ITS INSTITUTE
Intelligent Transportation Systems

Monitoring and Use of HOV and HOT Lanes

Final Report

Prepared by:

Eric Holec
Guruprasad Somasundaram
Nikolaos Papanikolopoulos
Vassilios Morellas

Digital Technology Center
Department of Mechanical Engineering
University of Minnesota

CTS 13-07

Technical Report Documentation Page

1. Report No. CTS 13-07	2.	3. Recipients Accession No.
4. Title and Subtitle Monitoring the Use of HOV and HOT Lanes		5. Report Date January 2013 6.
7. Author(s) Eric Holec, Guruprasad Somasundaram, Nikolaos Papanikolopoulos, and Vassilios Morellas		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Minnesota Digital Technology Center Department of Mechanical Engineering 499 Walter Library 117 Pleasant Street SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project #2011079 11. Contract (C) or Grant (G) No.
12. Sponsoring Organization Name and Address Intelligent Transportation Systems Institute Center for Transportation Studies University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplementary Notes http://www.its.umn.edu/Publications/ResearchReports/		

16. Abstract (Limit: 250 words)

This report presents the formulation and implementation of an automated computer vision and machine learning based system for estimation of the occupancy of passenger vehicles in high-occupancy vehicles and high-occupancy toll (HOV/HOT) lanes. We employ a multi-modal approach involving near-infrared images and high-resolution color video images in conjunction with strong maximum margin based classifiers such as support vector machines. We attempt to maximize the information that can be extracted from these two types of images by computing different features. Then, we build classifiers for each type of feature which are compared to determine the best feature for each imaging method. Based on the performance of the classifiers we critique the efficacy of the individual approaches as the costs involved are significantly different.

17. Document Analysis/Descriptors High occupancy vehicle lanes, High occupancy toll lanes, High occupancy vehicles, Monitoring, Vehicles, Traffic counting, Infrared detectors, Vehicle occupants		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report)	20. Security Class (this page)	21. No. of Pages 26	22. Price

Monitoring the Use of HOV and HOT Lanes

Final Report

Prepared by:

Eric Holec
Guruprasad Somasundaram
Nikolaos Papanikolopoulos
Vassilios Morellas

Digital Technology Center
Department of Mechanical Engineering
University of Minnesota

January 2013

Published by:

Intelligent Transportation Systems Institute
Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Avenue SE
Minneapolis, Minnesota 55455

Acknowledgments

The authors wish to acknowledge those who made this research possible. The study was funded by the Intelligent Transportation Systems (ITS) Institute, a program of the University of Minnesota's Center for Transportation Studies (CTS). Financial support was provided by the United States Department of Transportation's Research and Innovative Technologies Administration (RITA).

Table of Contents

List of Figures

List of Tables

Executive Summary

The goal of this project is to investigate the use of inexpensive commercially available video cameras coupled with sophisticated data processing software to automate the process of estimating the occupancy of passenger vehicles on high-occupancy vehicle and high-occupancy toll (HOV/HOT) lanes. Such a technology can potentially help enforce the HOV/HOT lanes more effectively. Violators could be reliably identified thereby increasing revenue. Statistics about violators and technologies like the one investigated in this work can prove useful while making crucial decisions about infrastructure management and planning. Our approach involves the use of two different imaging modalities: near-wave infrared spectrum and optical video. Dual-band near-wave infrared images proved to be useful for differentiating human skin from other infrared reflective materials. High-resolution color video images were useful for segmenting human-like objects based on skin color, face-like features etc. Our data processing software deployed state-of-the-art classification procedures on both types of images and it was used to obtain estimates of passenger vehicles occupancy.

