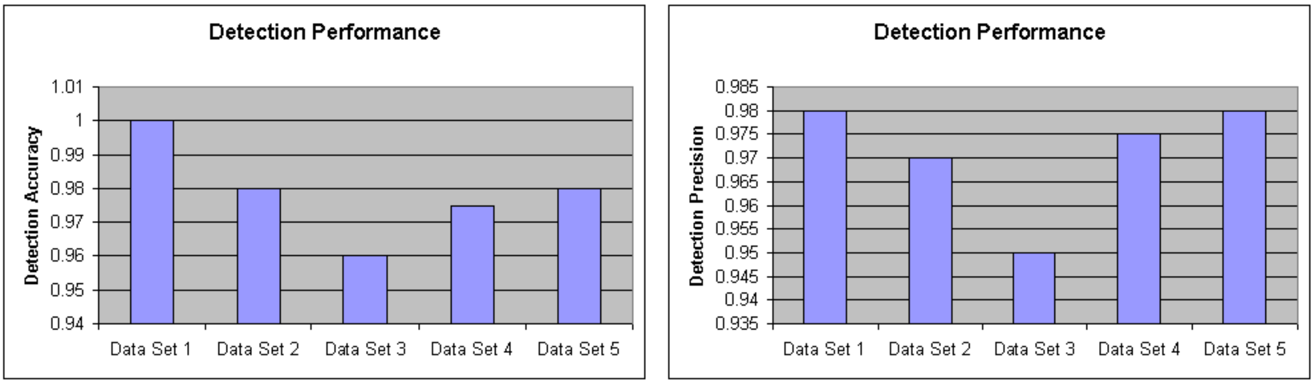


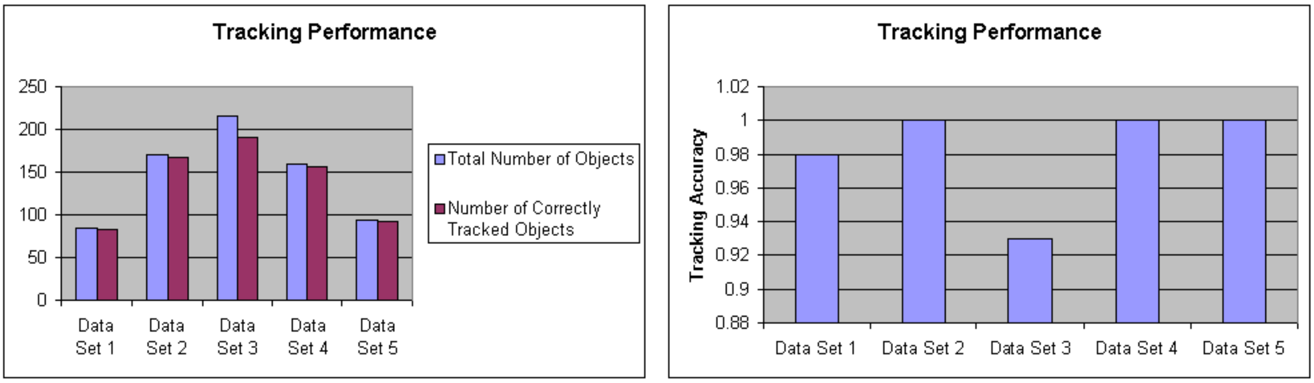
Figure 3. Detection Performance: The graph shows the number of correct detections compared to the total number of objects in the five data sets.



(a)

(b)

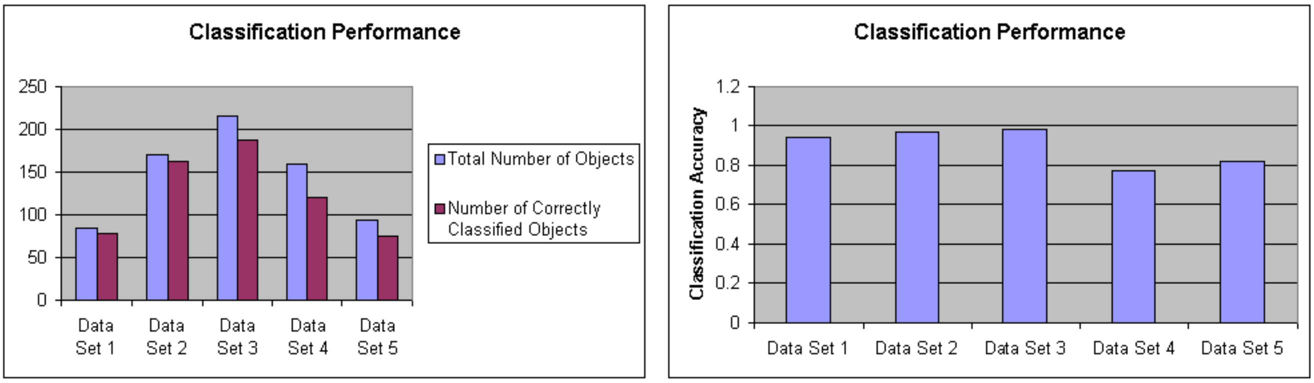
Figure 4. Detection Performance Plots. (a) This graph shows the detection accuracy measure in the five data sets. (b) Detection precision in each of the data sets.



(a)

(b)

Figure 5. Tracking Performance Plots. (a) This graph shows the number of correct tracking compared to the total number of objects in the five data sets. (b) Tracking accuracy in each of the data sets.



(a) (b)

Figure 6. Classification Performance Plots. (a) This graph shows the number of correct classifications compared to the total number of objects in the five data sets. (b) Classification accuracy in each of the data sets.

ACKNOWLEDGMENTS

This work was supported by the Florida Department of Transportation. The authors would like to thank Capt. Garland Young and Mr. Biscan for their cooperation and assistance, and Mr. Donald Harper and Mr. Robert Traub for the engineering of the system hardware.

REFERENCES

1. O. Javed, Z. Rasheed, K. Shafique, and M. Shah, "Tracking across multiple cameras with disjoint views," *The Ninth IEEE International Conference on Computer Vision, Nice, France* , 2003.
2. O. Javed, Z. Rasheed, O. Alatas, and M. Shah, "Knight^M: A real time surveillance system for multiple overlapping and non-overlapping cameras," *The fourth International Conference on Multimedia and Expo (ICME 2003), Baltimore, Maryland* , 2003.
3. C. Wren, A. Azerbayejani, T. Darrel, and A. Pentland, "Pfinder: Real time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , July 1997.
4. C. Stauffer and E. Grimson, "Learning pattern of activity using real time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , August 2000.
5. Y. Ricquebourg and P. Bouthemy, "Real time tracking of moving persons by exploiting spatiotemporal image slices," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , August 2000.
6. I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , August 2000.
7. T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real time robust background subtraction and shadow detection," *IEEE Frame Rate Workshop* , 1999.
8. O. Javed, K. Shafique, and M. Shah, "Hierarchical approach to robust background subtraction using color and gradient information," *Proc. IEEE Workshop on Motion and Computing* , 2002.
9. O. Javed and M. Shah, "Tracking and object classification for automated surveillance," *Proc. European Conference on Computer Vision* , 2002.