1. Introduction

The goal of this project is to develop a system that uses multiple imaging modalities to assess the occupancy of passenger vehicles in high-occupancy vehicle and high-occupancy toll (HOV/HOT) lanes. Previously, in a work that involved the University of Minnesota and Honeywell labs, the authors used a sophisticated tri-band near infrared imaging device to acquire high-quality images, from which the vehicle occupancy could be estimated. Based on their study, the skin reflects infrared in the band below 1400 nm and absorbs above it. Using this observation, they integrated into one system two co-registered near infrared cameras with suitable IR filters to capture images in the two bands. Subtraction of the images from the two IR bands was used to segment out the skin-like regions and then an estimate of the number of occupants could be calculated. Even though this system was very effective, it was rather expensive and complicated making its field deployment a difficult task. In this work we investigated an inexpensive system that produces similar results. While the cost of the proposed system is relatively lower, we faced a bigger challenge while processing the images owing to their inferior quality images acquired and to stereo artifacts. However, since the original effort by our team with Honeywell labs, there has been a rapid improvement in computer vision and machine learning methods for the problem of object classification in signals and images. Software image processing methods were thus used to compensate for the loss of information resulting from the use of inexpensive cameras. We have been able to apply powerful algorithms like support vector machines [1] to the acquired infrared images to extract as much discriminatory information as possible. We formulate the problem as a binary classification problem where the two classes are: (i) vehicles with just one occupant and (ii) vehicles with more than one occupant. We use different image features for infrared and color video images. The near infrared cameras were not low light cameras and could be used only in daylight conditions. In addition, their resolution was also poor. The color video cameras have a much higher quality and frame rate and this facilitated in-depth data processing. By comparative analysis of the performance of the two approaches, we are able to assess the efficacy of the imaging modalities. In short, what we lacked in terms of high-cost sophisticated equipment was compensated by the more advanced algorithms deployed.

Our report is organized as follows; we initially discuss some pertinent related work about the reflectance pattern of human skin and about object classification methods. We detail the hardware setup and our choice of sensors. This is followed by a description of our algorithmic approach. Finally we provide some experimental results and discussion thereof.

2. Related Work

Using conventional video, differentiating people from dummies is extremely challenging. A well designed dummy can be nearly indistinguishable from a real person and even simple printed images are capable of fooling people identification techniques. People detection in vehicles become even more challenging due to existing features in a vehicle such as seats with headrests whose silhouettes are similar to those of passengers. To overcome these challenges non-visible spectrum cameras can be used.

Human skin exhibits a unique reflectance pattern in the near-infrared spectrum [2]. From 900 nm to 1300 nm, skin's reflectance ranges from 25% to 60%. After 1300 nm, the reflectance of skin rapidly decreases to a low point of 10% at 1400 nm. These characteristics can be primarily attributed to the high content of water in skin. A more detailed analysis of this phenomenon can be found in [2].

Since other inanimate objects in a vehicle, such as leather or artificial dummies, do not contain significant amounts of water, they do not absorb nearly as much infrared light above 1400 nm. By taking near-infrared images both above and below 1400 nm, the parts of the image corresponding to skin can be identified.

Figure 1. A pair of images taken by the infrared cameras is shown. The left image captures the wavelengths from 900 nm to 1400 nm, while the right image is from 1400 nm to 1700 nm. Most objects appear similar in the two images, while skin is significantly less reflective in the higher wavelengths.

Previous work by Pavlidis and Morellas has used this difference in the infrared reflectance of skin to count occupants in vehicles [3]. In their previous system, two near-infrared cameras where mounted such as a beam splitter separated the lower and upper near-infrared bands. The thresholded difference of the two images was then supplied to a neural network classifier to determine whether there were one or two occupants within a vehicle.

Since the time that this work was published in 2000 camera technology, computers, and algorithms, have all undergone significant advancements. These new technological and software developments warrant a new research that challenges the previous system although this past work provides a strong reference for our implementation. We use support vector machines, which are extremely effective in performing classification in many different kinds of feature spaces using different kernel types [1] [4] [5]. SVM based approaches have been prominent in supervised learning applications. Not limited to object classification, SVMs have been an attractive choice for many other classification and regression applications and thus have been used with many different object features with success. A detailed report is available in [5].

3. Hardware and Experimental Setup

Infrared Image Acquisition

To differentiate people from other objects in a vehicle, images from the near infrared bands both above and below 1400 nm needed to be collected. An ideal camera would be able to capture these two bands simultaneously, but no such cameras are commercially available. Instead, a stereo pair of cameras was used.

A pair of Alpha NIR cameras was used in the final setup. They have a resolution of 320 by 256 pixels and can capture images at a rate of 30 Hz. The spectral response range for these cameras is from 900 nm to 1700 nm. These values are typical for cameras of this type and the Alpha NIR camera was chosen among several candidates, primarily due to cost and availability.

Each camera was fitted with an identical fixed focus lens. The focal length was 50 mm, providing an 18° field of view. This field of view was chosen so that a car 10 to 15 meters from the camera would occupy approximately half of the image. If the camera needed to be placed further from the roadway, a longer focal length (and therefore narrower field of view) lens could have been used. Additionally, the lenses were specifically designed to minimize distortions in near infrared imagery.

To separate the upper and lower infrared bands, each of the two cameras used a different filter placed in front of them. The first camera used a short wave-pass filter so that the camera would only receive the wavelengths below 1400 nm. The second camera used a long wave-pass filter so that it would register wavelengths above 1400 nm. With these filters, the first camera had a response range of 900 nm to 1400 nm and the second from 1400 nm to 1700 nm.

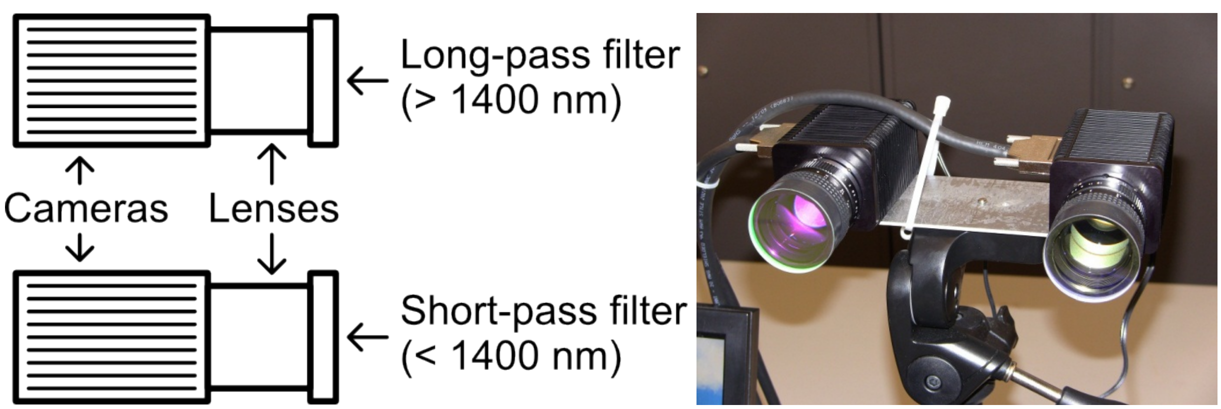

Figure 2. A diagram view and photograph of the experimental infrared camera rig. The two cameras have different filters to isolate the parts of the near-infrared spectrum of interest.

The cameras were mounted on a steel plate so that their optical axes are parallel. The plate was then mounted on a tripod. The cameras were positioned at chest height, to the side of the roadway. The assembly was angled so that a vehicle in the center of the image was approximately between 10 and 15 meters away. This positioning provided an unobstructed view of the front seats of each vehicle.

Figure 3. In the left a picture of the camera rig setup adjacent to a roadway is shown. The right image shows an image from one of the cameras in this setup. The location was chosen so that the cameras would have an unobstructed view of the front two seats in passing vehicles.

In addition to the cameras, a desktop computer was deployed to collect the images. Installed in the computer were two National Instruments PCI camera capture cards, necessary to communicate with the cameras. Other short wave infrared cameras have similar requirements. Some of the newer cameras did however have more modern communication methods such as USB. This would have allowed for a more compact setup, as a full desktop computer would not have been acquired.

The power requirements for the system came primarily from the desktop PC used. A typical PC would be expected to require 200 watts of power. The Alpha NIR camera consumes less than 5 watts each and includes power adaptors for standard 110V plugins.

Images were acquired near midday so that natural illumination would be at its peak. The location was chosen so that sunlight would enter the front of the vehicle. This did cause some issues as drivers often had their sunshades lowered or the structure of the roof blocked direct light from the upper half of the driver's face.

Color Image Acquisition

Typical color video images were obtained using a canon HF S 20 HD digital camcorder shown below.

Figure 4. Canon camcorder used for color video acquisition.

The cost of this camera is around $1000 dollars and it includes a hard disk from which the video data can be acquired offline and processed.

4. Algorithms and Software

We have two imaging modalities: the near-infrared and color images. Our approach for the two different image types is slightly different. However, in both cases we focused on the windshield area of the vehicle.

Near Infrared

For the near infrared we acquired a stereo pair of images providing information in the near-infrared wavelengths above and below 1400 nm. A sample pair is shown in figure 5.

Figure 5. A sample pair of near infrared images acquired using our hardware. We can notice the difference in the reflectance of the skin in the two images.

We manually extracted the windshields from the two images. Even though this step can be automated, our goal was to minimize the detection and registration error. At this point of time, our main focus is to evaluate the usefulness of infrared imagery vs. conventional video imagery. After identifying the windshields, an image registration step was performed to correct for the affine transformation between the two viewpoints. The scaling, rotation, and translation that minimized the squared difference (least squares problem) between the two images of the windshield were found via a gradient decent method. This made the delineation of the windshields much more robust to errors. After the registration the windshields are brought to the same view point and scale as shown in figure 6.

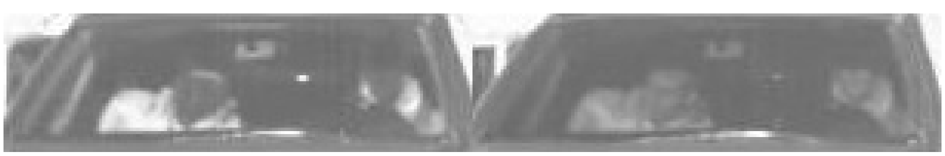

Figure 6. The pair of registered windshield images is shown. These images can be used in the same frame of reference.

Given the pair of images in the same frame of reference, we can compare them explicitly by taking their difference. The difference image shown in figure 7 can be used as a feature for estimating the number of passengers.

Figure 7. The difference between the registered images can be used directly as a classification feature.

We can compute different features with the image pair shown in figure 6 and the difference image henceforth referred to as the blob shown in figure 7. We compute three types of features: the vectorized blob image resized to different sizes, the vectorized image pair shown in figure 6 of the windshields, and the blob statistics feature F for the blob image:

$$F = \begin{bmatrix} m_u & \sigma & \frac{m_{u3}}{\sigma} & \frac{m_{u4}}{\sigma} & \frac{m_{u5}}{\sigma} \end{bmatrix}$$

Where m_u is the mean of each row in the image, σ is the standard deviation, $\frac{m_{u3}}{\sigma}$, $\frac{m_{u4}}{\sigma}$, and $\frac{m_{u5}}{\sigma}$ are the third, fourth, and fifth moments about the mean (third and fourth moments are the skew and kurtosis). The first two features encourage completely data driven classification whereas the third converts the data into statistical data.

For the vectorized feature, the size of the blob determines the length of the feature. So we resized all blobs to a fixed predefined size. This was also varied and determined based on the accuracy of the classification. More discussion is provided in the results section.

Color Images

For the color images, we extract the windshield area manually as we did with the infrared images. We have a much higher resolution color image shown in figure 8 as compared to figure 5. Given this image we can compute a number of popular image features such as vectorized gray scale image (similar to blob in near infrared), histogram of oriented gradients [6], skin color features [7].

Figure 8. A sample RGB color image of a windshield. Face has been intentionally masked out to obscure identity.

Figure 9. A skin color segmentation example Y > 60, 90 < Cb < 140, 120 < Cr < 165.

We estimate the skin color pixels according to the rule in [7] based on YCbCrRGB color space data:

(Y, Cb, Cr) is classified as skin if:

Y > ymin and cbmax > Cb > cbmin and crmax > Cr > crmin

The different thresholds ymin, cbmin, cbmax, crmin and crmax are determined based on evaluating the precision and recall of skin color segmentation when these values are varied. The values that yielded the best average results based on some training images are used. *Ymin = 60, cbmin = 90, cbmax = 140, crmin = 120, crmax =165.*

Histogram of oriented gradients (HoG) measures the distribution of edge intensities and orientations. These are considered powerful descriptors for the problem of object detection and classification especially for human detection and classification. We used an angle range of 180 degrees, 8 bins as the choice of parameters for HoG computation.

Classification

Based on the different features, classification between vehicles with one occupant and vehicles with more than one occupant can be obtained through many discriminative models. The KNN (K nearest neighbor) algorithm provides the simplest way to discriminate data of different types. Nearest neighbor classification is performed by saving all the training data, and classification is done by finding the K nearest matches based on a distance metric of choice. For our experiments we use the Euclidean distance.

The support vector machine acts differently. Instead of keeping the training data, it identifies the extreme or support vectors (data samples) of each class. These support vectors are in fact a projection of the original data points onto a decision boundary, which can be a linear, polynomial or exponential function. These functions are called kernels. The choice of the kernel and their parameters has a significant impact on the performance of the classifier. We determine the best parameters by a cross-validation. We split the labeled training data into parts (folds). The classifier is trained on two parts and tested on the remaining part. This process is repeated by changing the parts used for training and testing. Eventually we take the average accuracy across all parts. Note that we can also do cross-validation for KNN or any type of classifier.

5. Experimental Results and Discussion

Near Infrared Results

We manually extracted 158 near infrared windshield samples. 110 of those were single occupant examples and 48 samples had more than one occupant. After the registration, the blob feature vector based on the difference between the two bands of observation, the feature vector of the individual views concatenated without computing the difference, and finally the moment statistics of the blob features were computed. For each pass of testing, while computing the feature vector the sizes of the blobs and windshields were varied from 6 X 20 to 14 X 30 resulting in a feature length of 120 dimensions to 420 dimensions. We performed three-fold cross-validation using both KNN and SVM classifiers. For SVM we used linear, radial basis function and polynomial kernel to determine the best choice for our scenario. We report the best-case accuracies in table 1.

Table 1. Near infrared classification results with best parameter choices.

Feature	K nearest neighbor	Linear SVM	RBF SVM	Polynomial SVM
Blob vector	69.93%	82.91%	76.58%	79.74%
Both windshields vectorized	66.78%	72.78%	68.98%	70.88%
Blob statistics	64.57%	74.05%	74.68%	73.41%

We obtain best classification results with a linear SVM using the basic blob vectorized feature. The difference image by itself is a very good discriminant between single occupant and multiple occupant vehicles.

Color Results

For this experiment we used 122 labeled color windshield samples. Of these 40 samples were of vehicles with more than one occupant and 82 were single occupant vehicles. The gray scale, skin color segmented image feature and the HoG feature were computed. Since the resolution of the color images was approximately four times higher we scaled the windshield regions to sizes ranging between 24 X 80 and 56 X 120. For the HoG computation the blob was resized to size bet.

Table 2. Color image results with best parameter choices.

Feature	K nearest neighbor	Linear SVM	RBF SVM	Polynomial SVM
Gray Scale	71.39%	73.77%	**86.06%**	81.14%
Skin Color	80.74%	67.21%	79.50%	**88.52%**
HoG	68.44%	**78.68%**	76.22%	**78.68%**

The thresholds we determined for skin color segmentation, seems to have produced significantly good results. We obtained 88% accuracy with polynomial SVM using skin color features.

Based on both results we observe a common trend, K nearest neighbors yields inferior results in both cases. This is because just based on Euclidean distance the data is rarely separated. Also the color results are better than the infrared. Even though the resolution of the color images was superior, the cost of the color camcorder is around 1000 dollars as opposed to a full-fledged near infrared acquisition system, which costs more than 50000 dollars. We could have obtained significantly better results with an infrared illuminator albeit with additional costs.

Using only natural light, the usefulness of near infrared is very limited. Only during specific times of the day and weather conditions there was enough illumination to see into vehicles. Even during these times, strong shadows caused by sun blocks and car roofs limited to visible area within a vehicle. For a more practical system, artificial lighting is necessary. Near infrared has the distinct advantage over conventional video in that illumination can be provided without distracting the driver. It is invisible to the human eye and in general lighting situations, has no risk to individuals.

Recent changes in auto-glass regulations may limit the long-term viability of an infrared system. In 2009, California passed legislation that would mandate the use of infrared reflective glass in all vehicles sold in that state, starting in 2014 [8]. One supplier of such infrared reflective coatings is Southwall Technologies. Their XIR laminated glass rejects more than 94% of infrared radiation [9]. It would be virtually impossible to provide sufficient infrared illumination to see passengers in such a vehicle as the light must pass through the glass twice (both when entering the vehicle and exiting it). It has been our observation that some of the currently available vehicles already have similar infrared reflective glass as in our data collection; there were several instances where there was no apparent infrared illumination within the vehicles.

Figure 10. Infrared reflective glass prevents the majority of infrared illumination from entering a vehicle. Currently this technology is uncommon in the United States, but may increase in the future due to its energy savings.

6. Conclusions

We developed a system for acquiring near infrared images in wavelengths above and below 1400 nm separately. Using these two images we used a support vector machine algorithm for predicting whether a car has a single occupant or more. Based on practical experimentation we obtained an accuracy of 82%. We also developed an equivalent system with a HD color video camcorder. We obtained higher performance by using the color video images; as high as 88%. This leads us to debate the usefulness of the near infrared system when compared to the color video images given the enormous cost difference. Even though the color video cameras cannot be used in low light, we could not capture good quality images even with the IR cameras. They required significant IR illumination. The system can be more useful with an infrared illuminator although the costs incurred will be even higher.

Future Work

Due to the prohibitively high costs for renting the infrared equipment, only one view could be observed with the available equipment. In most cases it might be sufficient to determine if the front passenger seat is occupied by a person to ensure that the HOV lane rules are respected. Recently we have acquired a complete high-resolution tri-band camera system courtesy of Honeywell, which allows us to acquire better quality images, which can potentially be positioned to get a fair amount of the front windshield as well as the side window.

Due to logistics and planning the power requirements for the system, testing was conducted in the vicinity of a power source in a building. Also the equipment was available for too short a period (rental rates limited us to 2 weeks with the equipment) to plan scheduled testing on freeways. With the available tri-band system, we can conduct more extensive testing in freeways for evaluation in higher speeds. Also, since the existing methods were tested at ~20 - 40 mph, we do not anticipate a huge difference in the performance at higher speeds. This eventually will be more dependent on the quality of the camera (such as the shutter rate, aperture size, etc.) rather than the algorithm.

References

[1] C. Cortes and V. Vapnik, "Support Vector Networks," *Machine Learning*, pp. 273-297, 1995.

[2] A. Nunez and M. Mendenhall, "Detection of Human Skin in Near Infrared Hyperspectral Imagery," *International Geoscience and Remote Sensing Symposium*, Boston, MA, 2008.

[3] I. Pavlidis, V. Morellas and N. Papanikolopoulos, "A Vehicle Occupant Counting System Based on Near-Infrared Phenomenology and Fuzzy Neural Classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 2, pp. 72-85, 2000.

[4] H. Zhang, A. Berg, M. Maire and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition," *Computer Vision and Pattern Recognition*, vol. 2, pp. 2126-2136, 2006.

[5] J. Zhang, M. Marszalek, S. Lazebnik and C. Schmid, "Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, New York, NY, 2006.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[7] V. Vezhnevets, V. Sazonov and A. Andreeva, "A survey on pixel-based skin color detection techniques," *Proceedings Graphicon*, 2003.

[8] M. Coda, "California mandates use of IR reflective auto glass," *Glass Magazine*, 11 September 2009. [Online]. Available: http://www.glassmagazine.com/article/auto/california-mandates-use-ir-reflective-auto-glass. [Accessed 30 December 2011].

[9] "XIR Laminated Glass," Southwall Technologies, 2011. [Online]. Available: http://www.southwall.com/southwall/Home/Automotive/Products/XIRLaminatedGlass.html. [Accessed 30 December 2011].